Uncertainty of Reference Pixel Soil Moisture Averages Sampled at SMAP Core Validation Sites

FAN CHEN,[a] WADE T. CROW,[b] MICHAEL H. COSH,[b] ANDREAS COLLIANDER,[c] JUN ASANUMA,[d] AARON BERG,[e] DAVID D. BOSCH,[f] TODD G. CALDWELL,[g] CHANDRA HOLIFIELD COLLINS,[h] KARSTEN HØGH JENSEN,[i] JOSE MARTÍNEZ-FERNÁNDEZ,[j] HEATHER MCNAIRN,[k] PATRICK J. STARKS,[l] ZHONGBO SU,[m] AND JEFFREY P. WALKER[n]

[a] *SSAI/Hydrology and Remote Sensing Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland*
[b] *Hydrology and Remote Sensing Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland*
[c] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California*
[d] *University of Tsukuba, Tsukuba, Japan*
[e] *Department of Geography, Environment and Geomatics, University of Guelph, Guelph, Ontario, Canada*
[f] *Southeast Watershed Research Lab, Agricultural Research Service, USDA, Tifton, Georgia*
[g] *Nevada Water Science Center, U.S. Geological Survey, Carson City, Nevada*
[h] *Southwest Watershed Research Center, Agricultural Research Service, USDA, Tucson, Arizona*
[i] *Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark*
[j] *University of Salamanca, Villamayor, Spain*
[k] *Agriculture and Agri-Food Canada, Ottawa, Ontario, Canada*
[l] *Grazinglands Research Laboratory, Agricultural Research Service, USDA, El Reno, Oklahoma*
[m] *Faculty of Geo-Information Science and Earth Observations (ITC), University of Twente, Enschede, Netherlands*
[n] *Monash University, Clayton, Victoria, Australia*

(Manuscript received 26 February 2019, in final form 6 June 2019)

ABSTRACT

Despite extensive efforts to maximize ground coverage and improve upscaling functions within core validation sites (CVS) of the NASA Soil Moisture Active Passive (SMAP) mission, spatial averages of point-scale soil moisture observations often fail to accurately capture the true average of the reference pixels. Therefore, some level of pixel-scale sampling error from in situ observations must be considered during the validation of SMAP soil moisture retrievals. Here, uncertainties in the SMAP core site average soil moisture (CSASM) due to spatial sampling errors are examined and their impact on CSASM-based SMAP calibration and validation metrics is discussed. The estimated uncertainty (due to spatial sampling limitations) of mean CSASM over time is found to be large, translating into relatively large sampling uncertainty levels for SMAP retrieval bias when calculated against CSASM. As a result, CSASM-based SMAP bias estimates are statistically insignificant at nearly all SMAP CVS. In addition, observations from temporary networks suggest that these (already large) bias uncertainties may be underestimated due to undersampled spatial variability. The unbiased root-mean-square error (ubRMSE) of CSASM is estimated via two approaches: classical sampling theory and triple collocation, both of which suggest that CSASM ubRMSE is generally within the range of 0.01–0.02 $m^3 m^{-3}$. Although limitations in both methods likely lead to underestimation of ubRMSE, the results suggest that CSASM captures the temporal dynamics of the footprint-scale soil moisture relatively well and is thus a reliable reference for SMAP ubRMSE calculations. Therefore, spatial sampling errors are revealed to have very different impacts on efforts to estimate SMAP bias and ubRMSE metrics using CVS data.

1. Introduction

In theory, satellite-based soil moisture retrievals represent a spatial average of the surface-layer soil water content within a ground-projected satellite footprint. The performance of these retrievals is usually evaluated against a set of in situ observations acquired within a reference footprint area. In practice, this is based on establishing a ground-based network of sensors from which observations are spatially aggregated, or upscaled, to obtain an average soil moisture value within the reference pixel. This strategy is the dominant validation practice for recent satellite soil moisture missions (e.g., Jackson et al. 2010, 2012; Colliander et al. 2017a; Kerr et al. 2016)

Corresponding author: Fan Chen, fan.chen@ars.usda.gov

DOI: 10.1175/JHM-D-19-0049.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

and, in particular, forms the basis of the core validation site (CVS) strategy employed to validate surface and root-zone soil moisture retrievals acquired by the NASA Soil Moisture Active Passive (SMAP) mission launched in 2015.

