

An objective methodology for merging satellite- and model-based soil moisture products

M. T. Yilmaz,¹ W. T. Crow,¹ M. C. Anderson,¹ and C. Hain²

Received 23 November 2011; revised 13 September 2012; accepted 18 September 2012; published 2 November 2012.

[1] An objective methodology that does not require any user-defined parameter assumptions is introduced to obtain an improved soil moisture product along with associated uncertainty estimates. This new product is obtained by merging model-, thermal infrared remote sensing-, and microwave remote sensing-based soil moisture estimates in a least squares framework where uncertainty estimates for each product are obtained using triple collocation. The merged anomaly product is validated against in situ based soil moisture data and shows higher correlations with observations than individual input products; however, it is not superior to a naively merged product acquired by averaging the products using equal weighting. The resulting combined soil moisture estimate is an improvement over currently available soil moisture products due to its reduced uncertainty and can be used as a standalone soil moisture product with available uncertainty estimates.

Citation: Yilmaz, M. T., W. T. Crow, M. C. Anderson, and C. Hain (2012), An objective methodology for merging satellite- and model-based soil moisture products, *Water Resour. Res.*, 48, W11502, doi:10.1029/2011WR011682.

1. Introduction

[2] Agricultural drought is commonly defined as the presence of insufficient soil water available for adequate crop and forage production. As a result, agricultural drought is typically monitored using soil moisture anomaly products generated during the growing season at weekly or monthly time scales [Anderson *et al.*, 2011]. Consistent estimates of soil moisture for drought monitoring can be obtained in various ways; for example through remote sensing or through modeling of the land-surface water budget. However, these estimates are not perfect and each method has characteristic uncertainties [Koster *et al.*, 2009; Jackson *et al.*, 2010]. Therefore, it is frequently desirable to merge independent realizations to obtain a more accurate unified estimate. Theoretically, the more independent data that are merged, the larger the reduction in the noise of the merged product and past work, in a variety of fields, has demonstrated the benefits of combining geophysical estimates obtaining from a range of observational and modeling resources into a single hybrid estimate (see, e.g., Ebert [2001] or Xie and Arkin [1996]). However, it is important to weigh the products based on their relative accuracies in order to minimize errors.

[3] Data assimilation using Kalman filter-based methodologies is one of the most commonly used approaches for

merging different products while taking into account their relative uncertainties. Kalman filter theory can be shown to be a recursive solution of the least squares problem [Sorenson, 1970] for an appropriate time frame. The solution of Kalman [1960] enables propagation of the best estimate and its errors in time, whereas in ordinary least squares the solution is assumed constant in time. The goal for both solutions is obtaining an estimate with minimized error variance. However, both solutions also require prior knowledge of product uncertainties to obtain an optimal analysis. In land data assimilation studies, Kalman filter-based methodologies often rely on ad-hoc statistical descriptions of errors in assimilated observations, model parameters, or model forcings. As a result, the relative weighting applied to modeled and observed soil moisture information by a land data assimilation is arguably subjective and does not necessarily reflect an optimized integration of independent data sources [Crow and Van Loon, 2006; Reichle *et al.*, 2008]. Therefore, our goal here is the development of an objective merging that is less dependent on uncertain, user-defined error assumptions.

[4] Triple collocation is a method that objectively obtains error estimates for three or more independent products. This method was originally introduced by Stoffelen [1998] and Caires and Sterl [2003] to estimate near-surface wind speed errors, and later applied in many hydrological applications. In particular Scipal *et al.* [2008] estimated the errors in passive microwave-, active microwave-, and model-based soil moisture products. Miralles *et al.* [2010] estimated errors in passive microwave-, station-, and model-based soil moisture products and validated the error estimates using watershed scale station-based data. Dorigo *et al.* [2010] evaluated the uncertainties of global passive microwave-, active microwave-, and model-based soil moisture products. Hain *et al.* [2011] estimated errors in passive microwave-, thermal infrared-, and model-based soil moisture realizations and

¹Hydrology and Remote Sensing Laboratory, USDA, Beltsville, Maryland, USA.

²Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, USA.

Corresponding author: M. T. Yilmaz, Hydrology and Remote Sensing Laboratory, USDA, BARC-WEST, Beltsville, MD 20705, USA. (tugrul.yilmaz@ars.usda.gov)

found passive microwave- and thermal infrared-based soil moisture products yield complementary soil moisture information. *Parinussa et al.* [2011b] estimated errors in passive microwave-, active microwave-, and antecedent precipitation index-based soil moisture products, then compared the triple collocation-based errors with data assimilation-based error estimates [Crow, 2007], and found very high correlation between the error estimates of these two techniques.

[5] It is relatively straight forward to use triple collocation as a means to estimate observation error covariance parameters required by land data assimilation systems [Crow and van den Berg, 2010]. However, applying triple collocation to the parameterization of modeling error within a Kalman filter is more difficult, and—to our knowledge—has not yet been attempted. In particular, a Kalman filter requires covariance information regarding background errors that vary in time according to flow conditions and/or the frequency of assimilated observations. In contrast, triple collocation provides only a temporally constant value of error representing a continuous, nonupdated integration of the forecast model. Likewise, triple collocation provides only information about the magnitude of modeling errors and not their source and/or impact on the full model forecast covariance matrix. Such information is often important in land data assimilation applications [Crow and Van Loon, 2006]. While potentially resolvable, these challenges suggest that the initial use of triple collocation-based modeling errors should be based on a relatively simple least square framework as opposed to a full data assimilation approach.

[6] Here we propose an objective methodology that does not require any user-defined error parameters as input. In this approach, different anomaly soil moisture products are merged in a least squares framework that relies on product error estimates obtained from triple collocation. Specifically, we have merged weekly, growing-season soil moisture anomaly products obtained from thermal remote sensing via the atmosphere land exchange inversion (ALEXI [Anderson et al., 2007a]) model, microwave remote sensing via the land parameter retrieval model (LPRM [Owe et al., 2008]), and the Noah [Ek et al., 2003] land surface model. The least squares framework is also able to provide estimates of uncertainty in the merged product, which could be used to augment existing drought products. It should be stressed that the presented approach is particularly well suited for agricultural drought applications commonly based on the detection of growing-season soil moisture anomalies at weekly to monthly time scales. The proposed methodology can also potentially add value to soil moisture products derived from current and future soil moisture satellite missions (i.e., SMOS (soil moisture and ocean salinity); SMAP (soil moisture active passive)) by optimally merging them with various independent soil moisture estimates.

[7] The general least squares solution is briefly reviewed in section 2. Section 3 reviews the triple collocation equations, section 4 introduces the input data, section 5 presents the results, and section 6 summarizes our conclusions.

2. Least Squares Merging

[8] Least squares is an estimation theory that has been used in numerous studies since its initial applications by Gauss [1963] and Legendre [1806]. Kalman [1960] has

shaped the theory into its current form [Sorenson, 1970], which can be used to describe the basis of most modern data assimilation techniques [Talagrand, 1997]. The least squares solution has been derived in many studies; a brief review is given here to provide background for our proposed merging algorithm.

[9] Assume we have three independent realizations (S_x , S_y , and S_z) of a variable along with their respective zero-mean errors (ϵ_x , ϵ_y , and ϵ_z) and error variances (σ_x^2 , σ_y^2 , and σ_z^2). These realizations can be represented by

$$S_x = \alpha S_t + \epsilon_x, \quad (1)$$

$$S_y = \alpha S_t + \epsilon_y, \quad (2)$$

$$S_z = \alpha S_t + \epsilon_z, \quad (3)$$

where S_t is the true value of the variable and α is a measure of the relation between these realizations and the assumed truth. Although in some cases $\alpha = 1$, this is not a requirement; the least squares solution can be obtained as long as all realizations relate to the truth with the same coefficient. The desired merged estimate S_m is obtained as

$$S_m = w_x S_x + w_y S_y + w_z S_z, \quad (4)$$

where w_x , w_y , and w_z are the relative weights of S_x , S_y , and S_z , respectively. To have an unbiased merged estimate ($E[S_m - \alpha S_t] = 0$), it is required that $w_x + w_y + w_z = 1$. Given these constraints, the ultimate goal is to derive these weights as functions of the error variance of the three realizations and to find the error variance estimate of the merged product. The error estimate of the merged product is obtained as $\epsilon_m = S_m - \alpha S_t$ and the solution we seek minimizes a selected cost function (J) in a mean squares sense. Here we select J to be the error variance (σ_m^2) of the merged estimate in the form

$$J = \sigma_m^2 = w_x^2 \sigma_x^2 + w_y^2 \sigma_y^2 + w_z^2 \sigma_z^2, \quad (5)$$

$$J = \sigma_m^2 = w_x^2 \sigma_x^2 + (1 - w_x - w_z)^2 \sigma_y^2 + w_z^2 \sigma_z^2. \quad (6)$$

Setting $\partial J / \partial w_z = 0$ and $\partial J / \partial w_x = 0$ in equation (6) and solving for w_x , w_y , and w_z , we obtain

$$w_x = \frac{\sigma_y^2 \sigma_z^2}{\sigma_x^2 \sigma_y^2 + \sigma_x^2 \sigma_z^2 + \sigma_y^2 \sigma_z^2}, \quad (7)$$

$$w_y = \frac{\sigma_x^2 \sigma_z^2}{\sigma_x^2 \sigma_y^2 + \sigma_x^2 \sigma_z^2 + \sigma_y^2 \sigma_z^2}, \quad (8)$$

$$w_z = \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 \sigma_y^2 + \sigma_x^2 \sigma_z^2 + \sigma_y^2 \sigma_z^2}. \quad (9)$$

The solution is intuitive since the weights are proportional to the uncertainty of the other two estimates. If two

realizations are available instead of three, then the least squares solution can be applied similarly with a cost function selection of

$$J = \sigma_m^2 = w_x^2 \sigma_x^2 + (1 - w_x)^2 \sigma_y^2, \quad (10)$$

with weights

$$w_x = \frac{\sigma_y^2}{\sigma_x^2 + \sigma_y^2}, \quad (11)$$

$$w_y = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2}. \quad (12)$$

3. Error Estimation Using Triple Collocation

[10] For a given set of realizations, optimal merging based on the least squares technique described here requires an estimate of the relative uncertainties of the input products. In this study, the error variances of these estimates are obtained using triple collocation. In particular, we apply assumptions underlying the triple collocation approach of *Stoffelen* [1998],

$$S_1 = \alpha_1 S_t + e_1, \quad (13)$$

$$S_2 = \alpha_2 S_t + e_2, \quad (14)$$

$$S_3 = \alpha_3 S_t + e_3, \quad (15)$$

where S_t is the true soil moisture anomaly with variance σ_t^2 ; S_1 , S_2 , and S_3 are three soil moisture anomalies related to truth with α_1 , α_2 , and α_3 coefficients, with zero-mean errors e_1 , e_2 , and e_3 , and with error variances σ_1^2 , σ_2^2 , and σ_3^2 respectively. Here σ_t^2 does not imply the truth has errors, but rather it is the true soil moisture variance in time.

