Evaluating the Operational Application of SMAP for Global Agricultural Drought Monitoring

Iliana E. Mladenova, John D. Bolten, Wade T. Crow, Nazmus Sazib, Michael H. Cosh, Compton J. Tucker, and Curt Reynolds

***Abstract*—Over the past two decades, remote sensing has made possible the routine global monitoring of surface soil moisture. Regional agricultural drought monitoring is one of the most logical application areas for such monitoring. However, remote sensing alone provides soil moisture information for only the top few centimeters of the soil profile, while agricultural drought monitoring requires knowledge of the amount of water present in the entire root zone. The assimilation of remotely sensed soil moisture products into continuous soil water balance models provides a way of addressing this shortcoming. Here, we describe the assimilation of NASA's soil moisture active passive (SMAP) surface soil moisture data into the United States Department of Agriculture Foreign Agricultural Service (USDA FAS) Palmer model and assess the impact of SMAP on USDA FAS drought monitoring capabilities. The assimilation of SMAP is specifically designed to enhance the model skill and the USDA FAS drought capabilities by correcting for random errors inherent in its rainfall forcing data. The performance of this SMAP-based assimilation system is evaluated using two approaches. At global scale, the accuracy of the system is assessed by examining the lagged correlation agreement between soil moisture and the normalized difference vegetation index (NDVI). Additional regional-scale evaluation using *in situ*-based soil moisture estimates is carried out at seven of the SMAP core Cal/Val sites located in the USA. Both types of analysis demonstrate the value of assimilating SMAP into the USDA FAS Palmer model and its potential to enhance operational USDA FAS root-zone soil moisture information.**

***Index Terms*—Agricultural drought, data assimilation, hydrologic modeling, soil moisture, soil moisture active passive (SMAP), United States Department of Agriculture Foreign Agricultural Service (USDA FAS).**

Manuscript received March 5, 2019; revised May 13, 2019; accepted June 4, 2019. Date of publication July 8, 2019; date of current version September 29, 2019. *(Corresponding author: Iliana E. Mladenova.)*

I. E. Mladenova is with the Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA, and also with Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740 USA (e-mail: iliana.e.mladenova@nasa.gov).

J. D. Bolten is with the Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA (e-mail: john.bolten@nasa.gov).

W. T. Crow and M. H. Cosh are with Hydrology and Remote Sensing Lab Agricultural Research Service, United States Department of Agriculture, Beltsville, MD 20705 USA (e-mail: wade.crow@ars.usda.gov; michael.cosh@ars.usda.gov).

N. Sazib is with the Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA, and also with Science Application International Corporation, Lanham, MD 20706 USA (e-mail: nazmus.s.sazib@nasa.gov).

C. J. Tucker is with the NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA (e-mail: compton.j.tucker@nasa.gov).

C. Reynolds is with the International Production Assessment Division Foreign Agricultural Service, United States Department of Agriculture, Beltsville, MD 20705 USA (e-mail: curt.reynolds@fas.usda.gov).

Digital Object Identifier 10.1109/JSTARS.2019.2923555

I. Introduction

THE US Department of Agriculture (USDA) Foreign Agricultural Service (FAS) is tasked with enhancing the international market competitiveness of United States agricultural exports. A major part of this effort is based on extensive global data analysis that provides USDA FAS with prompt information on the current state of agricultural commodity markets and allows USDA FAS to predict changes in international market conditions. Thus, the timely detection of environmental issues and variations in global weather or climate patterns that can affect crop production is essential. Agricultural drought is associated with a reduction in water supply and typically monitored by tracking changes in root-zone soil moisture (SM) (RZSM) conditions. As a result, SM, along with other essential agrometeorological parameters (e.g., precipitation, actual and potential evapotranspiration, daily temperature, and crop characteristics), is used by USDA FAS to properly quantify the impact of weather on crop growth and development.

Historically, USDA FAS has generated RZSM estimates using a 2-layer Palmer model (PM) soil water balance approach. The accuracy of SM estimates derived from water balance modeling is strongly dependent on the quality of precipitation forcing data applied to the model, which, in turn, is typically derived using multiple sources, including remotely sensed observations corrected by rain gauge observations [1], [2]. Densely instrumented regions are, therefore, associated with the highest accuracy precipitation forcing, while poorly instrumented areas are challenging due to the lack of ground-based observations for proper calibration of the large-scale precipitation products used to drive the hydrologic models. Additionally, global soil water balance models suffer deficiencies related to model parameter uncertainties, oversimplified model physics, and initialization errors.

Along with the application of soil water balance modeling, remote sensing advances made in the past two decades have enabled the timely monitoring of surface SM using satellite-based microwave sensors. These efforts culminated with the 2015 launch of the NASA SM active passive (SMAP) mission. However, remote sensing alone provides information about the SM conditions for only the top few centimeters of the soil profile, whereas agricultural applications require knowledge of the amount of water present in the entire root zone. In addition, these retrievals are prone to error in conditions of high biomass and lack continuity in both time and space.

Given the individual shortcomings of modeled and remotely sensed SM products, data assimilation strategies have been developed for the optimal integration of satellite SM retrievals into continuous water balance models. If properly constructed, these strategies yield optimal RZSM estimates with continuous coverage in both time and space.

The use of data assimilation for improved hydrologic modeling is not new. Numerous papers have already demonstrated the benefit of assimilating satellite-retrieved observations into physically-based land surface models [3]–[10]. Much of this past research utilized an ensemble Kalman filter (EnKF) approach. As discussed in [11], the EnKF is often preferred as a land data assimilation approach due to its flexibility in terms of parameterizing land surface modeling errors. Given the proven added value of data assimilation for land surface modeling, there has been significant recent interest in the development of operational applications that promote the use of the available satellite data for near-real-time decision making. This paper describes a fully developed and implemented satellite-model application approach that assimilates SMAP SM products into operational activities conducted within the USDA FAS Crop Forecasting System.

The application developed to enhance the USDA FAS drought monitoring and forecasting capabilities utilizes SM observations derived from SMAP and their assimilation into the USDA FAS PM using a one-dimensional (1-D) EnKF (hereinafter referred to as the "PM+SMAP" data assimilation system). This system represents one of the first truly operational uses of SMAP SM products. USDA FAS analysts are currently using PM-EnKF SM estimates to quantify the severity of agricultural drought and better anticipate the impact of such drought on global agricultural production. Therefore, our study objectives are to: 1) assess the value of SMAP for improved operational agricultural monitoring; 2) evaluate the performance of the operational SMAP enhanced PM system (PM+SMAP); and 3) demonstrate the benefit of assimilating satellite-based observations into the PM.

This paper is structured as follows. Section II provides background on the: 1) EnKF data assimilation approach; 2) PM; and 3) evaluation methodology we apply. PM-EnKF SMAP-based results are then presented and discussed in Section III and summarized in Section IV.

II. Data and Methodology

A. Data

1) SMAP Level 2 SM: NASA launched the SMAP mission in January 2015 and operational data collection started in late March 2015. The satellite observes the earth's SM and freeze/thaw states from a near-polar, sun-synchronous orbit twice a day at approximately 6 A.M. and 6 P.M. local solar time. The mission is designed to carry aboard two L-band microwave instruments. A radar (active centered at 1.26 GHz) and a radiometer (passive centered at 1.41 GHz) are both nested in a large 6-m spinning mesh antenna that measures the radio frequency energy "spot" over an area with a diameter of 40 km. This area is further reduced to 1 km using radar aperture synthesis. Unfortunately, the SMAP radar ceased operation in July 2015.

The current SMAP passive microwave (PMW) data archive spans from March 31, 2015 to the present. Based on these passive microwave observations, SMAP provides multiple SM products developed using different algorithms. Here, we utilized the baseline Level 2 (L2) SMAP SM product generated using the single channel algorithm (SCA) and SMAP V-pol brightness temperature observations [12]. This product is distributed at a 36-km × 36-km EASE-grid projection. For our purposes here, it was projected and resampled to match the model grid of the PM (regular 0.25° latitude/longitude grid) using conversion tools provided by the National Snow and Ice Data Center (NSIDC).