Although extensive efforts have been made to maximize the sampling coverage of the reference pixels, the spatial density of the sensor network is generally limited by financial, siting and maintenance constraints. This is particularly true for cases of significant subgrid variability in soil texture, land cover/land use (e.g., irrigation, mixed croplands and forests), complex topography, or in the presence of strong contrasts in land surface conditions/land use. A range of upscaling strategies has been proposed to mitigate such limitations (Crow et al. 2012). For example, studies have been conducted to find locations manifesting temporal stability that can be used directly, or with some bias correction, to accurately capture the large-scale spatial mean of a soil moisture field (Cosh et al. 2006, 2008; Brocca et al. 2009). However, even a statistically representative measurement site can miss heterogeneous wetting or drying events within the pixel and deviate from the true average for the duration, which makes it a less stable and desirable validation source than a network average acquired from dense random sampling (e.g., Ran et al. 2017). In addition, the temporal stability method is not recommended when cropped land is present, as representative stations in agricultural areas are often misidentified as temporally unstable due to unpredictable irrigation events (Yee et al. 2016). Other potential techniques include using block kriging, data collected during intensive sampling campaign, and land surface modeling to derive optimal upscaling functions (Crow et al. 2012), as well as the application of random forest regression (e.g., Clewley et al. 2017). However, despite these efforts, a perfect representation of the true soil moisture average of the satellite pixel remains elusive and some level of underlying uncertainty must generally be tolerated.

Over the past few decades, dense gravimetric sampling has been conducted to calibrate probe-based soil moisture networks at several USDA experimental watersheds and examine the temporal stability of network sensors (e.g., Cosh et al. 2006; Bosch et al. 2006; Cosh et al. 2008, 2013). Other studies have examined key processes driving soil moisture spatial variability over different spatial scales using extensive (but short-term) field experiment data (Famiglietti et al. 1999; Choi et al. 2007; Famiglietti et al. 2008). Based on empirical relationships derived from observed spatial variance and mean at the footprint scale, the number of soil moisture samples needed to meet certain accuracy levels can be calculated (Famiglietti et al. 2008). These results establish the reliability of using such in situ networks for the calibration and/or validation of satellite products and thus provide guidance on the construction of a soil moisture network for such a purpose. However, the uncertainty in the network average itself, as well as its impact on key validation metrics, has not been widely explored.

Soil moisture data collected from a set of established SMAP CVS provides an opportunity to examine the uncertainty in the sensor network spatial averages due to spatiotemporal sampling limitations, and thus clarify the implications of this uncertainty on the validation of satellite soil moisture products. The footprint-scale of interest here is the 33-km grid cells representing the ground spatial support of the SMAP level 2 enhanced soil moisture product (Chan et al. 2018; Jackson et al. 2018). At each SMAP CVS, a core site average soil moisture reference value (CSASM) is typically obtained as the weighted spatial average of point-scale ground observations acquired within the domain of the corresponding soil moisture retrieval grid cell (i.e., the reference pixel). Note that, in this context, the CSASM is an error-prone estimate of the reference pixel true soil moisture average rather than the truth itself.

The goal of this paper is to quantify the uncertainty, or random sampling error, present in the current SMAP 33-km-scale CSASM dataset, for example, deviation between CSASM θ_{WA} and the true soil moisture average θ_T of the reference pixel. Two useful descriptions—based on both first- and second-order error statistics—of this uncertainty are examined. First, the uncertainty in the long-term CSASM mean values expressed as confidence intervals, which determine the significance of the sampled SMAP biases at CVS locations. Second, estimates of the unbiased root-mean-square error (ubRMSE) between daily CSASM and θ_T. If available, such estimates can be applied to correct the raw ubRMSE metric in a validation analysis and thus recover the true SMAP-versus-θ_T ubRMSE from SMAP-versus-CSASM ubRMSE by removing the root-mean-square differences contributed by CSASM errors. The implications of these different types of CSASM uncertainty for the CSASM-based validation of remote sensing soil moisture products will be discussed.