[11] In general, we may expect products to have differences in their relationship with the truth. Therefore, rescaling is required to ensure each data set has a consistent relationship with the assumed truth via reconciling equations (13)–(15) and equations (1)–(3). We rescale these realizations using

$$S_1^* = \alpha S_t + e_1^*, \quad (16)$$

$$S_2^* = \alpha S_t + e_2^*, \quad (17)$$

$$S_3^* = \alpha S_t + e_3^*, \quad (18)$$

where S_1^* , S_2^* , and S_3^* are the rescaled realizations and e_1^* , e_2^* , and e_3^* are the relative errors of the realizations with variances σ_1^{*2} , σ_2^{*2} , and σ_3^{*2} . Rescaled values are related to the initial estimates as $S_1^* = S_1 c_1$, $S_2^* = S_2 c_2$, and $S_3^* = S_3 c_3$, where c_1 , c_2 , and c_3 are the rescaling factors. By arbitrarily selecting any of the data sets as a reference (by setting $\alpha = \alpha_1$), and assuming error covariances between products

are zero and the representativeness errors described by *Stoffelen* [1998] are zero, these factors can be found as

$$c_1 = 1, \quad (19)$$

$$c_2 = \frac{\overline{S_1^* S_3^*}}{\overline{S_2^* S_3^*}} = \frac{\alpha_1}{\alpha_2}, \quad (20)$$

$$c_3 = \frac{\overline{S_1^* S_2^*}}{\overline{S_3^* S_2^*}} = \frac{\alpha_1}{\alpha_3}, \quad (21)$$

where overbar indicates averaging in time. In general, the accuracy of the rescaling in matching the data sets with the reference data set or the truth is tied to the strength of the linear relationship between the products assumed in equations (1)–(3). When compared to more nonlinear systems, highly linear systems are expected to have smaller sampling errors and require fewer observations to obtain similar levels of accuracy. Also, note that this rescaling step can be performed independently for each area or time period of interest, hence it may vary spatially and/or temporally.

[12] Assuming the errors of the products are independent from each other and from the truth, and assuming a mutual linear relationship between these estimates and the true soil moisture, the final error variances of the rescaled realizations are found as

$$\sigma_1^{*2} = \overline{(S_1^* - S_2^*)(S_1^* - S_3^*)}, \quad (22)$$

$$\sigma_2^{*2} = \overline{(S_2^* - S_1^*)(S_2^* - S_3^*)}, \quad (23)$$

$$\sigma_3^{*2} = \overline{(S_3^* - S_1^*)(S_3^* - S_2^*)}. \quad (24)$$

[13] Once these error variance estimates are obtained, they are inserted into equations (7)–(9) (or (11) and (12)) to obtain least squares weights at each time step. While the obtained error variance estimates are constant in time, these weights are not. When all three realizations are available, the least squares solution for three data sets (equations (7)–(9)) is used; when two out of three realizations are available then the least squares solution for two data sets (equations (11) and (12)) is used. Accordingly, equation (5) or (10) is used to calculate the reported error variance of the merged product at each time step.

[14] In the triple collocation system of equations presented above ((13)–(15)), there are seven unknown parameters (α_1 , α_2 , α_3 , σ_t^2 , σ_1^2 , σ_2^2 , and σ_3^2) constrained by three equations. By selecting a reference data set (i.e., assuming $\alpha = \alpha_1$) and rescaling other data sets to this reference, our goal in equations (16)–(18) becomes seeking a solution for four unknown parameters ($\alpha^2 \sigma_t^2$, σ_1^{*2} , σ_2^{*2} , and σ_3^{*2}), rather than seven. This system, with four unknowns and three equations, is still under-determined. We are able to solve for these four unknowns only after assuming all error related cross covariances are zero. Without this assumption, the $\alpha^2 \sigma_t^2$ term remains in the system of equations as an unknown.

[15] However, in the absence of any other independent information, we cannot decompose the $\alpha^2 \sigma_t^2$ estimate into

estimates of α^2 and σ_i^2 ; meaning we can never know the true σ_i^2 . Different reference data set selections result in different $\alpha^2 \sigma_i^2$ as well as different σ_1^{*2} , σ_2^{*2} , and σ_3^{*2} . Therefore the triple collocation equations described above provide only the relative accuracy of these realizations (how the noisiness of one product compares against that of another product), whereas the absolute values of the error variances themselves are dependent on the reference data set selection. While triple collocation is not ideal for capturing absolute errors, its representation of relative errors between input products is independent from the arbitrary choice of a single data set as a scaling reference. Fortunately, this type of relative information is all that is required in order to determine optimal least squares averaging.

4. Data

4.1. Input Data Sets

[16] The study area is selected as the contiguous United States (CONUS), between 125° and 67°W and 25° and 50°N. Consistent with our stated focus on the monitoring of agricultural drought during the growing season, daily data sets are obtained from 2002 to 2010 for the growing season months of April through October. Large-scale soil moisture information is currently available from three independent sources: retrievals derived from thermal-infrared remote sensing, retrievals derived from microwave remote sensing, and estimates derived from water balance models forced with micrometeorological observations. Here, all three sources of soil moisture data are used as input into the triple collocation analysis. In particular, this study utilizes an ALEXI energy balance model soil moisture proxy obtained from thermal infrared remotely sensed images, LPRM soil moisture estimates that are obtained from passive microwave remote sensing images, and Noah land surface model soil moisture simulations. The methodology was applied at a grid space of 0.25°; data sets at higher native resolution have been aggregated to this common grid. All data sets were averaged to weekly composites from their native temporal resolution.

[17] ALEXI is a two-source (soil and vegetation) model that solves for the latent and sensible heat components of the surface energy balance by taking advantage of measurements of morning land-surface temperature rise obtained by geostationary satellites [Anderson *et al.*, 1997; Mecikalski *et al.*, 1999; Anderson *et al.*, 2007a]. Using the obtained fluxes, a strong relationship was found between the ratio of actual to potential evapotranspiration fluxes (f_{PET}) and the fraction of available water (f_{aw}) in the soil column [Anderson *et al.*, 2007a, 2007b, 2011]. Following these studies, Hain *et al.* [2009] proposed unique relationships between f_{PET} and f_{aw} , evaluated this relation using soil moisture observations from the Oklahoma Mesonet Network, and showed ALEXI has valuable information about f_{aw} , which serves as a proxy for the root-zone soil moisture in the vegetated areas. Here we utilize ALEXI-based f_{PET} retrievals following the approach described by Hain *et al.* [2011]. Note that ALEXI f_{PET} represents a surface–root-zone merged soil moisture estimate, yielding a proxy estimate of water availability for evapotranspiration (i.e., water in the surface layer for bare soil evaporation, and water in the root zone for canopy transpiration). ALEXI f_{PET} values

have been aggregated from 10 km to 0.25° resolution. Given its reliance on the thermal remote sensing-based observations, current ALEXI retrievals are limited to clear-sky conditions, which is a major limitation to data availability particularly over the northern US. To fill the entire grid, it is necessary to average daily f_{PET} fields over time to create time composites. Detailed information about the derivation of an ALEXI-based soil moisture proxy can be found in the above mentioned studies.

[18] Noah (version 2.7) LSM data were obtained from the global simulations generated using Global Land Data Assimilation System (GLDAS [Rodell *et al.*, 2004]) forcing data. The Noah model employs a coupled surface water and energy balance, computing multilayer soil moisture as the storage component of the soil water balance. More details about these Noah simulations can be found at <http://disc.sci.gsfc.nasa.gov/hydrology/documentation>. These hourly simulations were performed at 0.25° spatial resolution, hence spatial aggregation was not needed. Since the ALEXI soil moisture proxy has mixed vertical support over sparsely and densely vegetated surfaces, a Noah soil moisture estimate was computed that mimics this vertical support. The second-layer (10–40 cm depth) and the third-layer (40–100 cm depth) soil moisture simulations were averaged into a root-zone soil moisture estimate (Noah_{root}) by weighting each layer volumetric soil moisture proportional to respective soil layer depths. This root-zone product and the surface (0–10 cm) soil moisture simulations (Noah_{surf}) were later combined into an adjusted soil moisture estimate (Noah_{adj}) following the method of Hain *et al.* [2011]

$$\text{Noah}_{\text{adj}} = (1 - f_{\text{vc}})\text{Noah}_{\text{surf}} + f_{\text{vc}}\text{Noah}_{\text{root}}, \quad (25)$$

where f_{vc} is the fractional vegetation cover based on remote sensing-based observations of leaf area index acquired by the moderate resolution imaging spectroradiometer (MODIS) as shown in Figure 1. As a result of equation (25), Noah_{adj} estimates are essentially surface soil moisture estimates over areas with no vegetation cover, and are root-zone soil moisture estimates over areas with dense vegetation cover. Here vegetation-adjusted soil moisture reflects soil moisture conditions in the part of the root zone that is in direct interaction with the atmosphere (by supplying water for evapotranspiration). Therefore, it is representative of the soil layer depth for sampling the moisture associated with vegetation and surface water stress and thus a natural metric for monitoring the severity of agricultural drought.