SMAP implemented a carefully designed calibration and validation (Cal/Val) plan (https://smap.jpl.nasa.gov/science/validation/) [12], [13]. The plan, developed in accordance with mission objectives and product requirements, included two stages—pre-launch and post-launch. The pre-launch stage focused on ensuring that there were reliable data resources available to validate, calibrate, and test the proposed SMAP algorithms and build the post-launch Cal/Val infrastructure and protocols. The Cal/Val post-launch objectives were focused on ground-based validation activities and efforts to ensure that the SMAP products meet the predefined mission requirements. Most importantly, in preparation for SMAP, a large number of *in situ* SM network partners were engaged to produce an extensively quality checked and comprehensive database for SMAP's Cal/Val activities. This database was further extended by leveraging additional ground data collected during a number of carefully planned field campaigns (https://smap.jpl.nasa.gov/science/validation/fieldcampaigns/). These data have been extensively assessed for their reliability and have been used (post-launch) to assess the accuracy of the SMAP SM products [12], [14]–[16]. Since the SMAP Cal/Val data are fully independent from the PM, this dataset is an objective resource for evaluating the quality of the EnKF SM predictions. Here, we utilized available SMAP Cal/Val data at seven core validation sites in the USA (described below in greater detail).

2) MODIS NDVI: NASA moderate resolution imaging spectroradiometer (MODIS)-based vegetation indices, such as the normalized difference vegetation index (NDVI), leaf area index (LAI), etc., have long demonstrated their potential for monitoring vegetation variability, phenology, and dynamics [17]. The NDVI dataset used in this paper has been developed by the Global Inventory Modeling and Mapping Studies Group at NASA Goddard (http://glam1.gsfc.nasa.gov/), funded through the Global Agricultural Monitoring Project by USDA's FAS and NASA Applied Sciences Program. The product provides global coverage at 250-m resolution and is derived using daily MODIS data, which is processed to 8-day composites. These data were aggregated up to 0.25° to match the model grid and binned to monthly composites. Following the approach outlined in [8], NDVI are used to evaluate global RZSM products. It should be noted that MODIS NDVI data were not used in the generation of any SM product considered here.

TABLE I
LIST OF THE SMAP SOIL MOISTURE CORE VALIDATION SITES USED FOR VALIDATION OF THE SMAP ASSIMILATION RESULTS

Watershed Site	Location	Climate	IGBP Land Cover
Walnut Gulch (WG)	Arizona	Semi-Arid	Shrub open
Little Washita (LW)	Oklahoma	Temperate	Grasslands /croplands
Fort Cobb (FC)	Oklahoma	Temperate	Croplands
Little River (LR)	Georgia	Temperate	Croplands/ natural mosaic
St. Josephs (StJ)	Indiana	Cold	Croplands
South Fork (SF)	Iowa	Cold	Croplands
Reynolds Creek (RC)	Idaho	Semi-Arid	Grasslands

All sites located in USA.

3) ARS Watershed Data: Currently, the SMAP SM validation network consists of partner sites spread worldwide that cover various land cover conditions and climate regimes [14]. Our ground-based validation was conducted over seven dense SM networks located within the USDA Agricultural Research Service (ARS) watershed sites and capturing predominantly cropland and grassland land covers (see Table I). These USA-based networks have been well calibrated and provide temporally continuous (and spatially dense) SM data at a 5-cm depth. All SM comparisons were made against watershed averages obtained and thus, reflect the mean SM within an area between approximately 150 and 820 km^2. These averages were generated by aggregating individual sites within each network via scaling functions designed to improve the spatial representativeness of the *in situ* stations with regard to a remote sensing footprint[14].

B. Methodology

1) PM and Data Assimilation: The modified PM used by USDA FAS is a simple 2-layer model driven by daily observations of precipitation and minimum and maximum air temperature [18]. The model is run at 0.25° grid spacing, with forcing data provided by the U.S. Air Force 557th Weather Wing (USAF557, formerly known as Air Force Weather Agency, AFWA). The model computes daily surface and subsurface SM estimates using a simple bucket-type water balance approach. Values are provided as depth of water (mm), where the maximum water holding capacity of the first layer is assumed to be 25.4 mm. The amount of water that can be stored in the second model layer is variable and is modeled as a function of specific soil properties found within each grid cell. The amount of soil water in the second model layer is increased only after the first layer reaches the maximum holding capacity. Once both layers are filled, the excess water is modeled as runoff.

Under dry-down conditions when the soil water in the top layer drops below two-thirds of the maximum available water, the model allows upward movement of water from the subsurface layer at a fraction of the potential evapotranspiration. This simple bucket-type model enables daily estimates of surface- and RZSM with water loss being driven primarily by evapotranspiration and runoff fluxes. No baseflow is assumed in the model. Modifications introduced by the USDA FAS include the implementation of the Penman–Monteith equation (which replaced the Thornthwaite equation) and the adjustment of the simple diffusion function that models the transfer of water between the two model layers. The shallow first layer of the model is ideal for the 1-D EnKF approach discussed below since it can be used as an appropriate proxy for the sensing depth of SMAP L2 SM retrievals.

The daily USAF557 precipitation estimates are generated by incorporating data from satellite, model, and World Meteorological Organization (WMO) ground stations. As discussed in Section I, PM SM estimates are heavily dependent on the quality of precipitation forcing applied to the model. This quality tends to degrade over areas of the world where WMO rain gauges are not available to correct for biases in satellite-based precipitation products. This shortcoming can be addressed through data assimilation, whereby satellite-derived SM is used to correct for random errors in the model SM estimates associated with low-quality precipitation forcing.

Here, we explore the value of SMAP for agricultural monitoring by assimilating SMAP L2 SM retrievals into the PM using a 1-D EnKF (Figure 1). The EnKF is a Monte Carlo-based, sequential data assimilation technique suitable for moderately nonlinear dynamic systems [19]. Its application to land data assimilation has been well documented [3]–[10]. Therefore, a detailed description of the EnKF technique is not included here.

The PM is set to run at a regular 0.25° latitude/longitude grid defined by the resolution of the model's forcing datasets. Thus, the SMAP L2 product was preprocessed prior to assimilation to match the model grid by taking the following steps. First, the 36-km EASE-Grid was disaggregated to the 1-km EASE-Grid. Next, the resulting 1-km fields were up-scaled up to 25 km by averaging all 1-km points that fall within each 25-km grid box.

2) Filter Parameterization: As discussed in [11], an effective way to evaluate EnKF performance is examining the time series of the filtering innovations, defined as the difference between the assimilated observations and the background realized during the analysis cycle. A properly constructed filter should yield an innovation time series that is serially uncorrelated, stationary with a normalized variance of one and mean zero. In an EnKF framework, the model forecasts are updated in response to satellite observations using the so-called Kalman gain (K). K is a function of the error covariance of the satellite observations, their error estimate (R), and the cross correlation between the observations and each forecasted state variable. The forecast realization is defined by the model uncertainty matrix (Q). Generally, Q reflects the model's accuracy, including physics and parameterization, and our trust in the forcing data, while R represents the accuracy of the satellite-derived SM observations, our confidence in the performance of the retrieval algorithm and its parameterization, and the sensor's sensitivity and calibration accuracy. Thus, tuning of the filter to satisfy the above-specified constraints depends on the proper parameterization of Q and R, which is a nontrivial task given the nonlinear dynamics of the system.