This paper is organized as follows. Section 2 describes the classical statistical analysis and triple collocation approach used to derive the sampling uncertainties for CSASM obtained at SMAP CVS. Section 3 describes the in situ, remote sensing, and land surface modeling soil moisture datasets applied in the analyses. Results are presented and discussed in sections 4 and 5, respectively.

2. Estimate uncertainty of CSASM

As discussed above, suboptimal sampling density and upscaling functions leads to error in CSASM that then

directly impacts the reliability of the SMAP validation metrics calculated against the CSASM values. Here, CSASM error statistics are estimated from two perspectives: 1) a statistical analysis based on classical spatial sampling theory that estimates sampling uncertainty in CSASM temporal average and 2) a triple collocation (TC) analysis that quantifies random error variances in CSASM estimates of true soil moisture spatial averages.

a. *Spatial sampling uncertainty in time-averaged CSASM*

The CSASM is essentially a sampled mean of observations drawn from discrete sampling points within the reference pixel. In practice, purely random or stratified sampling across the domain is impossible. Instead, installed sites are typically selected based (in part) on practical considerations (e.g., accessibility and land ownership) and often preferentially sample certain land cover types. Nevertheless, the resulting network is generally assumed to provide a representative sample of spatial variability within the reference pixel. Thus, at least initially, the soil moisture observations acquired from the network are assumed here to be based on random spatial sampling. Based on this assumption, classical statistical analysis tools can be applied to estimate the probability distribution of their sample mean.

To start, a classical statistical analysis used to obtain confidence intervals for time average values of the CSASM due to random error in sparse spatial sampling is described. This time-averaged error in the CSASM directly reflects uncertainty in CSASM-based bias estimates (i.e., the temporal average of differences between satellite retrievals and CSASM). Given the difficulty of directly measuring CSASM sampling error, sampling uncertainty is estimated by applying the central limit theorem, based on an assumption of unbiased sampling with sufficient spatial coverage to guarantee independent sampling errors (Crow et al. 2005; see discussion above). This spatial sampling uncertainty, in turn, provides confidence intervals for the CSASM-based SMAP bias estimates.

Given an in situ soil moisture network consisting of N sensors, the weighted-average soil moisture (CSASM or θ_{WA}) at time j is obtained by

$$\theta_{\mathrm{WA},j} = \sum_{i=1}^{N} \theta_{i,j} w_i \,, \tag{1}$$

where w_i is the normalized weight assigned to the ith sensor with $\sum_{i=1}^{N} w_i = 1$, and $\theta_{i,j}$ is the observed volumetric soil moisture at location i and time j. Equation (1) is a typical form of upscaling function applied at SMAP CVS (Colliander et al. 2017a).

The spatial variance of time-averaged soil moisture for each of the N sensors, adjusted for uneven weighting, is then given as

$$\sigma^2_{\mathrm{WA}} = \frac{\sum_{i=1}^{N} w_i \sum_{i=1}^{N} \left[w_i (\overline{\theta_i} - \overline{\theta_{\mathrm{WA}}})^2 \right]}{\left(\sum_{i=1}^{N} w_i \right)^2 - \sum_{i=1}^{N} w_i^2} = \frac{\sum_{i=1}^{N} \left[w_i (\overline{\theta_i} - \overline{\theta_{\mathrm{WA}}})^2 \right]}{1 - \sum_{i=1}^{N} w_i^2}, \tag{2}$$

where the overbar for θ_i and θ_{WA} denotes temporal averaging. The standard error of CSASM is then obtained as

$$\mathrm{SE}_{\mathrm{WA}} = \sqrt{\sigma^2_{\mathrm{WA}} / N_{\mathrm{eff}}}, \tag{3}$$

where N_{eff} is the effective sample size (Kish 1965) and an approximation of the size of a simple random-sampling sample that produces the same spatial variance σ^2_{WA} obtained from the CVS network given by

$$N_{\mathrm{eff}} = \left(\sum_{i=1}^{N} w_i \right)^2 \Big/ \sum_{i=1}^{N} w_i^2 = 1 \Big/ \sum_{i=1}^{N} w_i^2 . \tag{4}$$