[19] Advanced microwave scanning radiometer EOS (AMSR-E) microwave remote sensing-based brightness temperature observations have been used in numerous passive microwave-based algorithms [Jackson, 1993; Owe *et al.*, 2001; Njoku and Chan, 2006; Lu *et al.*, 2009], and the resulting soil moisture products have been extensively validated under a wide range of ground conditions and climate regimes [Draper *et al.*, 2009; Mladenova *et al.*, 2011; Parinussa *et al.*, 2011a]. Here we utilize LPRM surface soil moisture retrievals [Owe *et al.*, 2008] obtained from Vrije University Amsterdam (VUA). LPRM soil moisture estimates are obtained using a one-layer model based on radiative transfer at the surface. This retrieval model uses soil related information as ancillary data, and solves for soil moisture, vegetation optical depth, and effective

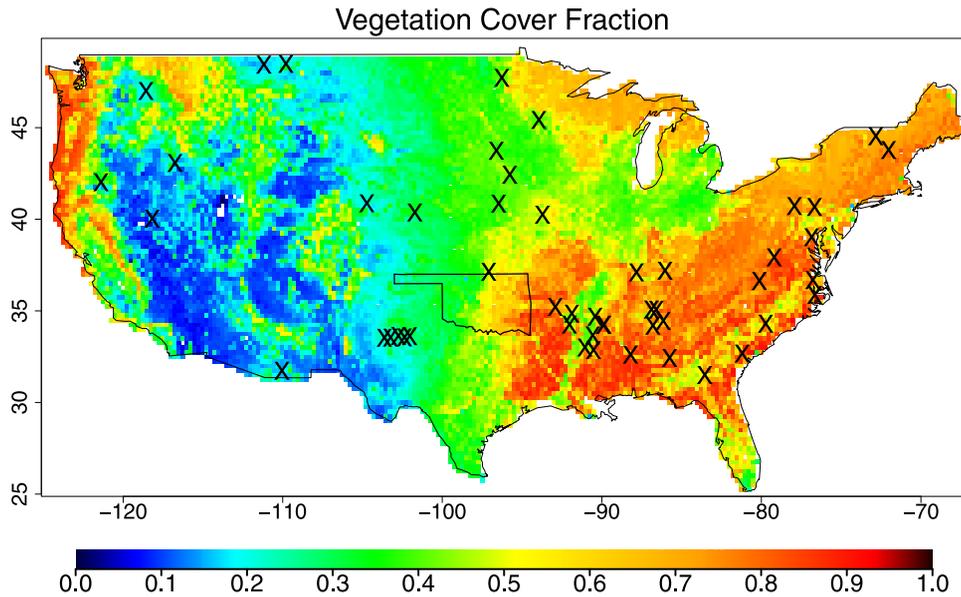


Figure 1. Average fraction of vegetation cover climatology over CONUS from April to October. The locations of SCAN stations used in the study are shown with symbol X. The state of Oklahoma, where all MESONET stations are distributed within, is shown with a black polygon.

physical soil temperature. The model uses the relationship between microwave polarization difference index, vegetation optical depth, and soil dielectric constant [Owe *et al.*, 2008], and solves for the skin temperature using a regression-based model based on Ka-band vertical polarization AMSR-E brightness temperature data [Holmes *et al.*, 2009]. Soil moisture retrievals used in this study are based on C- and X-band (mixed) descending AMSR-E brightness temperature observations. However, X-band observations are only also used in areas of the world where C-band observations are affected by radio frequency interference. Only the descending AMSR-E observations are used in this study primarily because they may result in more accurate soil moisture estimates than the observations acquired during ascending orbits [Draper *et al.*, 2009]. The LPRM soil moisture estimates refer to the top 3 cm of the soil profile. AMSR-E-based brightness temperature observations are obtained at native spatial resolutions of 56 and 38 km for the C- and X-bands, respectively. The operational LPRM product has been regridded to 0.25° spatial resolution by taking advantage of the multiple footprint centers that fall within the same 0.25° grid.

[20] Note that all three parent soil moisture data sets are obtained from different algorithms driven by different input data. These differences support the error independence assumption underlying application of triple collocation. On the other hand, these products also have different systematic and random relationships with the truth. However, here it is stressed that as long as highly linear relationships exist between the products, the data set retrieval method does not present a problem in a triple collocation framework. This issue will be revisited in section 5.4.

[21] In terms of timing, ALEXI provides a direct estimate of the soil moisture conditions at shortly before the local noon on days with clear morning conditions. LPRM soil moisture retrievals are obtained using microwave remote

sensing-based observations collected at 1:30 am local time. On the other hand, Noah soil moisture (SM) estimates are temporally continuous, and available at an hourly time interval. Accordingly, there could be inconsistencies between the weekly composites for each product due to differences in the temporal representativeness of each product. However, the impact of these inconsistencies should be minimized by the temporal averaging and standardization (see section 4.3 below) performed to obtain weekly composites.

[22] Given orbit patterns and typical frequency of masked retrievals, ALEXI and LPRM provide about 3.0 and 4.0 retrievals per week (respectively) over CONUS during the growing season (Figure 2). ALEXI has best temporal coverage over southwestern CONUS and LPRM has better temporal coverage over the midwest and southern CONUS (Figure 2). In addition to the satellite overpass availability, cloud cover and dense vegetation are the primary factors affecting the data availability of ALEXI and LPRM, respectively.

4.2. Validation Data Sets

[23] The merged product has been evaluated in comparison with in situ soil moisture observations from the Oklahoma MESONET Network [Brock *et al.*, 1995; Basara and Crawford, 2000] and the soil climate analysis network (SCAN [Schaefer *et al.*, 2007]) within the CONUS. In Oklahoma, an integrated network of 135 meteorological stations has been installed during the past two decades (Figure 1). Among these stations, around 100 have calibrated soil moisture monitoring devices taking measurements at 5, 25, 60, and 75 cm depths. Collected data undergo automated and manual quality controls conducted by University of Oklahoma during the conversion of 30 min raw data into daily soil moisture averages [Illston *et al.*, 2008]. There are over 150 SCAN stations (Figure 1) spread throughout the CONUS taking hourly soil moisture measurements at 5, 10,

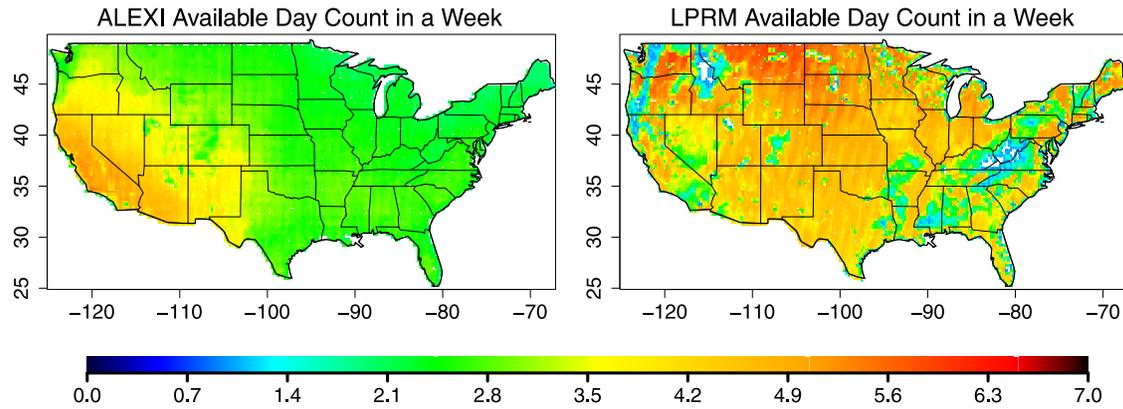


Figure 2. Average number of ALEXI and LPRM retrievals per week obtained between April and October (2002–2010).

20, 50, and 100 cm depths [Schaefer *et al.*, 2007], which later undergo quality control procedures.

[24] In a manner analogous to equation (25), a vegetation correction has been applied to the station measurements to ensure consistent soil moisture estimates between the merged products and the validation data sets. More specifically, the first layer (top 5 cm) MESONET data have been taken as surface soil moisture and a weighted average of the second to the fourth layers as a root zone. MODIS-based vegetation cover fraction (Figure 1) information on a 0.25° grid was interpolated from a native 8 day temporal resolution to weekly and is assumed to be of the station location. Vegetation correction to the MESONET data was carried out using equation (25). Similarly a vegetation correction was also applied to SCAN soil moisture values; the first layer soil moisture values were used as surface values and average soil moisture values of the second to the fifth layers, weighted by their depths, were used as root-zone values. The merged soil moisture estimates were validated using these vegetation-cover adjusted soil moisture observations. Because the MESONET and the SCAN station data were adjusted for vegetation cover fraction, the number of available station data points depends on the availability of both the surface and the root-zone observations. Since root-zone observations are not as readily available as surface observations, there are approximately only 50 MESONET and SCAN stations (Figure 1) available for verification.

[25] The skill of the triple collocation-based weights was also evaluated by comparing the performance of the merged estimate against the performance of a naively merged product performance, which simply assumes equal weights for each available product.