As discussed above, soil water balance models are highly sensitive to the quality of the precipitation data applied to force

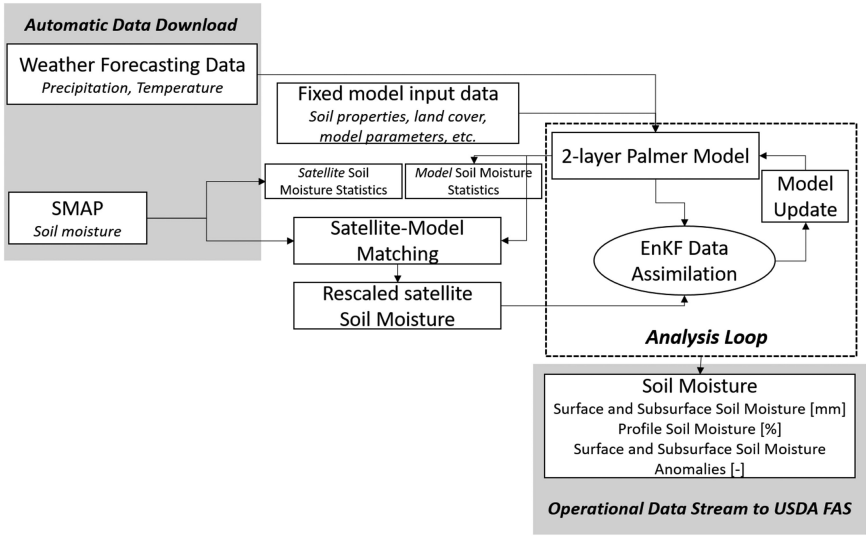


Fig. 1. Schematic representation of the data flow within the EnKF data assimilation scheme.

them. Given that USAF557 is a merged product that includes rain gauge data, considered closest to reality, Q was modeled as a function of gauge density and distance to the nearest rain gauge station, where higher confidence is given to the model estimates over areas with denser gauge networks such as the USA and Europe. Specifically, Q is set to gradually increase as the distance between the model grid point and the nearest station increases.

There are a few different approaches that can be adopted to model the observation error R. In any of the available approaches, the underlying logic is based on the well known dependence of the quality of the satellite retrievals on the density of the vegetation canopy layer. Over densely vegetated areas, the microwave signal cannot fully penetrate through the vegetation layer. The canopy masks or dampens the signal coming from the water present in the soil layer leading to less-reliable SM retrievals. Passive-based microwave retrieval approaches take this canopy dependence into consideration and correct through the optical depth (τ) parameter. τ can be either modeled using an NDVI-based climatology (as is the case in the SMAP baseline single channel algorithm [21]) or estimated simultaneously with SM through consideration of observations acquired at different microwave frequencies or polarizations. For most vegetation types, canopy density and the amount of water present in the leaves and stems are not static in time; therefore, it is expected that the accuracy of the satellite retrievals will change not only as a function of land cover type but also temporally throughout the year. Thus, R can be parameterized as a function of land cover type and/or as a function of vegetation characteristics as described by the NDVI climatology or remotely observed τ.

When modeled as a function of land cover, R varies spatially but it is fixed in time, while modeling R as a function of NDVI or τ introduces seasonal variability in the observation error, which can be either climatology-based or provided in near real time. Alternatively, R can be defined using actual error estimates obtained through statistical error analysis. The main difference between these approaches is the presence and detail of temporal information in R.

Detailed comparison of these approaches and evaluation of their impact on the data assimilation results are beyond the scope of our paper. Nevertheless, preliminary analyses indicate that adding a seasonal component to R does not substantially impact the accuracy of EnKF SM estimates [22]. Therefore, the observation error was modeled by considering only spatial variations in land cover and neglecting inter- and intra-annual temporal variability (for the specific values see SMAP Cal/Val report: https://nsidc.org/sites/nsidc.org/files/technical-references/L2SMPE_Asmt_Rpt_EOPM_v5c_Jun2018.pdf). Specific error variance values assigned to each land cover class represent the square of unbiased root-mean-square-error (*ubRMSE*) values computed against actual *in situ* data. These values were further rescaled to transform observation errors into the climatology of the model as discussed in [23].

3) Lagged Comparisons With NDVI: Water stress causes decreased leaf water potential and reduction in the stomatal openings, which, in turn disrupts the plant's photosynthetic capability [26]. However, plants have a number of physiological and biochemical coping and adaptation mechanisms to adjust and adapt to a variety of environmental stresses such as light deficiency, extreme heat, abnormally low temperatures, diminished or insufficient water supplies, etc. [27], [28]. Thus, there is typically some delay in the plant's reaction and response to water stress. This delay suggests that, in water-limiting situations, knowledge of a change in SM conditions can be used to predict future changes in local vegetation status and predict the extent areas experiencing agricultural drought.

Here, the reliability of the global RZSM products is assessed by examining the relationship between water availability and plant status using a time-lagged cross-correlation analysis between NDVI and the modeled RZSM estimate [9], [29]. Negative anomalies in RZSM should foretell near-future negative anomalies in NDVI. Therefore, following [8], our assumption is that lagged cross correlation between RZSM and NDVI will

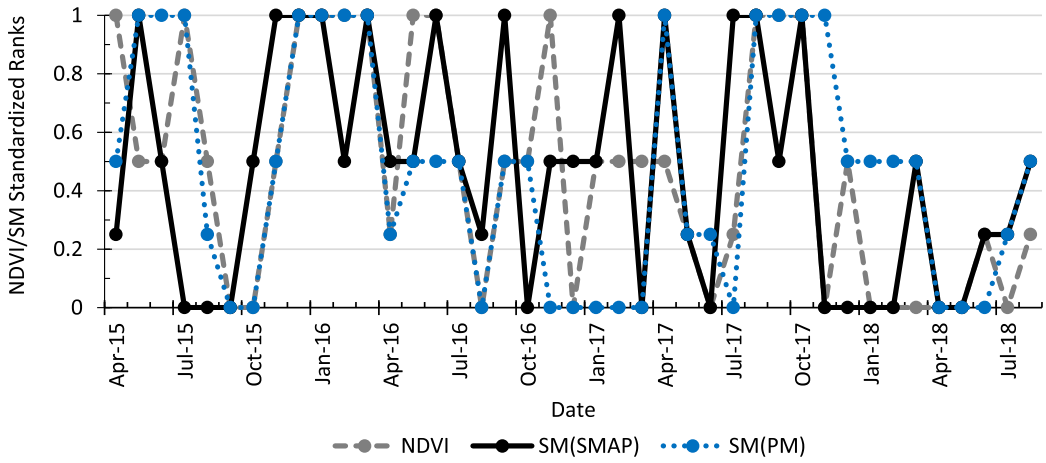


Fig. 2. Time series of standardized ranked NDVI and soil moisture observations over the Little Washita watershed area (Lat. 34.93°, Long. 98.17°). Grey line shows the temporal change in vegetation conditions as captured by NDVI, while the soil moisture response from the model (PM) alone and SMAP are depicted by the black and blue lines, respectively.

increase as the precision of the SM estimates is improved. Based on this premise, we used the SM and NDVI cross correlation as a metric for the accuracy of PM and PM+SMAP RZSM estimates.

This evaluation analysis was conducted using monthly SM and NDVI composites at 0.25°. Correlation coefficient values were computed using ranks, where ranking was performed separately for each month of the year. That is, all monthly composites from all years (April 2015–October 2018) were grouped according to their particular month of the year and ranked relative to each other. The corresponding ranks were then normalized so that the final range for all months was the same. Each rank for a specific month represents how dry (rank $= 0$) / wet (rank $= 1$) a particular month (e.g., June 2016) is relative to the *same month of the year* in all the other years (e.g., June 2015, June 2017, and June 2018). In this case, the NDVI-SM agreement is not impacted by seasonality effects. In addition, as clarified above, the NDVI-validation data and the climatological NDVI data used to compute the τ values are independent. Therefore, the month-based ranking approach adopted here ensures no influence of any residual effects related to the use of climatological NDVI information within the SMAP SCA to parameterize τ.