With the assumption that the weighted soil moisture observed at the CVS sensor network is randomly sampled within the reference pixel, the distribution of sample mean (CSASM) approaches a normal distribution. The 95% confidence interval for $\overline{\theta_{\mathrm{WA}}}$ can therefore be obtained as

$$\overline{\theta_{\mathrm{WA}}} \pm t_{0.025, N_{\mathrm{eff}}-1} \mathrm{SE}_{\mathrm{WA}} \,, \tag{5}$$

where $t_{0.025, N_{\mathrm{eff}}-1}$ is the critical value at 0.025 for the t distribution with $N_{\mathrm{eff}} - 1$ degrees of freedom. Stated in a different, but equivalent way, (5) indicates that, if SMAP retrievals were bias free, 95% of the sampled SMAP bias should fall within the range defined by $\pm t_{0.025, N_{\mathrm{eff}}-1} \mathrm{SE}_{\mathrm{WA}}$.

For time-averaged CSASM, an evenly spaced, fixed-location sampling network of appropriate density would provide sampled statistics that are unbiased with respect to a spatially random, independent sampling scheme, repeated many times temporally. Applying different weights within SMAP CVS is often aimed at compensating for uneven spatial sampling and to obtain a more representative (and reliable) estimate of the reference pixel mean. For example, in SMAP validation weights are generally assigned to be proportional to the area of Voronoi polygons determined for associated sampling points (Colliander et al. 2017a). Therefore, it is a corrective procedure applied to

compensate for nonrandom sampling. However, this correction comes at a cost since using N_{eff} to obtain the t value penalizes and inflates the random sample uncertainty of CSASM (since $N_{\text{eff}} \leq N$). Nevertheless, it is worth noting that (4) is not the only method to estimate N_{eff} and there is controversy over the definition and calculation of N_{eff} (e.g., Thiébaux and Zwiers 1984). Given this, CSASM uncertainty estimates have been calculated here using both N and N_{eff} in (5) and N_{eff} has been replaced with $N - 1$ in (3) when N is used. These two slightly different estimates of CSASM uncertainty are presented and discussed further in section 4a.

b. *The unbiased root-mean-square error (ubRMSE) of CSASM*

In addition to bias, ubRMSE has become a critical accuracy metric for recent satellite soil moisture assessments. By removing a fixed, long-term temporal average from both the reference and retrieval time series (prior to taking a root-mean-square error), the ubRMSE describes the combination of random error and time-varying seasonal differences in a given remote sensing product relative to the reference time series. However, ubRMSE of a remote sensing product calculated against CSASM (i.e., $\text{ubRMSE}_{\text{RS,WA}}$) reflects not only the true error of the product but also the random error in the CSASM itself. Given that the goal of any satellite validation project is estimating errors in the satellite product versus an unknown truth (i.e., $\text{ubRMSE}_{\text{RS},T}$), it is useful to estimate errors in the CSASM time series (i.e., $\text{ubRMSE}_{\text{WA},T}$) in order to isolate $\text{ubRMSE}_{\text{RS},T}$ from $\text{ubRMSE}_{\text{RS,WA}}$. If remote sensing and CSASM errors are assumed to be independent, these three quantities are linked as

$$\text{ubRMSE}_{\text{RS,WA}} = \sqrt{\text{ubRMSE}^2_{\text{RS},T} + \text{ubRMSE}^2_{\text{WA},T}}, \tag{6}$$

where the subscript RS represents a remote sensing product, WA represents CSASM (weighted average soil moisture of the reference pixel), and T denotes the true footprint-average soil moisture value. As noted earlier, two different approaches are applied below to estimate $\text{ubRMSE}_{\text{WA},T}$.