4.3. Data Standardization

[26] Weekly composites were standardized, so that their time mean (across years) is zero and time variance is unity for a given pixel and week

$$\mu_{w,\text{lon},\text{lat}} = \sum_y^{\text{nar}} \text{SM}_{y,w,\text{lon},\text{lat}} / \text{nar}, \quad (26)$$

$$\sigma_{w,\text{lon},\text{lat}} = \left[\sum_y^{\text{nar}} (\text{SM}_{y,w,\text{lon},\text{lat}} - \mu_{w,\text{lon},\text{lat}})^2 / \text{nar} \right]^{1/2}, \quad (27)$$

$$\text{SMS}_{y,w,\text{lon},\text{lat}} = \frac{\text{SM}_{y,w,\text{lon},\text{lat}} - \mu_{w,\text{lon},\text{lat}}}{\sigma_{w,\text{lon},\text{lat}}}, \quad (28)$$

where y , w , lon , and lat denote year, week, longitude, and latitude, respectively; SM denotes one of the three soil moisture products used in this study (ALEXI, Noah, and LPRM); SMS are the standardized soil moisture anomalies that are merged; and nar is the number of available realizations out of 9 years for the given week, longitude, and latitude. The climatologies are removed with the standardization process so that data sets have zero mean (consistent with the solution of Stoffelen [1998]) and unity standard deviation. Consequently, the triple collocation analysis and the merging process were performed solely using the standardized soil moisture anomalies defined above. Even though the merged product presented in this study is an anomaly product, it can be linearly transformed into absolute value products by selecting a reference data and applying the inverse of equation (28) using the standard deviation and mean values of the reference data set.

[27] There are only 9 years of data (9 points) available to calculate the mean and the standard deviation statistics for each week during the standardization process. The accuracy of these statistics could be improved by using longer time windows (i.e., 5 weeks). However, this may induce artificial autocorrelation in the standardized products and reduce the degrees of freedom for the triple collocation studies. Hence there is likely a trade-off between the accuracy of the standardized products and the accuracy of the triple collocation-based errors. In order to examine this issue, results were calculated using both a 1 and 5 week sampling window and separately evaluated.

4.4. Vertical Support

[28] The output product produced by ALEXI is a vegetation-adjusted (surface–root-zone merged) soil moisture estimate representing a proxy estimate of water available for evapotranspiration. Using equation (25), Noah soil moisture predictions can be converted into a variable with the same vertical support. However, LPRM soil moisture is associated only with the surface (0 to 3 cm) and therefore has a different vertical support than Noah and ALEXI soil moisture products over vegetated areas.

[29] To examine this effect, additional triple collocation analyses were performed using vegetation-adjusted LPRM values obtained similar to equation (25) using LPRM-based surface (native) and root-zone products. LPRM-based root-zone products required as input into equation (25) were obtained using the exponential smoothing methodology described by *Wagner et al.* [1999] and *Albergel et al.* [2008] to estimate root-zone soil moisture retrievals from superficial observations

$$\text{LPRM}_{\text{root}}(t) = \frac{\sum_i \text{LPRM}_{\text{srfc}}(t_i) e^{-(t-t_i)/\tau}}{\sum_i e^{-(t-t_i)/\tau}}, \quad (29)$$

where $t_i \leq t$, $\text{LPRM}_{\text{srfc}}$ is the surface LPRM soil moisture estimate at time t_i ; $\text{LPRM}_{\text{root}}$ is the root-zone soil moisture estimate; and τ is the characteristic time length. Specifically, three vegetation-adjusted LPRM products were estimated using three separate root-zone LPRM values obtained via assigning τ values of 4, 7, and 14 days. Accordingly, we have performed four parallel triple collocation analyses that use the same ALEXI and Noah data sets but different LPRM-based soil moisture values (one LPRM-surface product and three vegetation-adjusted LPRM products).

[30] In this study we also used CLM (version 2.0) simulations, solely for the investigation of surface–vegetation-adjusted soil moisture values coupling and not in the triple collocation merging methodology (section 5 below). Like Noah, CLM is a soil-vegetation-atmosphere transfer model that solves for the water and the energy balance at the surface [*Dai et al.*, 2003], and is driven here by GLDAS forcing data [*Rodell et al.*, 2004]. CLM simulations have 1° spatial resolution and utilize 10 soil layers with 2, 3, 4, 8, 12, 20, 34, 55, 92, and 113 cm depths, respectively. Consistent with Noah soil layer depths, vegetation-adjusted CLM soil moisture values were obtained (equation (25)) by using surface soil moisture estimates defined as the weighted average of the first to the third layers (0–9 cm) and using root-zone soil moisture estimates defined as the weighted average of the fourth to seventh layers (10–83 cm).

4.5. Additional Considerations

[31] For cross comparisons of the linear relation between parent products, cross correlations were calculated without setting any threshold for the availability of the products. The resulting correlation values were then masked if a significant correlation was not found. For the triple collocation we have set a minimum number (100) of mutually available data sets. If at least 100 mutually available soil moisture values were not found, then the triple collocation analysis was not performed and all error variance estimates were assumed equal. Here we note that these estimates are used at all time steps for the weight calculation process regardless of whether or not any of the products are missing.

[32] Even though the triple collocation-based errors are constant in time, weights used for merging products at each time step are calculated separately depending on the availability of the data sets for that particular time step. If all products are available, weights are calculated using equations (7)–(9); if only two products are available then weights are calculated using equations (11) and (12), while

the missing product is assigned a weight of zero; or if only a single product is available, then this product is assigned a weight of one and the two other missing products are assigned zero weight. Weights change in time only due to the availability of the products at any given time step. If there are not enough mutually available products, meaning a triple collocation-based estimate is not available and equal error variances are assigned, then products are merged using equal weights.

[33] The merged product at any given time can be based on anywhere between one and three realization(s). Hence, the uncertainty of the merged product (equation (6)) at any given location may not be constant in time; dates with more missing soil moisture values have higher uncertainty compared to dates with less missing values. Accordingly, for each available merged product, its uncertainty is also given as a separate product. There will also be temporal changes in the basis of merged predictions as the availability of ALEXI and LPRM changes over time. The lack of temporal consistency brought on by the intermittent availability of observations is a generic problem in land data assimilation. However, the issue is arguably more acute for our particular system since these variations will manifest themselves as abrupt variations in our analysis (rather than being smoothed over time as they would be in a full data assimilation system).

[34] Triple collocation-based error estimates are also dependent on the availability of the daily products, which influences the uncertainty of the sampled weekly composites. The more frequently a data set is available, the less noisy its weekly composites become. Noah weekly estimates are based on 168 separate hourly Noah soil moisture predictions generated each week (i.e., 24 estimates/day times 7 days), while ALEXI and LPRM have on average 3–4 estimates per week. Hence, Noah has better “weekly” temporal support than do the other products. However, it should be noted that poor support is simply one component of the total random error detected by triple collocation and therefore poses no particular challenge for our proposed merging strategy.

5. Results

5.1. Correlations and Weights

[35] ALEXI-, Noah-, and LPRM-based soil moisture anomaly estimates were used to calculate the error variances of each product in a triple collocation framework. As triple collocation-based error estimates require a mutual linear relationship between products, we have evaluated the linearity between the three products by analyzing their cross correlations (Figure 3) using anomaly products. Significant correlations between LPRM and ALEXI, and between LPRM and Noah over large parts of Eastern CONUS are not found (consistent with *Hain et al.* [2009]), which is partly due to the nonavailability of LPRM soil moisture estimates caused by the strong attenuation of the microwave signal over densely vegetated areas. On the other hand, there are strong cross correlations over areas of central CONUS, indicating a strong mutual linear relationship between various soil moisture products.

[36] The triple collocation-based errors were computed using equations (22)–(24) and were used in the least

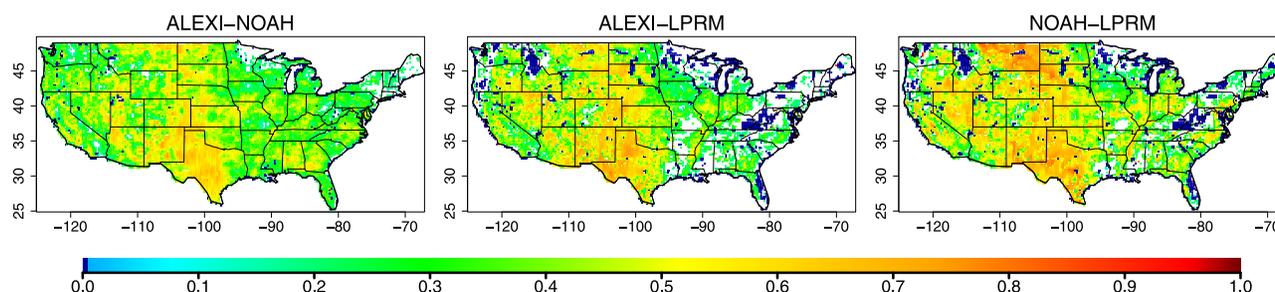


Figure 3. Cross correlations (r) between weekly ALEXI, Noah, and LPRM composites during 2002–2010 using months April through October. Regions lacking significant correlation (at 95% confidence level) are plotted as white and the regions where either correlated product is consistently missing are given in dark blue.

squares framework to obtain weights using equations (7)–(9). The resulting weights shown in Figure 4 are intuitively consistent with the cross correlations of the products (Figure 3); the product that has the highest cross correlation with its pairs also has the largest estimated weights. For example, the correlations between Noah and ALEXI and between Noah and LPRM are higher than the correlations between ALEXI and LPRM over North Dakota; therefore, Noah weighting is relatively higher than both ALEXI and LPRM over this area (Figure 4, top row). Similarly, the correlations between LPRM and ALEXI and between LPRM and Noah are higher than the correlations between ALEXI and Noah over Montana; therefore, the optimal weighting applied to LPRM retrievals is higher than the weighting of ALEXI and Noah over this area. In general, ALEXI performs better over southern CONUS, which can be attributed to the lower temporal coverage of ALEXI over northern CONUS due to clouds [Hain *et al.*, 2011].

[37] This study focused on the warm season mainly because agricultural drought is associated with the warm season over the study area; however, it is possible to perform the methodology for any time period using various data sets. In general, we may expect remote sensing-based weekly soil moisture averages retrieved during winter to have higher sampling errors than estimates retrieved during summer due to larger data gaps (both temporally and spatially) partly caused by snow and frozen soil conditions. Hence, a single set of weights for the entire year may not reflect the error characteristics as well as monthly derived weights. Unfortunately, the estimation of monthly weights likely requires longer time series than are currently available.