This procedure was applied to both the modeled SM and the NDVI time series. Correlation values between the ranked NDVI and SM time series were then computed as

$$R(L) = R\left(\mathrm{Rank}_{\mathrm{NDVI}_m}, \mathrm{Rank}_{\mathrm{SM}m+L}\right) \tag{1}$$

where $R(L)$ is lagged (L) rank correlation, $\mathrm{Rank}_{\mathrm{NDVI}}$ and $\mathrm{Rank}_{\mathrm{SM}}$ are the ranked NDVI or SM time series for month m. Therefore, $L = -1$ [month] indicates that the SM rank precedes the NDVI rank by one month. As explained in [10], a lag of one month generally agrees well with the time scales important for operational agricultural activities. Global $R(L)$ values for each model grid point were computed including only months that fall within the growing season (i.e., April–October over the Northern Hemisphere and October–April over the Southern Hemisphere). An example of ranked SM and NDVI time series illustrating the validation approach is shown in Fig. 2.

4) In Situ-Based Analysis: The global SM-NDVI lag correlation evaluation described above was supplemented with additional *in situ*-based accuracy assessments. In particular, the surface estimates from the PM before and after assimilating SMAP was compared against the *in situ* SM data collected at the seven Cal/Val ARS sites described in Section II. The analysis was based on sampling the temporal correlation (R) between *in situ* and SMAP-based surface SM estimates and computing the unbiased root-mean-square error $(ub\mathrm{RMSE})$ between observed and modeled SM

$$R = R\left(A_{SM[\mathrm{In\,Situ}]}, A_{\mathrm{SM}[\mathrm{PM}^*]}\right) \tag{2}$$

$$ub\mathrm{RMSE} = \sqrt{\sigma^2_{\mathrm{PM}^*} + \sigma^2_{\mathrm{In\ Situ}} - 2R\sigma_{\mathrm{PM}^*}\sigma_{\mathrm{In\,Situ}}} \tag{3}$$

where A represents SM time series of standardized anomalies, PM^* is the output from the PM alone or the assimilation run, $\mathrm{PM} - \mathrm{EnKF\,SMAP}$, σ^2 is the time variance of the *in situ* and modeled SM [30]. Since the PM is constrained using a different soil property dataset from the one used for the calibration of the individual stations within the watersheds (i.e., global, coarse resolution soil data versus *in situ* collected and analyzed soil texture samples), extreme outliers (i.e., values lower/higher than the 2.5/97.5 percentile) were excluded.

III. Results and Discussion

A. Filter Whitening and Evaluation

Our EnKF implementation was static in the sense that it utilizes fixed, predefined *Q* and *R* values and does not interactively adjust *Q* and *R* within the analysis cycle to optimize innovation statistics. Nevertheless, as described below, reasonable filtering performance was achieved using *a priori* knowledge to define both parameters.

The final model (*Q*) and observation error (*R*) maps are shown in Fig. 3. By design, the spatial variability captured in the model error map (see Fig. 3, top plot) closely resembles the WMO

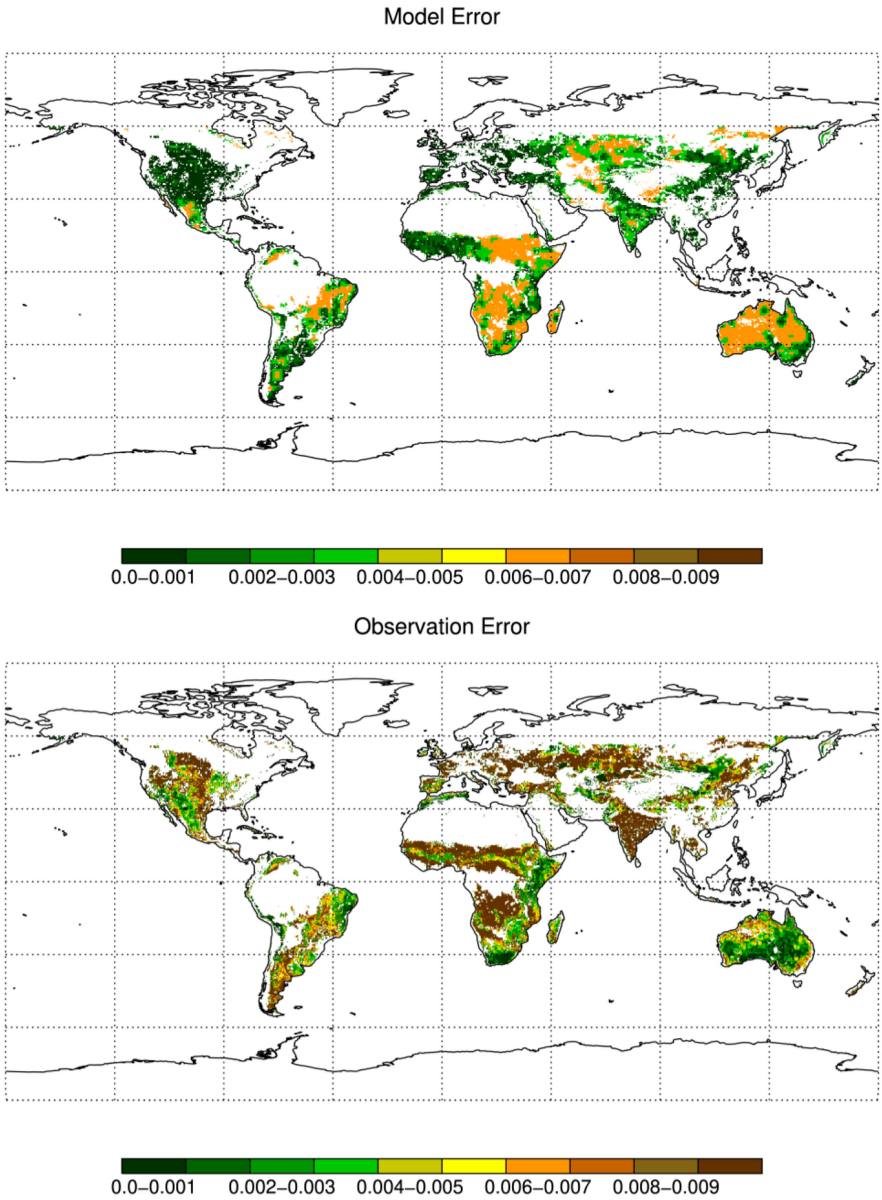
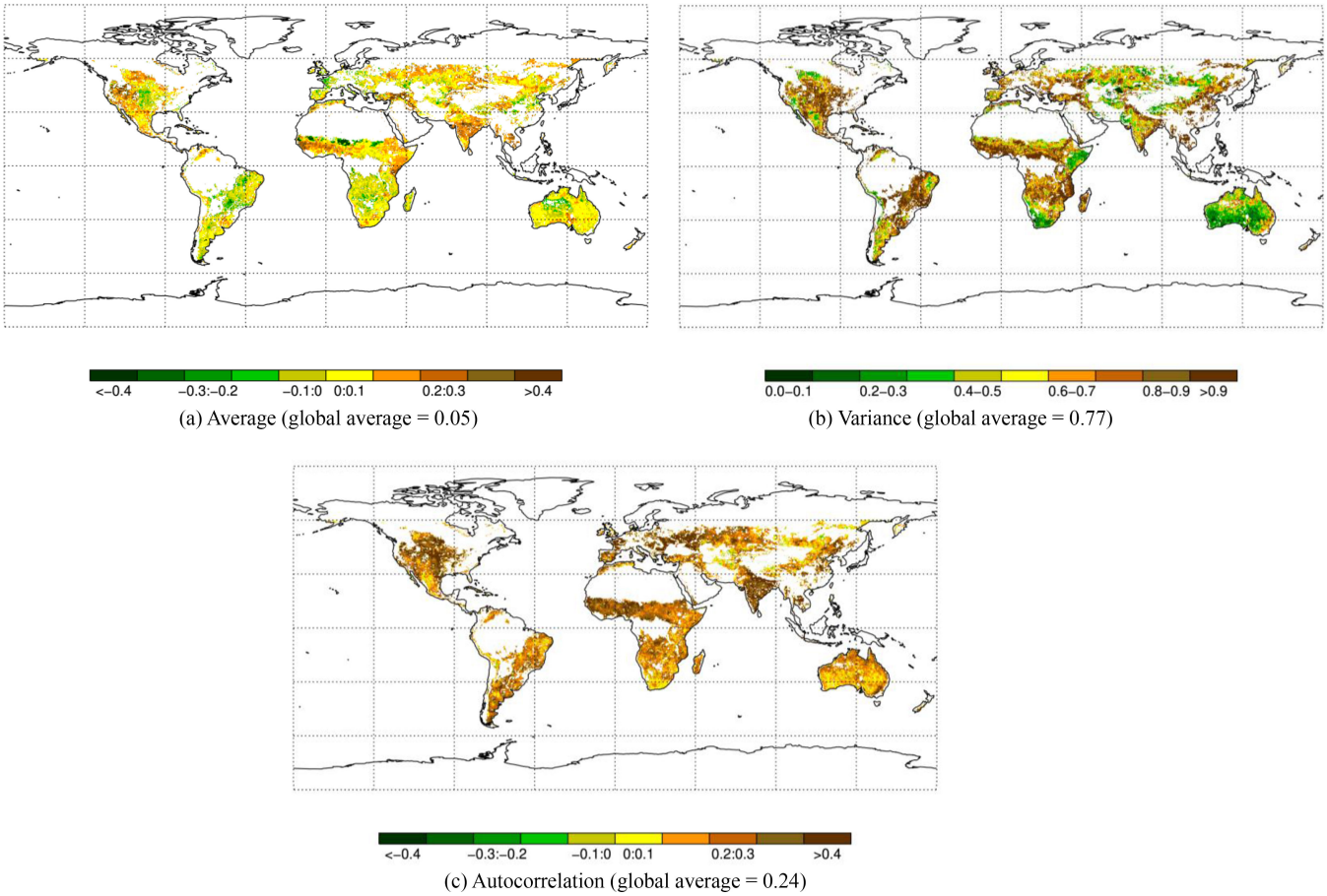