1) $\text{ubRMSE}_{\text{WA}}$ BASED ON SAMPLING THEORY

While $\text{ubRMSE}_{\text{WA},T}$ cannot be directly obtained from the available observations, it can be approximated by the temporal mean of the sampled spatial variance of temporal anomalies (relative to a single, long-term temporal mean) for cases where spatial sampling is assumed to be random such that

$$\text{ubRMSE}_{\text{WA},T} \approx \sqrt{\frac{\sum_{j=1}^{N_j}\left\{\sum_{i=1}^{N} w_i\left[(\theta_{i,j} - \overline{\theta_i}) - (\theta_{\text{WA},j} - \overline{\theta_{\text{WA}}})\right]^2\right\}}{(N_{\text{eff}} - 1)N_j}}, \tag{7}$$

where N_j is the number of temporal observations obtained by the soil moisture network. Ground-based observations at SMAP CVS are commonly obtained at high temporal frequencies ($\leq$30 min). However, here only the daily observations closest to each SMAP retrieval time [approximately 0600 and 1800 local solar time (LST)] were used in the calculation. As the CVS sampling networks are designed with the intention to best represent the corresponding reference pixel, (7) can generally be considered to give a close approximation of $\text{ubRMSE}_{\text{WA},T}$. The 95% confidence interval for $\text{ubRMSE}_{\text{WA},T}$ is calculated based on Faber (1999); that is, the uncertainty of the mean squared error of prediction can be approximated by a χ^2 distribution with the degrees of freedom equal to the number of samples.

2) $\text{ubRMSE}_{\text{WA}}$ BASED ON TRIPLE COLLOCATION ANALYSIS

Despite our assumption of random sampling, it should be noted that obtaining an unbiased sample of reference pixel conditions is challenging for sites with substantial heterogeneity in soil, vegetation, and/or topography—especially when soil moisture sensors are installed at permanent locations that favor certain land cover conditions. In such cases, systematic bias in the CSASM relative to the true reference pixel average is unavoidable, which cannot be estimated by statistical analysis using classical sampling theory alone. Neither can statistical analysis reveal the error variance of the CSASM due to the spatial variability not captured by the existing (potentially biased) spatial sampling scheme.

The direct measurement of total sampling error (including both sample bias and/or random error variance) requires access to data collected from (more) spatially intensive field work incorporating a much denser sampling network. Unfortunately, such data are not commonly available. In its absence, triple collocation (TC) analysis provides an alternative estimate of the "total" ubRMSE in the CSASM relative to the reference pixel truth, which includes time-varying errors contributed by instrumentation errors and unsampled spatial variability. In this way, the need to assume a purely random spatial sample is avoided. However, TC is based on a different set of assumptions—principally, mutually independent time-varying errors in each of its three input datasets.

Given this contrast in their underlying assumptions, it is of interest to compare the CSASM error estimates obtained from TC and classical sampling theory.

TC has been widely applied in remote sensing validation and assessment studies to estimate the error variance of a data product in the absence of an error-free benchmark (e.g., Draper et al. 2013). The application here is based on an assumption that independent estimates of θ_T can be obtained from ground-based observations via θ_{WA}, a land surface model θ_{LSM}, and remote sensing θ_{RS}, each of which is linearly related to θ_T by

$$\theta = \alpha\theta_T + \beta + \varepsilon, \tag{8}$$

where α and β are multiplicative and additive bias terms, and ε is random error. TC solves the error variance of the input datasets when the following principle assumptions are met: 1) zero error cross correlation, 2) zero correlation between errors and the truth, and 3) stationary signal and error statistics (Gruber et al. 2016; Draper et al. 2013; Zwieback et al. 2012).

In practice, the three measurement sources often present differences in representative depths, contributing area, dynamic ranges and seasonalities, as well as nonstationary α and β parameters that result in violation of the TC assumptions. Therefore, it is often necessary to transform estimates into anomalies by removing slowly varying temporal averages and/or climatologies present in each dataset (i.e., to remove β). Chen et al. (2017) demonstrates that the most robust TC results are achieved after the removal of seasonal trends in the form of 30-day moving averages from the input time series. In addition, to account for differences in the scaling factor α, one of the three datasets must be selected as a scaling reference (Chen et al. 2017).