5.2. Merged Estimate and Station Data

[38] All subsequent merging results are based on the case of no LPRM smoothing (i.e., the top row in Figure 4). For the merging methodology, the weights in Figure 4 are used only when all three the data sets are available; for missing days, weights were calculated using the error estimates of the available days. Parent products (ALEXI, Noah, and LPRM), the merged estimate (merged realization using least squares), and the uncertainty of the merged estimate for the 19th week (7–13 May) of 2007 is shown in Figure 5. For this particular week, the standard deviation of the error estimate is around 0.40 (unitless as all products are standardized anomalies), and the soil moisture anomalies range between -2.6 and $+2.7$ standard deviations around the climatology

of the given local pixel. Note that the merged product is a standardized anomaly product that should generally range between about -3 and $+3$. However, it is emphasized that this can be easily linearly transformed into absolute value products by using the statistics (i.e., mean and standard deviation) of any reference data set. In fact, such linear transformations are very common in hydrological applications (i.e., variance matching, cumulative distribution function matching, etc.), particularly in land data assimilation studies [i.e., Reichle and Koster, 2004].

[39] Time series of the parent products, the merged estimate, and the uncertainty of the merged estimate are shown together with data from two individual MESONET and SCAN stations in Figure 6. The weights of the parent products are similar at these station points; hence, the merged estimates fall between the three parent products without closely following any one in particular. On average, 82% of the time the station data (both MESONET and SCAN) fell between the $\pm 2\sigma$ lines (while we expect around 95% of the time), indicating that the triple collocation-based errors are slightly underestimating the uncertainty of the products and/or the difference is due to the representativeness errors of the station data (Figure 6). Average station data correlations with the parent products and the merged estimate are summarized in Table 1; the significance of these correlations, the correlation comparisons of parent products, and the merged estimate are given in Table 2. On average, parent products are better correlated with the MESONET data than with the SCAN data (first three rows of Table 1). The number of stations that have significant correlations with the parent products and the merged estimate are higher for the MESONET data than for the SCAN data (first four rows of Table 2). The merged estimates are better correlated with the station data than the individual parent products (first four rows of Table 1), particularly better than both ALEXI and LPRM (fifth and seventh rows of Table 2)—implying the merged product is more accurate than its parents products individually. Although, on average, the merged estimate has better correlation with the MESONET (but not SCAN) than the best correlation of the parent products, the improvement is not significant for the majority of the stations (rows 11 and 12 of Table 2).

[40] Here the station-based vegetation-adjusted soil moisture is better correlated with the surface than root-zone soil moisture, while we expect vegetation-adjusted soil moisture to converge to the surface soil moisture over nonvegetated

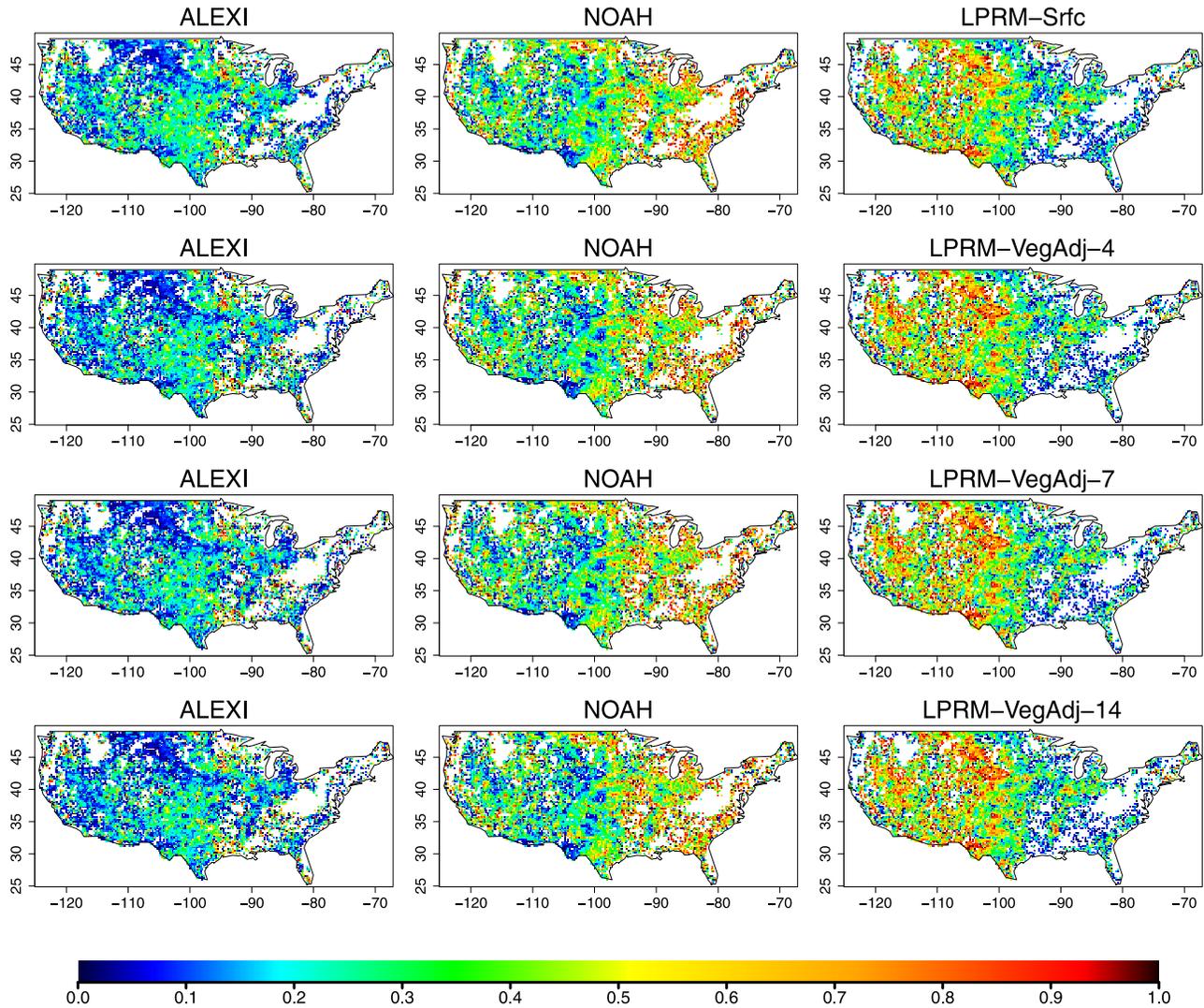


Figure 4. Weights of soil moisture estimates obtained from triple collocation. All four rows used the same ALEXI and Noah products in the triple collocation analysis. The first row used the native LPRM surface soil moisture product, whereas the second to fourth rows used exponentially filtered LPRM-based root-zone soil moisture products with characteristic time lengths of 4, 7, and 14 days, respectively. White areas correspond to equal 1/3 weighting where triple collocation results are unavailable.

areas (Table 1). In general the vegetation-adjusted product is highly related to both surface and the root-zone soil moisture products, while the relative magnitude of these relations are driven by the vegetation cover fraction of the area of interest. Given the average fraction of vegetation over the study area is 0.37 (Figure 1), it is expected that the vegetation-adjusted soil moisture will on average be more strongly associated with the surface than with the root zone in this experiment.

[41] The above results were obtained by calculating the anomalies using a weekly sampling window (see section 4.3). Sampling errors in the mean and standard deviation estimates that were used to create these anomalies could potentially degrade the quality of results. Therefore, we repeated this analysis using a longer 5 week (versus 1 week) sampling window centered on the week-of-interest to calculate Noah, ALEXI, and LPRM anomalies. While a longer sampling window decreases sampling errors in estimated anomalies, it also

obstructs the effective characterization of short-term soil moisture anomalies and decreases the number of effective degrees of freedom in the anomaly time series (by inducing temporal autocorrelation). Since results from this sensitivity analysis (not shown) suggest that the negative impacts of a longer sampling window outweigh the positives (i.e., all cross correlations referred in Table 2 are degraded), we kept the standardization process as described in section 4.3.

5.3. Comparisons to Naive Merging

[42] Similar to the triple collocation-based merging scheme, naive merging also leads to an integrated product that was generally better than any of its three parent products in isolation (Table 2). However, the triple collocation-based merged estimate did not generally lead to an integrated product that was demonstrably superior to naive merging (i.e., merging with equal weighting) (Table 2). Here we stress that the analyses in Table 2 exclude stations