Fig. 3. Spatial variability of assumed variances [m6/m6] in model forecasts (top plot) and the observations (bottom plot) error, Q and R, respectively. Water and energy limited areas, such as the Amazon and the Sahara regions were masked out in addition to applying the recommended SMAP flags for unreliable retrievals and retrievals over unfavorable ground conditions (i.e., frozen soils, dense vegetation, and mountainous terrain). These areas are masked out in all plots included in the manuscript.

gauge density map, where higher confidence is given to the model over areas with denser gauge coverage. The green end of the bar indicates higher confidence in the precipitation data, and hence, the SM estimates derived from the PM model estimates (due to the higher density of available rain gauge stations to correction satellite-based precipitation estimates), while the orange/brown colors are associated with lower gauge density and higher modeling uncertainty. The specific values of R (see Fig. 3, bottom plot) were determined using error values derived from published SMAP Cal/Val results (see the SMAP L2 SM Cal/Val report: https://nsidc.org/sites/nsidc.org/files/technical-references/L2SMPE_Asmt_Rpt_EOPM_v5c_Jun2018.pdf). The green end of the bar indicates higher confidence in the observations. The observation error map indicates expected spatial variability displaying higher confidence in the observations over barren/sparsely vegetated and shrubland areas. Observation errors increase over croplands (i.e., China, parts of Europe), which is expected given the high biomass, and therefore, sharp attention of soil signals, within summertime crop cover.

The corresponding innovation statistics are shown in Fig. 4. The overall global performance of the standardized innovations is within its expected limits. As described above, an auto-correlation of zero and a normalized innovation variance of one are expected for an idealized filter. However, such results are based on an assumption of purely white observation errors. In the presence of auto-correlated observing errors, optimal filter performance is instead associated with auto-correlation values that are slightly positive and normalized innovation variances



(a) Average (global average = 0.05)

(b) Variance (global average = 0.77)

(c) Autocorrelation (global average = 0.24)

Fig. 4. Evaluation of the average, m3/m3 (a), normalized variance (b), and unit-less autocorrelation (c) of the standardized innovations for the USDA FAS EnKF system.

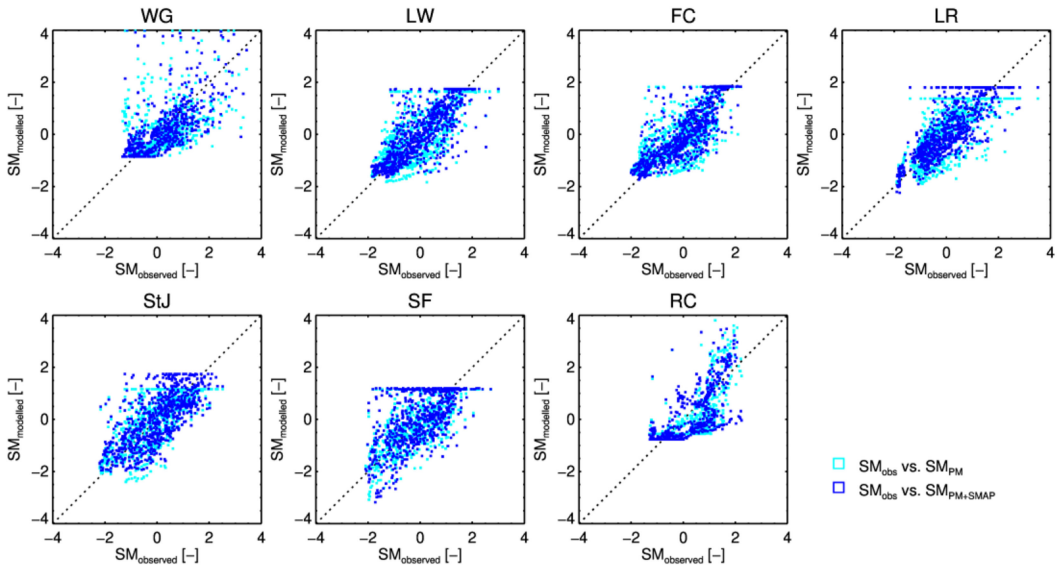


Fig. 5. Scatter plots of the standardized watershed soil moisture anomaly values against the modeled output plotted on the *x*- and *y*-axis, respectively. Each plot shows a separate watershed area (see Table I for the abbreviations of the watershed names) and includes two clusters; the cyan colored squares show the agreement between the watershed data and the model alone, and the blue colored squares show the same agreement but after assimilating SMAP.