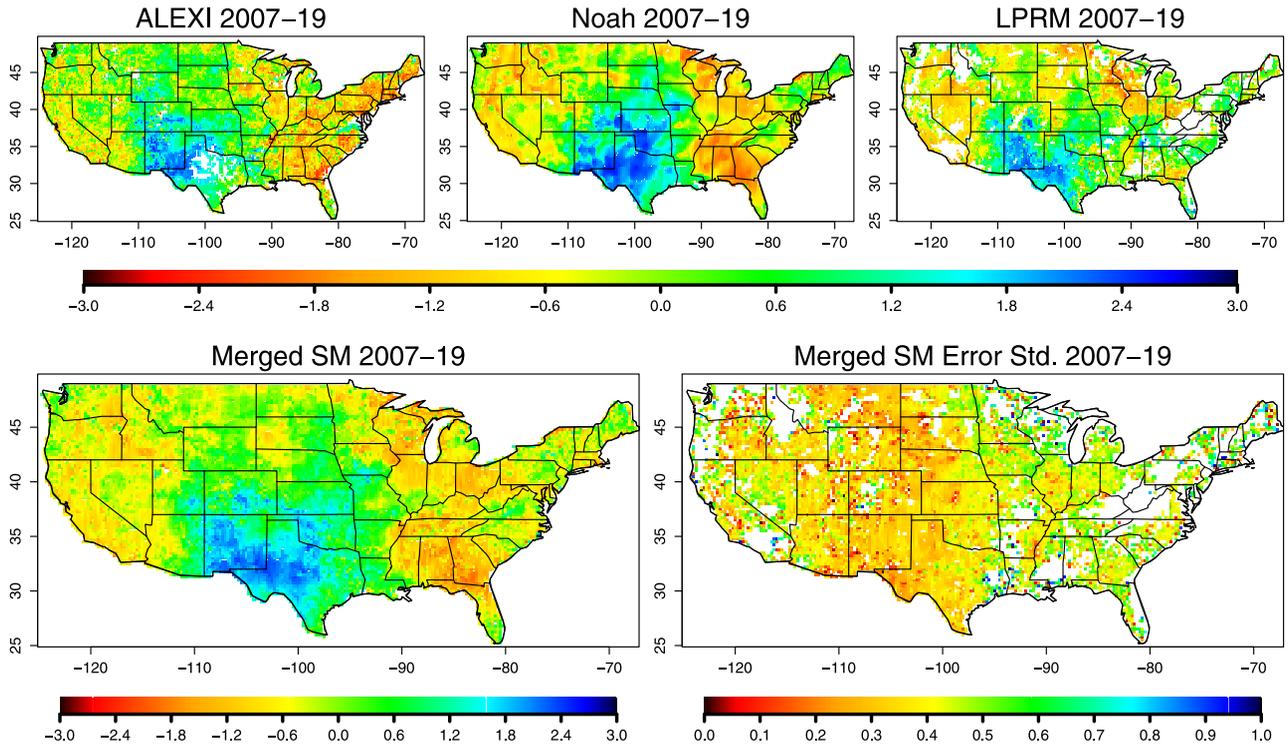


Figure 5. Example weekly composites of ALEXI, Noah, LPRM, merged soil moisture, and its uncertainty estimates for the 19th week of 2007. Soil moisture estimates and the error of the merged product are presented in terms of standard normal deviates. White areas in the error image correspond to areas where triple collocation results are unavailable.

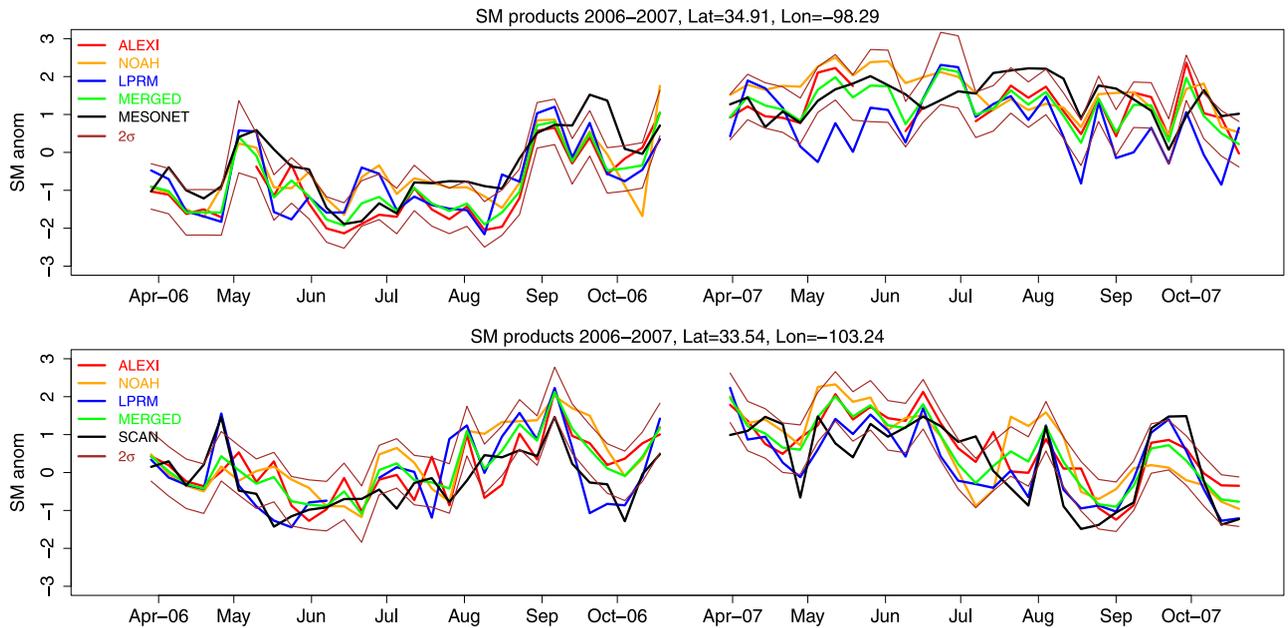


Figure 6. Weekly soil moisture composite time series in terms of standard normal deviates for the Apache (MESONET) station (Figure 6, top) and the crossroads (SCAN) station (Figure 6, bottom). ALEXI, Noah, and LPRM values are obtained from pixels that include the available station. Brown lines indicate the 2σ confidence interval of the merged product calculated using the uncertainty estimates obtained from equation (5) or (10).

Table 1. Parent Products (ALEXI, Noah, and LPRM), Merged Estimate, and Station Data (MESONET or SCAN) Cross Correlations With the Station Data^a

	MESONET			SCAN		
	Surface	Veg. Adj.	Root	Surface	Veg. Cor.	Root
ALEXI	0.46	0.48	0.38	0.36	0.38	0.34
Noah	0.54	0.54	0.33	0.41	0.42	0.33
LPRM	0.52	0.55	0.43	0.51	0.54	0.51
MERGED	0.61	0.63	0.46	0.55	0.58	0.51
NAIVE	0.61	0.64	0.46	0.55	0.57	0.50
MESONET or SCAN (Surface)	1.00	0.91	0.37	1.00	0.91	0.67
MESONET or SCAN (Veg. adj.)	0.91	1.00	0.60	0.91	1.00	0.78
MESONET or SCAN (Root)	0.37	0.60	1.00	0.67	0.78	1.00

^aThree layers of station soil moisture data are considered: surface, vegetation adjusted, and root zone. NAIVE refers to the merged product obtained by giving equal weight to each parent product.

where root-zone station-based soil moisture observations (primary reason) and triple collocation error estimates are not available. This indicates that the lack of difference between the triple collocation results and naive (i.e., equal) weighting results is due to the approximate equal weighting by triple collocation (and not a lack of data availability). Potential reasons for the lack of significant improvement in our merged product versus the baseline parent and the naively merged products include (1) station data are point data and may have high representativeness errors [Ryu and Famiglietti, 2005; Miralles et al., 2010; Cosh et al., 2006], and/or (2) triple collocation-based errors may not be optimal due to inadequate mutually available data (limited temporal extent of parent products), and/or (3) the weights are optimal, but the parent products may have similar skills and therefore merging them in a naive way produces estimates that are only marginally different from the optimally merged estimates obtained via triple collocation.

[43] In particular the station observations are point data, and thus very susceptible to representativeness errors and the weights obtained through triple collocation are very sensitive to the number of mutually available data. It is our experience that the number of mutually available triplets in this study may not be sufficient for highly accurate triple collocation estimates on weekly or monthly time scales

[Zwieback et al., 2012]. However, as longer time series become available through remote sensing techniques and modeling, and as improved station data (with less representativeness errors via better selection of station and/or sensor locations) are collected, it is expected that the merged estimates will result in higher improvements over the parent products.

[44] The difference between the optimal solution and the naive method was also evaluated by investigating the sensitivity of the optimal solution to data availability and averaging. Specifically, the triple collocation-based weights and the cross correlations for various averaging window lengths (weekly or monthly) were calculated (Table 3) to evaluate the sensitivity of derived optimal weights to aggregation period and retrieval availability. To do this, the daily data were averaged into either weekly or monthly composites by using all the available daily data for averaging (i.e., the “all available scenario”) or using only the days when all three products are available (i.e., the “mutually available scenario”). Applying the mutually available scenario guarantees that equal numbers of daily products are used in weekly or monthly composites analyzed via triple collocation. Later, the weights and the correlations were spatially averaged into a single value, for each scenario separately, by filtering out the areas that do not have reliable triple

Table 2. Results of Product Versus Ground-Data Cross-Correlation Analysis for Various Scenarios^a

Scenario	Product	MESONET				SCAN			
		Total	Neg	Non	Pos	Total	Neg	Non	Pos
Correlations significantly different than 0	ALEXI	50	0	2	48	44	0	5	39
	Noah	50	0	1	49	44	0	4	40
	LPRM	50	0	1	49	44	0	7	37
	MERGED	50	0	0	50	44	0	2	42
MERGED estimate correlations better than individual products (no significance test)	ALEXI	50	5	–	45	44	4	–	40
	Noah	50	12	–	38	44	13	–	31
	LPRM	50	4	–	46	44	7	–	37
NAIVE estimate correlations better than individual products (no significance test)	ALEXI	50	3	–	47	44	3	–	41
	Noah	50	10	–	40	44	17	–	27
	LPRM	50	7	–	43	44	10	–	34
MERGED best significantly	ALL	50	0	49	1	44	0	44	0
MERGED best	ALL	50	0	19	31	44	0	23	21
NAIVE best significantly	ALL	50	0	47	3	44	0	43	1
NAIVE best	ALL	50	0	18	32	44	0	27	17

^a“Total” refers to the number of ground stations considered. “Neg” and “Pos” refer to statistically significant negative and positive results, respectively, for the scenarios given in the left column, and “Non” refers to neither a positive result nor a negative result. For the significance tests, a 95% confidence level is used.