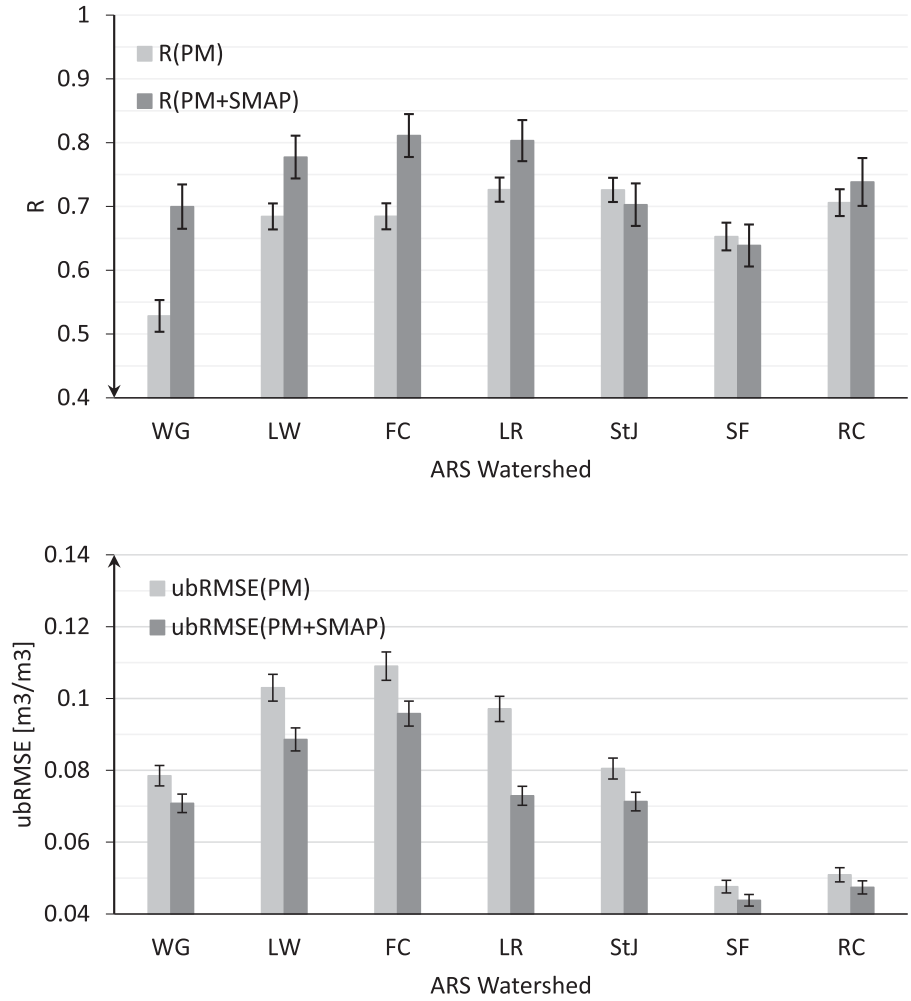


Fig. 6. Summary plot of the *in situ* model soil moisture anomaly correlation results. Light grey and dark grey bars show the correlation values before assimilation and after assimilating SMAP, respectively. Error bars represent 95% error associated with the time sampling uncertainty at all seven ARS watershed sites (see Table I for the abbreviations of the watershed names). Note that SMAP assimilation consistently improved the modeled (PM) soil moisture information.

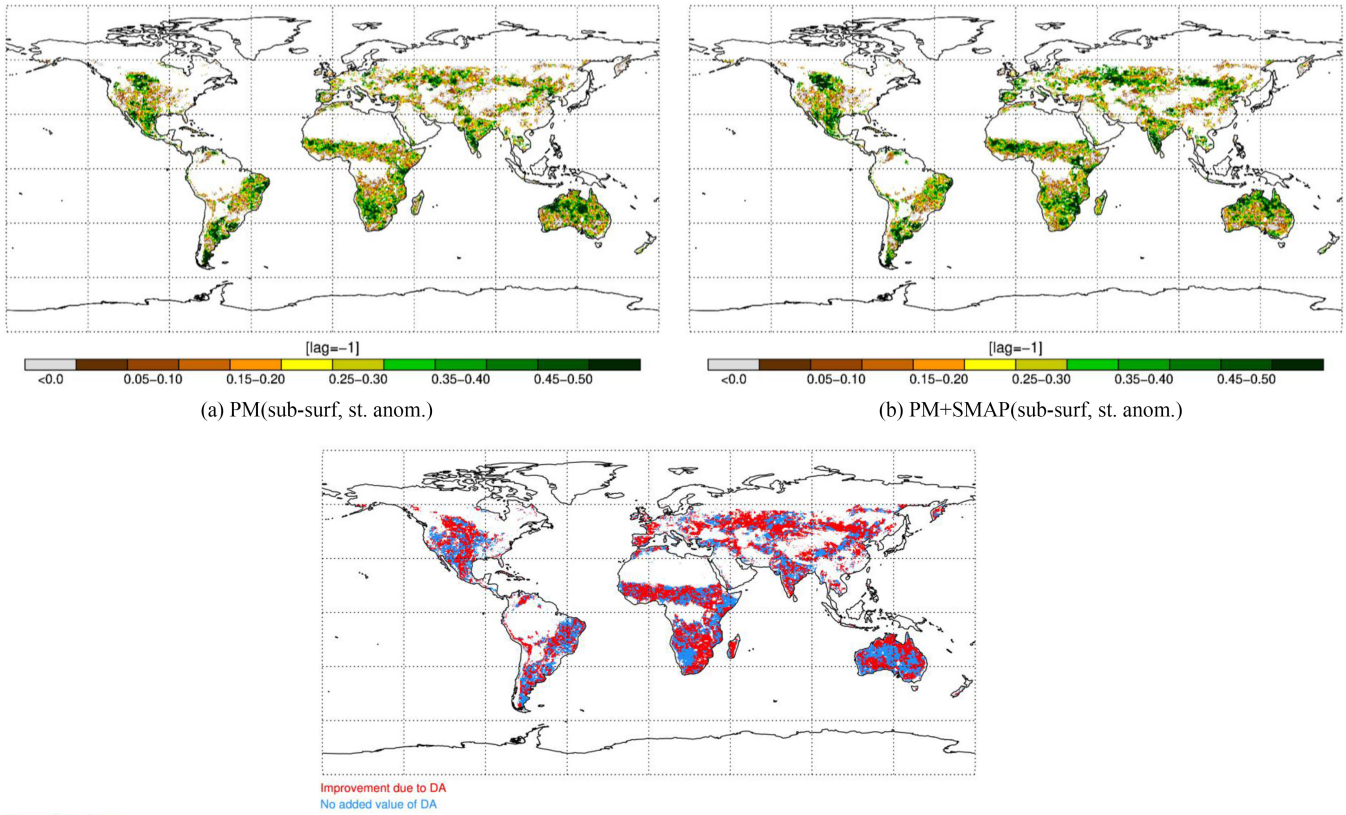
that are slightly less than one [11]. Therefore, the innovation results in Fig. 4 suggest a generally well parameterized filter, given the known tendency for SMAP SCA retrievals to have temporally auto-correlated errors [29]. Nevertheless, certain areas may require further filter tuning. Furthermore, one should consider the rather short SMAP data record and, as a result, the presence of nonsignificant levels of sampling error present in innovation statistics presented in Fig. 4. Given these considerations, Fig. 4 suggests that our EnKF system is functioning reasonably well [11], [29].

B. PM-EnKF Accuracy Evaluation

1) Point Analysis: If properly calibrated and screened, ground-based SM data can be considered a "true" reference when evaluating global satellite-based products due to their relative reliability versus large-scale SM estimates. Therefore, despite the sporadic spatial coverage of the available networks, ground-based validation is invaluable when assessing any global product.

Summary results from the ground-based analysis computed using SM observations from the USDA ARS watersheds are presented in Figs. 5 and 6. The scatter plots displayed in Fig. 5 show the agreement between the standardized anomaly SM time series of the watershed data (x-axis) and model (y-axis) output (obtained both before and after the assimilation of SMAP data). Note that watershed data points are based on the spatial averaging of multiple, ground-based sites into a single watershed-averaged value. Evaluation results obtained before (PM) and after (PM+SMAP) assimilating SMAP are plotted in cyan and blue color, respectively. All observations are well aligned along the one-to-one line indicating good agreement between the two datasets. The scatter plots of some of the watersheds such as Little Washita, for example, show tightening of the blue cluster (corresponding to PM+SMAP results) along the one-to-one line indicating that SMAP enhances the SM provided by the PM.

This tendency is even clearer in Fig. 6, where the initial PM correlation values (no assimilation) are shown in grey, while the PM+SMAP values achieved after assimilating SMAP are shown in black. The initial R values at all watershed sites are



(a) PM(sub-surf, st. anom.)

(b) PM+SMAP(sub-surf, st. anom.)

(c) PM+SMAP vs. PM(sub-surf.)