Table 3. Mean Weights and Cross Correlations Over CONUS for Different Data Compositing Strategies^a

	Weights		
	ALEXI	Noah	LPRM
Mutually available weekly	0.27	0.33	0.41
Mutually available monthly	0.33	0.29	0.37
All available weekly	0.24	0.38	0.37
All available monthly	0.31	0.34	0.35
	Correlations		
	ALEXI-Noah	ALEXI-LPRM	Noah-LPRM
Mutually available weekly	0.38	0.40	0.43
Mutually available monthly	0.44	0.45	0.45
All Available weekly	0.40	0.38	0.44
All Available monthly	0.46	0.44	0.46

^aMutually and all available scenarios refer to the parent products that are used in data standardization (section 4.3).

collocation estimates (for weight averages) or have insignificant correlations (for correlation averages). In general, the differences in product weights were higher than the differences between product cross correlations for weekly all available scenario, while the weight differences were much less for the weekly mutual scenario and both monthly scenarios (Table 3). This implies that the weighting favors products with higher temporal availability (=model) for weekly scenarios, but the effect of this retrieval frequency is reduced when data sets are averaged for longer time periods. This reduced difference in weights and correlations can also explain the similarity between the performance of merged products based on triple collocation and naive weighting (i.e., naive merging is equal to having 0.33 weights, hence the smaller the difference between weights of the products, the more it is similar to naive merging). Since the skills of the parent products are very similar, the naive averaging approach simply follows the optimal solution obtained via triple collocation.

5.4. Vertical Support

[45] As discussed above, the final merged soil moisture estimate is a mixed product that reflects the soil moisture

layer that is actively interacting with the atmosphere via evapotranspiration. Hence, using the surface-only microwave remote sensing product over sparsely vegetated areas is consistent with the properties of the mixed product. However, over vegetated areas this mixed vertical support is inconsistent with microwave-based soil moisture retrievals, which are strictly limited to the near-surface layer (surface to 3 cm). Consequently, over densely vegetated areas there is a potential inconsistency in the vertical support of LPRM soil moisture retrievals relative to ALEXI and Noah products (see above). A series of analyses has been performed to test the effect of using the surface-only microwave remote sensing product on triple collocation results over vegetated areas.

[46] Since the parameter of interest is the vegetation-adjusted soil moisture value (rather than root-zone soil moisture), we have narrowed our focus to this parameter. High correlations between surface and vegetation-adjusted soil moisture values at weekly time scales over densely vegetated areas imply a strong linear relation between the surface and the vegetation-adjusted soil moisture simulations, similar to the triple collocation equations (equations (16)–(18)) where we assume a linear relation between each data set and the truth. Therefore the applicability of these equations to soil moisture products obtained at different vertical depths is determined by the linearity of the relationship between surface and vegetation-adjusted soil moisture. The depth variations pose a problem to this application only if they manifest themselves in a nonlinear or a hysteric relationship between products. Conversely, if the relationship is linear, it simply folds into the linear rescaling step which underlies the application of triple collocation. Therefore the impact of vertical consistency (between LPRM and Noah/ALEXI-based soil moisture products) will hinge on the degree to which soil moisture estimates at various depths can be linearly related.

[47] Correlations were computed between the surface and vegetation-adjusted soil moisture values from both Noah and CLM (Figure 7) and both MESONET and SCAN station data (Table 4). Very high correlations (i.e., linear relationships) were found between the surface and the vegetation-adjusted station-based soil moisture data from station-based analysis (in Table 4, 0.91 for both MESONET

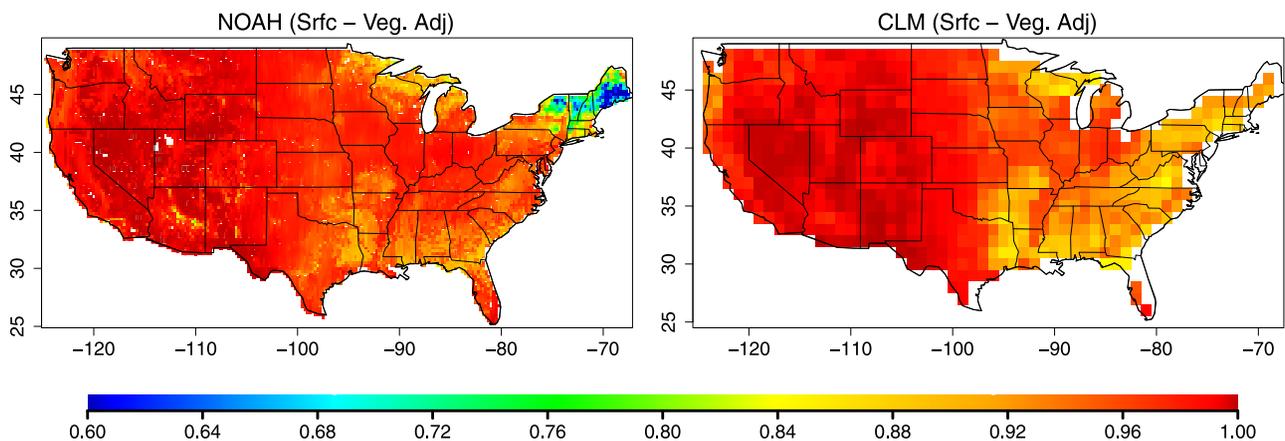


Figure 7. Weekly composite correlations between surface and vegetation-adjusted Noah and CLM soil moisture estimates.

Table 4. Noah, CLM, and Station Cross Correlations Between Surface and Vegetation-Adjusted Weekly Soil Moisture Composite Values at Multiple Locations^a

Surface–Veg. Adj.	MESONET Stations	SCAN Stations	CONUS	CONUS-East	CONUS-West
Noah	0.95	0.96	0.96	0.96	0.99
CLM	0.96	0.96	0.96	0.92	0.99
MESONET	0.91	–	–	–	–
SCAN	–	0.91	–	–	–

^aCONUS-East lays between 88° and 75° W, and 32° and 41° N and CONUS-West between 116° and 103° W and 29° and 36° N.

and SCAN data) and from model simulations (in Table 4, 0.96 correlations over CONUS for Noah and CLM, respectively). This suggests that at weekly time scales vertical inconsistencies in support can be effectively resolved via linear rescaling.

[48] Another way to test the potential impact of the surface-only LPRM data products is to transform it into an approximated root-zone product using a low-pass filter. In general, the differences between triple collocation analyses that use different LPRM products (corresponding to various amounts of temporal smoothing via equation (29)) are minimal (Figure 4), suggesting the nonlinearities due to vertical support differences do not have a major impact on estimated weights (supporting the correlation results given above).

6. Discussion and Conclusions

[49] Model error covariance estimates in many hydrological data assimilation applications are obtained through random perturbation of forcings and states without any rigorous justification for the magnitude of these perturbations [Reichle *et al.*, 2008; Crow and van den Berg, 2010]. Since ensemble spread tends to be a stronger function of forcing spread than initial condition spread [Yilmaz *et al.*, 2012], this results in a merging scheme that is dependent on the user to accurately characterize modeling errors which, in turn, determine the relative weight applied to model background and observations at update times.

[50] Here we introduce a methodology that is completely objective and does not assume any arbitrary assumptions concerning the error characteristics of its input data sets. Specifically, error variances of three independently estimated soil moisture data sets are obtained using a triple collocation method and different soil moisture products are merged in an ordinary least squares framework. With the objective analysis introduced here, we are also able to estimate the uncertainty of the merged soil moisture as a separate product, which could be particularly useful for applications which require information concerning the product reliability.

[51] Disadvantages of this framework when compared to traditional data assimilation techniques include the limitation that estimated model errors are assumed to be constant in time and corrective information obtained via the merger is not temporally propagated forward in time (as in sequential filtering). On the other hand, the triple collocation/least squares approach is simple, highly transparent, and based on an objective calculation of relative errors in various soil moisture estimates. As discussed above, it is likely possible (although not trivial) for triple collocation-based estimates of modeling error to be incorporated into a full data assimilation system. This analysis can be viewed as a preliminary step in that direction.

[52] In addition, there are three necessary requirements for the triple collocation analysis at the center of our approach: the independence of errors, the availability of long-enough time series, and the mutual linearity of products. The first assumption can be justified for many geophysical variables (i.e., soil moisture, soil temperature, potential evaporation, etc.) as there are numerous independent satellite- and model-based estimates of each. The second assumption (adequate data-series length) is required to minimize sampling errors in triple collocation estimates. While the 9 year data sets used here are adequate for this preliminary analysis, it should be noted that better triple collocation results may be obtained if longer data sets are results. Additionally the availability of longer time series will also enable estimation of separate sets of weights for seasonal or subseasonal time scales to partly address the issue of non-stationary weighting of products. The third assumption can be easily checked and the linearity can be justified via simple correlation calculations, as was done in this study. Nevertheless, the methodology is very flexible and can be applied in many hydrological applications given the availability of appropriate input data sets.

[53] In this study we have applied a triple collocation-based merging strategy to integrate soil moisture anomaly information acquired from microwave remote sensing, thermal remote sensing, and land surface modeling. The final merged product can be used as a standalone product particularly for agricultural drought monitoring applications. The approach also provides the ability to estimate uncertainties associated with the merger estimate. When compared to ground-based soil moisture observations, the merged product improves upon the accuracy of its three parent products but is not superior to merged products obtained using naive equal weighting. Given the small differences found between cross correlations of products and between weights of products, the lack of difference between our results and naive weighting appears attributable to the marginal skill differences that exist between ALEXI-, Noah-, and LPRM-based soil moisture estimates over the CONUS. We expect the differences between the skills of triple collocation- and naive method-based merged products would be higher over study areas where the differences between the skills of the parent products are higher.

[54] **Acknowledgments.** We thank Jeffrey Basara of University of OK and Micheal Cosh of U.S. Department of Agriculture for the MESONET and SCAN soil moisture data sets. We also thank the anonymous reviewers for their constructive comments, which led numerous clarifications in the final version of the manuscript. Research was partially supported by NASA Terrestrial Hydrology Program grant NNX06AG07G entitled “Monitoring of Root-Zone Soil Moisture Using Multi-Frequency Observations of Surface Soil Moisture and Evapotranspiration.” U.S. Department of Agriculture is an equal opportunity provider and employer.