Fig. 7. Top row shows global analysis of (unitless) lag correlations between the ranked monthly NDVI and soil moisture (subsurface layer). The left plot (a) displays the results obtained from the open loop (PM, no assimilation) at $L = -1$ [month], while the right plot (b) shows the correlations after assimilating SMAP into the PM (PM+SMAP) at $L = -1$ [month]. The image on the bottom row (c) displays the difference between the assimilation run (PM+SMAP) and the model alone (PM) at $L = -1$ [month]. Red shading indicates positive impact of the SMAP assimilation into the subsurface layer of the PM layer.

greater than 0.6, except for Walnut Gulch—indicating the relatively good baseline performance of the PM. Nevertheless, these baseline values are improved consistently at all locations after assimilating SMAP. Error bars plotted in Fig. 6 (representing the uncertainty encountered with the time sampling) demonstrate that the achieved results have rather low uncertainty at all sites. Agreement is the lowest at the Walnut Gulch site, located in Arizona—the only site covered by shrubland vegetation. Over grass- and crop-land, the magnitude of the correlation values appear to be similar and independent of vegetation type ($\bar{R}^{\mathrm{PM}}_{\mathrm{grass/crop}} = 0.70/0.73$; $\bar{R}^{\mathrm{PM+SMAP}}_{\mathrm{grass/crop}} = 0.77/0.75$, where values represent the average metric sampled across all watershed sites characterized by either grass or crop land cover—see Table I).

It should be noted that improvements at the USA-based USDA ARS watershed sites are likely to be less than global averages since the background PM is relatively accurate at these sites due to the excellent availability of real-time rain gauge data at sites within the USA. Larger relative improvements are, therefore, expected in data-poor areas where the background PM model contains much higher levels of random error. Nevertheless, this ground-based verification demonstrates the benefit of SMAP and its potential to improve the accuracy of the USDA FAS SM information.

2) Global Analysis: In order to geographically expand the scope of our evaluations, we also applied the SM/NDVI correlation analysis described earlier in Section II. Global maps of the resulting lag correlation coefficients between SM and EnKF are shown in Fig. 7. The cross-correlation maps show the NDVI-SM (subsurface) agreement for the PM alone in the top plot and the corresponding agreement after assimilating SMAP in the bottom plot. The PM+SMAP product shows higher agreement with NDVI as compared to the PM (no assimilation) product, which is the default USDA FAS SM product. Therefore, the assimilation of SMAP into the PM appears to improve the ability of the resulting SM product to anticipate NDVI anomalies (see Fig. 8). As discussed earlier in the paper, it is expected that the optimal NDVI-SM agreement would be achieved around $L = -1$ (month) [8]–[10]. Overall, globally our analysis demonstrates similar performance for PM+SMAP data at $L = -1$ and $L = 0$ (month). This tendency was also reported by [8]. Bolten and Crow [8] linked the performance of the assimilated results to the quality of the precipitation data. Unfortunately, the data record offered currently by SMAP currently only covers three full growing seasons, which is a highly marginal length of data from an agricultural forecasting perspective.

Nevertheless, the PM+SMAP case (see Fig. 7, top row right image) shows improvement over most of the globe relative to the

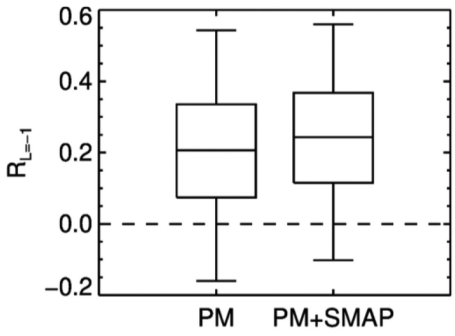


Fig. 8. Global average and overall variability in SM-NDVI lag correlation at $L = -1$ [month]. Box plots were generated by excluding correlations that fall below and above the 5th and the 95th percentile. The assimilation of SMAP into the USDA FAS PM (right box plot) increased the global agreement between the modeled SM of the subsurface layer and NDVI relative to the base run (left box plot).

results obtained from the case of the PM alone (see Fig. 7, top row left image). The bottom image in Fig. 7 offers a more detailed picture of the areas impacted by the integration of SMAP. The assimilation of SMAP improves the RZSM of the PM over large portions of Africa, shrub- and grass-land covered areas in North America and Asia and parts of Australia.

IV. Summary

The work presented here describes the first implementation of SMAP SM products into an operational crop forecasting system. Accurate and on-time global RZSM anomaly information is essential for USDA FAS as it aids their decision-making capabilities related to short- and long-term agricultural drought impacts on expected yield production and global food security. SMAP data have been assimilated into the USDA FAS PM, where the integration of the satellite data was specifically designed to improve the accuracy of the USDA FAS RZSM information. The system was assessed using two different and complementary approaches: 1) a ground-based validation analysis carried out using *in situ* data acquired at seven of the core SMAP Cal/Val sites located in the USA and 2) a global evaluation performed by assessing the lagged agreement between the modeled SM data (acquired both before and after assimilating) SMAP against NDVI, where NDVI is used as a proxy of vegetation health. Both analyses confirm the underlying utility of assimilating SMAP SM products into the PM and the benefit of our data assimilation approach to USDA FAS.

SMAP also provides a Level 4 RZSM product. However, the approach and product described here have been specifically designed to meet USDA FAS needs (e.g., specific spatial resolution and anomaly products) and requirements (short latency and matched climatology to preexisting USDA FAS PM climatology). It should be emphasized that the USDA FAS data assimilation system is run operationally and provides near-real-time SM data to USDA FAS. The data products generated from this system have been fully adopted by USDA FAS and are an integral part of the agency's CADRE BDMS (https://ipad.fas.usda.gov/cropexplorer/).

Having demonstrated the added benefit that SMAP-based observations provide to the operational USDA FAS crop forecasting system, it is envisaged that a next logical step is to leverage, in a more strategic manner, the near-real-time satellite-based monitoring of both SM and vegetation for improved global assessment and forecasting of crop conditions and improved estimation of crop yield. A topic that is yet to be explored in detail is the synergistic merging of SM and NDVI within a crop systems model at strategic times of the growing cycle to isolate signals most useful for indicating future crop yield. The added benefit of multivariable monitoring of end-of-season crop yield based on satellite-based evapotranspiration, SM, vegetation estimation, etc. was recently demonstrated in [30]. To this end, the results presented here provide a robust start to help pinpoint locations and times where SMAP-based SM estimates can help provide critical information on global agricultural drought and help with this more targeted holistic approach to global agricultural monitoring and forecasting.