References

- Albergel, C., C. Radiger, T. Pellarin, J.-C. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, B. Piguet, and E. Martin (2008), From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations, *Hydrol. Earth Syst. Sci.*, 12(6), 1323–1337.
- Anderson, M. C., J. M. Norman, G. Diak, W. P. Kustas, and J. R. Mecikalski (1997), A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing, *Remote Sensing Environ.*, 60, 195–216.
- Anderson, M. C., J. M. Norman, J. R. Mecikalski, J. A. Otkin, and W. P. Kustas (2007a), A climatological study of evapotranspiration and moisture stress across the continental United States: 1. Model formulation, *J. Geophys. Res.*, 112, D10117, doi:10.1029/2006JD007506.
- Anderson, M. C., J. M. Norman, J. R. Mecikalski, J. A. Otkin, and W. P. Kustas (2007b), A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 2. Surface moisture climatology, *J. Geophys. Res.*, 112, D11112, doi:10.1029/2006JD007507.
- Anderson, M. C., C. Hain, B. Wardlow, A. Pimstein, J. R. Mecikalski, and W. P. Kustas (2011), Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental United States, *J. Clim.*, 24(8), 2025–2044.
- Basara, J. B., and T. M. Crawford (2000), Improved installation procedures for deep layer soil moisture measurements, *J. Atmos. Ocean. Technol.*, 17, 879–884.
- Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eiltsm (1995), The Oklahoma mesonet: A technical overview, *J. Atmos. Ocean. Technol.*, 12, 5–19.
- Caires, S., and A. Sterl (2003), Validation of ocean wind and wave data using triple collocation, *J. Geophys. Res.*, 108(C3), 3098, doi:10.1029/2002JC001491.
- Cosh, M. H., T. J. Jackson, P. Starks, and G. Heathman (2006), Temporal stability of surface soil moisture in the Little Washita river watershed and its applications in satellite soil moisture product validation, *J. Hydrol.*, 323, 168–177.
- Crow, W. T. (2007), A novel method for quantifying value in spaceborne soil moisture retrievals, *J. Hydrometeorol.*, 8, 56–67.
- Crow, W. T., and M. J. van den Berg (2010), An improved approach for estimating observation and model error parameters in soil moisture data assimilation, *Water Resour. Res.*, 46, W12519, doi:10.1029/2010WR009402.
- Crow, W. T., and E. Van Loon (2006), Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture, *J. Hydrometeorol.*, 7, 421–432.
- Dai, Y., et al. (2003), The common land model, *Bull. Am. Meteorol. Soc.*, 84(8), 1013–1023.
- Dorigo, W. A., K. Scipal, R. M. Parinussa, Y. Y. Liu, and W. Wagner (2010), Error characterisation of global active and passive microwave soil moisture datasets, *Hydrol. Earth Syst. Sci.*, 14, 2605–2616.
- Draper, C. S., J. P. Walker, P. J. Steinle, R. A. De Jeu, and T. R. Holmes (2009), An evaluation of AMSR-E derived soil moisture over Australia, *Remote Sensing Environ.*, 113(4), 703–710.
- Ebert, E. E. (2001), Ability of a poor man's ensemble to predict the probability and distribution of precipitation, *Mon. Weather Rev.*, 129(10), 2461–2480.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayand, and J. D. Tarpley (2003), Implementation of Noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model, *J. Geophys. Res.*, 108(D22), 8851, doi:10.1029/2002JD003296.
- Gauss, C. (1963), *Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*, 326 pp., Dover, New York.
- Hain, C. R., J. R. Mecikalski, and M. C. Anderson (2009), Retrieval of an available water-based soil moisture proxy from thermal infrared remote sensing. Part I: Methodology and validation, *J. Hydrometeorol.*, 10, 665–683.
- Hain, C. R., W. T. Crow, J. R. Mecikalski, M. C. Anderson, and T. Holmes (2011), An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling, *J. Geophys. Res.*, 116, D15107, doi:10.1029/2011JD015633.
- Holmes, T. R. H., R. A. M. De Jeu, M. Owe, and A. J. Dolman (2009), Land surface temperature from ka band (37 GHz) passive microwave observations, *J. Geophys. Res.*, 114, D04113, doi:10.1029/2008JD010257.
- Illston, B. G., J. B. Basara, C. A. Fiebrich, K. C. Crawford, E. Hunt, D. K. Fisher, R. Elliott, and K. Humes (2008), Mesoscale monitoring of soil moisture across a statewide network, *J. Atmos. Ocean. Technol.*, 25(2), 167–182.
- Jackson, T. J. (1993), Measuring surface soil moisture using passive microwave remote sensing, *Hydrol. Processes*, 7(2), 139–152.
- Jackson, T. J., et al. (2010), Validation of advanced microwave scanning radiometer soil moisture products, *IEEE Trans. Geosci. Remote Sensing*, 48(12), 4256–4272.
- Kalman, R. (1960), A new approach to linear filtering and prediction problems, *Trans. ASME J. Basic Eng.*, 82D, 35–45.
- Koster, R. D., Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma (2009), On the nature of soil moisture in land surface models, *J. Clim.*, 22, 4322–4335.
- Legendre, A. (1806), *Nouvelles Methodes Pour la Determination des Orbites des Cometes*, Courcier, Paris.
- Lu, H., T. Koike, H. Fujii, T. Ohta, and K. Tamagawa (2009), Development of a physically-based soil moisture retrieval algorithm for spaceborne passive microwave radiometers and its application to AMSR-E, *J. Remote Sensing Soc. Jpn.*, 29(1), 253–262.
- Mecikalski, J. R., G. R. Diak, M. C. Anderson, and J. M. Norman (1999), Estimating fluxes on continental scales using remotely sensed data in an atmospheric-land exchange model, *J. Appl. Meteorol.*, 38, 1352–1369.
- Miralles, D. G., W. T. Crow, and M. H. Cosh (2010), Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations, *J. Hydrometeorol.*, 11, 1423–1429.
- Mladenova, I., V. Lakshmi, T. J. Jackson, J. P. Walker, O. Merlin, and R. A. De Jeu (2011), Validation of AMSR-E soil moisture using l-band airborne radiometer data from national airborne field experiment 2006, *Remote Sensing Environ.*, 115(8), 2096–2103.
- Njoku, E. G., and S. K. Chan (2006), Vegetation and surface roughness effects on AMSR-E land observations, *Remote Sensing Environ.*, 100, 190–199.
- Owe, M., R. De Jeu, and J. P. Walker (2001), A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index, *IEEE Trans. Geosci. Remote Sensing*, 39, 1643–1654.
- Owe, M., R. De Jeu, and T. Holmes (2008), Multisensor historical climatology of satellite-derived global land surface moisture, *J. Geophys. Res.*, 113, F01002, doi:10.1029/2007JF000769.
- Parinussa, R., A. G. C. A. Meesters, Y. Y. Liu, W. Dorigo, W. Wagner, and R. de Jeu (2011a), Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model, *Geosci. Remote Sensing Lett. IEEE*, 8(4), 779–783.
- Parinussa, R. M., T. R. H. Holmes, M. T. Yilmaz, and W. T. Crow (2011b), The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations, *Hydrol. Earth Syst. Sci.*, 15, 3135–3151.
- Reichle, R. H., and R. D. Koster (2004), Bias reduction in short records of satellite soil moisture, *Geophys. Res. Lett.*, 31, L19501, doi:10.1029/2004GL020938.
- Reichle, R. H., W. T. Crow, and C. L. Keppenne (2008), An adaptive ensemble Kalman filter for soil moisture data assimilation, *Water Resour. Res.*, 44(3), W03423, doi:10.1029/2007WR006357.
- Rodell, M., et al. (2004), The global land data assimilation system, *Bull. Am. Meteorol. Soc.*, 85(3), 381–394.
- Ryu, D., and J. S. Famiglietti (2005), Characterization of footprint-scale surface soil moisture variability using Gaussian and beta distribution functions during the southern great plains 1997 (sgp97) hydrology experiment, *Water Resour. Res.*, 41, W12433, doi:10.1029/2004WR003835.
- Schaefer, G. L., M. H. Cosh, and T. J. Jackson (2007), The USDA natural resources conservation service soil climate analysis network (scan), *J. Atmos. Ocean. Technol.*, 24, 2073–2077.
- Scipal, K., T. Holmes, R. de Jeu, V. Naeimi, and W. Wagner (2008), A possible solution for the problem of estimating the error structure of global soil moisture data sets, *Geophys. Res. Lett.*, 35, L24403, doi:10.1029/2008GL035599.
- Sorenson, H. W. (1970), Least-squares estimation: From Gauss to Kalman, *IEEE Spectrum*, 7, 63–68.
- Stoffelen, A. (1998), Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, *J. Geophys. Res.*, 103(C4), 7755–7766.
- Talagrand, O. (1997), Assimilation of observations: An introduction, *J. Meteorol. Soc. Jpn.*, 75, 191–209.

- Wagner, W., G. Lemoine, and H. Rott (1999), A method for estimating soil moisture from ERS scatterometer and soil data, *Remote Sensing Environ.*, *70*(2), 191–207.
- Xie, P., and P. Arkin (1996), Analysis of global monthly precipitation using gauge observations, satellite estimates, and numerical model prediction, *J. Clim.*, *9*, 840–858.
- Yilmaz, M. T., T. Delsole, and P. R. Houser (2012), Reducing water imbalance in land data assimilation: Ensemble filtering without perturbed observations, *J. Hydrometeorol.*, *13*(1), 413–420, doi:10.1175/JHM-D-11-010.1.
- Zwieback, S., K. Scipal, W. Dorigo, and W. Wagner (2012), Structural and statistical properties of the collocation technique for error characterization, *Nonlinear Process. Geophys.*, *19*, 69–80.