References

[1] R. Reichle and R. D. Koster, "Bias reduction in short records of satellite soil moisture," *Geophys. Res. Lett.*, vol. 31, 2004, Art. no. L19501.
[2] R. Reichle and R. D. Koster, "Global assimilation of satellite surface soil moisture retrievals into the NASA catchment land surface model," *Geophys. Res. Lett.*, vol. 32, pp. 2353–2364, 2005.
[3] R. H. Reichle, "Data assimilation methods in the earth sciences," *Adv. Water Resour.*, vol. 31, no. 11, pp. 1411–1418, 2008.
[4] R. D. Koster, Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, "On the nature of soil moisture in land surface models," *J. Climate*, vol. 22, no. 16, pp. 4322–4335, Aug. 2009.
[5] M. Drusch, "Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set," *J. Geophys. Res.*, vol. 112, no. D3, 2007, Art. no. D03102.
[6] W. Crow and D. Ryu, "A new data assimilation approach for improving hydrologic prediction using remotely-sensed soil moisture retrievals," *Hydrol. Earth Syst. Sci.*, vol. 13, pp. 1–16, 2009.
[7] A. J. W. de Wit and C. A. van Diepen, "Crop model data assimilation with the ensemble Kalman filter for improving regional crop yield forecasts," *Agricultural Forest Meteorol.*, vol. 146, no. 1/2, pp. 38–56, Sep. 2007.
[8] J. D. Bolten and W. T. Crow, "Improved prediction of quasi-global vegetation conditions using remotely-sensed surface soil moisture," *Geophys. Res. Lett.*, vol. 39, no. 19, Oct. 2012, Art. no. L19406.
[9] W. T. Crow, S. V. Kumar, and J. D. Bolten, "On the utility of land surface models for agricultural drought monitoring," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 9, pp. 3451–3460, Sep. 2012.
[10] E. Han, W. T. Crow, T. Holmes, and J. Bolten, "Benchmarking a soil moisture data assimilation system for agricultural drought monitoring," *J. Hydrometeorol.*, vol. 15, no. 3, pp. 1117–1134, Feb. 2014.
[11] W. T. Crow and M. J. Van den Berg, "An improved approach for estimating observation and model error parameters in soil moisture data assimilation," *Water Resour. Res.*, vol. 46, 2010, Art. no. W12519.
[12] S. Chan, "Soil moisture active passive (SMAP) mission enhanced level 2 passive soil moisture product specification document," Jet Propulsion Lab., Pasadena, CA, USA, Rep. JPL D-56291, Nov. 2016.
[13] S. K. Chan *et al.*, "Development and assessment of the SMAP enhanced passive soil moisture product," *Remote Sens. Environ.*, vol. 204, pp. 931–941, Jan. 2018.
[14] A. Colliander *et al.*, "Validation of SMAP surface soil moisture products with core validation sites," *Remote Sens. Environ.*, vol. 191, pp. 215–231, Mar. 2017.
[15] M. S. Burgin *et al.*, "A comparative study of the SMAP passive soil moisture product with existing satellite-based soil moisture products," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2959–2971, May 2017.
[16] Cai Xitian *et al.*, "Validation of SMAP soil moisture for the SMAPVEX15 field campaign using a hyper-resolution model," *Water Resour. Res.*, vol. 53, no. 4, pp. 3013–3028, Apr. 2017.

[17] N. Pettorelli, J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth, "Using the satellite-derived NDVI to assess ecological responses to environmental change," *Trends Ecol. Evol.*, vol. 20, no. 9, pp. 503–510, Sep. 2005.
[18] W. C. Palmer, "Meteorological drought," U.S. Weather Bureau, Washington, DC, USA, Res. Paper 45, 1965.
[19] G. Evensen, "The ensemble Kalman Filter: Theoretical formulation and practical implementation," *Ocean Dyn.*, vol. 53, pp. 343–367, 2003.
[20] T. J. Jackson, "Measuring surface soil moisture using passive microwave remote sensing," *Hydrol. Process.*, vol. 7, pp. 139–152, 1993.
[21] R. A. M. de Jeu and M. Owe, "Further validation of a new methodology for surface moisture and vegetation optical depth retrieval," *Int. J. Remote Sens.*, vol. 24, no. 22, pp. 4559–4578, 2003.
[22] I. E. Mladenova, J. D. Bolten, W. Crow, and R. de Jeu, "Evaluating the application of microwave-based vegetation observations in an operational soil moisture data assimilation system," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 5190–5193.
[23] M. T. Yilmaz and W. T. Crow, "The optimality of potential rescaling approaches in land data assimilation," *J. Hydrometeorol.*, vol. 14, no. 2, pp. 650–660, Oct. 2012.
[24] N. C. Turner and P. J. Kramer, *Adaptation of Plants to Water and High Temperature Stress*. Hoboken, NJ, USA: Wiley, 1980.
[25] M. M. Chaves *et al.*, "How plants cope with water stress in the field? Photosynthesis and growth," *Ann. Botany*, vol. 89, no. 7, pp. 907–916, Jun. 2002.
[26] Y. Osakabe, K. Osakabe, K. Shinozaki, and L.-S. P. Tran, "Response of plants to water stress," *Frontiers Plant Sci.*, vol. 5, Mar. 2014, Art. no. 86.
[27] E. Peled, E. Dutra, P. Viterbo, and A. Angert, "Technical note: Comparing and ranking soil drought indices performance over Europe, through remote-sensing of vegetation," *Hydrol. Earth Syst. Sci.*, vol. 14, no. 2, pp. 271–277, Feb. 2010.
[28] D. Entekhabi, R. H. Reichle, R. D. Koster, and W. T. Crow, "Performance metrics for soil moisture retrievals and application requirements," *J. Hydrometeorol.*, vol. 11, no. 3, pp. 832–840, Jun. 2010.
[29] J. Dong, W. T. Crow, and R. Bindlish, "The error structure of the SMAP single and dual channel soil moisture retrievals," *Geophys. Res. Lett.*, vol. 45, no. 2, pp. 758–765, Jan. 2018.
[30] I. E. Mladenova *et al.*, "Intercomparison of soil moisture, evaporative stress, and vegetation indices for estimating corn and soybean yields over the U.S.," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1328–1343, Apr. 2017.

Iliana E. Mladenova received the M.S. degree in hydrology from Free University, Amsterdam, The Netherlands, and the Ph.D. degree in geological sciences from the University of South Carolina, Columbia, SC, USA, in 2006 and 2009, respectively.

Her major expertise are in the areas of land surface hydrology and remote sensing. Her research interests include the use of satellite-based microwave techniques, land surface modeling, and data assimilation for improved monitoring of the different components of the water cycle, and their potential integration for enhanced earth system modeling.

John D. Bolten (M'05) received the M.S. and Ph.D. degrees in geological science with an emphasis in remote sensing from the University of South Carolina, Columbia, SC, USA, in 2001 and 2005, respectively.

He is currently a Physical Scientist with the Hydrological Sciences Lab, NASA Goddard Space Flight Center, Greenbelt, MD, USA. His primary research interests involve the application of satellite-based remote sensing and land surface hydrological modeling for improved ecological and water resource management.

Wade T. Crow (M'03) received the Ph.D. degree from Princeton University, Princeton, NJ, in 2001.

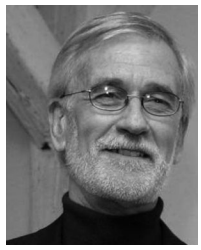
He is currently a Research Physical Scientist with the Hydrology and Remote Sensing Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD, USA. His research interests include the development of land data assimilation techniques to enhance the utility of remote sensing observations for hydrologic and agricultural applications.

Nazmus Sazib received the M.S and Ph.D. degrees in civil engineering from the University of Louisiana and Utah State University, Logan, UT, USA, in 2012 and 2016, respectively.

He is a Research Scientist with the Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD, USA. His research interest includes remote sensing, data assimilation, and the development of software and tools to advance the utility of remote sensing data for hydrologic application.

Micheal H. Cosh (M'02–SM'17) received the Ph.D. degree in civil and environmental engineering from Cornell University, Ithaca, NY, USA, in 2002.

He is a Research Hydrologist with the U.S. Department of Agriculture, Agricultural Research Service, Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA. His current research interests include the monitoring of soil moisture from both *in situ* resources and satellite products.



Compton J. Tucker recevied the B.S. degree in biology, and the M.S. and Ph.D. degrees in forestry from Colorado State University, Fort Collins, CO, USA, in 1969, 1973, and 1975, respectively.

Since 1975, he has been with NASA. His research interests include the use of satellite data to study deforestation, desertification, land degradation, ecologically coupled deceases, and terrestrial primary production.

Curt Reynolds received the B.S. degree in civil and environmental engineering from the University of Wisconsin-Madison, WI, USA, in 1983, and the M.S. degree in civil engineering/water resources management and the Ph.D. degree in agricultural and biosystems engineering from the University of Arizona, Tucson, AZ, USA, in 1993 and 1998, respectively.

He is a Geospatial Crop Analyst with the U.S. Department of Agriculture's Foreign Agricultural Service (FAS), where he has worked for the past 20 years. He has developed several operational systems for global crop monitoring, such as the FAS Crop Explorer and Commodity Explorer websites; the USDA/FAS and NASA Global Agriculture Monitoring (GLAM) and the Global Reservoir and Lake Monitor (G-REALM) web systems. His research interests include spatial hydrology and crop yield modeling.