Note that, applying the same ubRMSE-correction strategy via (6) when only point-scale observations (i.e., from sparse network stations) are available is not recommended (Chen et al. 2017) because, unlike the CSASM which can be assumed to be close to calibrated (i.e., $\alpha \approx 1$ and $\beta \approx 0$), an unknown α of the point-scale observation introduces additional uncertainty in the "corrected" ubRMSE of the satellite product via (6).

Because the error variance of the CSASM $\sigma^2_{\varepsilon_{WA}}$ is of interest here, the θ_{WA} anomaly (i.e., θ^*_{WA}) was chosen as the reference dataset, with θ_{RS} and θ_{LSM} anomalies rescaled to θ_{WA} (i.e., $\theta^*_{RS\to WA}$, $\theta^*_{LSM\to WA}$) using the optimal rescaling method described in Yilmaz and Crow (2014). Following this transformation, the unbiased-RMSE of the CSASM time series was estimated via TC:

$$\mathrm{ubRMSE}_{WA,T} = \sqrt{\sigma^2_{\varepsilon_{WA}}} \approx \sqrt{\sigma^2_{\theta^*_{WA}} - \frac{\mathrm{COV}(\theta^*_{WA}, \theta^*_{RS\to WA})\mathrm{COV}(\theta^*_{WA}, \theta^*_{LSM\to WA})}{\mathrm{COV}(\theta^*_{RS\to WA}, \theta^*_{LSM\to WA})}}. \tag{9}$$

See McColl et al. (2014) for background on the derivation of (9).

It is important to note that, even if these underlying assumptions are met, subtle differences exist between the TC-based estimates of $\mathrm{ubRMSE}_{WA,T}$ obtained from (9) and the corresponding $\mathrm{ubRMSE}_{WA,T}$ derived via (7). First, the two differ in that (7) reflects error after the removal of a single, long-term temporal mean, while (9) reflects error after the removal of 30-day moving averages following the recommendation in Chen et al. (2017). Values of ubRMSE calculated based on 30-day anomalies lack sensitivity to low-frequency and/or seasonally varying errors. To demonstrate the impact of excluding seasonality (and/or low-frequency variability) at varying temporal scales on TC-based $\mathrm{ubRMSE}_{WA,T}$ estimates, TC results based on soil moisture anomalies relative to different averaging window lengths are also presented and discussed below.

A second difference lies in the range of error sources considered by (7) and (9). Error statistics calculated using (7) depend entirely on the observed spatial variability derived from the scaling function and sampling network, whose actual spatial support may not perfectly align with the intended reference pixel. Likewise, random sampling errors are the only source of uncertainty captured. In contrast, TC theoretically solves the total random CSASM error statistic at the footprint scale (provided at least one of the other two datasets used in TC has a spatial resolution close to the footprint size; see discussion in Draper et al. 2013). Hence, $\mathrm{ubRMSE}_{WA,T}$ estimates acquired by TC in (9) reflects a broader range of error sources (including, e.g., random error due to instrumentation noise and nonrandom sampling effects) than $\mathrm{ubRMSE}_{WA,T}$ calculated via (7).

3) MOVING BLOCK BOOTSTRAP

Uncertainties in estimates of $\mathrm{ubRMSE}_{WA,T}$ acquired from both (7) and (9) were estimated by applying a moving-block bootstrap (MBB) resampling method. Since autocorrelation reduces the effective sample size of a time series, applying ordinary bootstrapping methods to autocorrelated variables (like soil moisture) will underestimate the probability that the sampled confidence interval contains the true statistical

property (Zwiers 1990; von Storch and Zwiers 1999). MBB was proposed as a better resampling method than ordinary bootstrap for estimating confidence intervals of correlation coefficients for autocorrelated time series (Mudelsee 2003, 2010). MBB was applied here to estimate the confidence intervals of ubRMSE because 1) the ubRMSE metric is mathematically related to correlation (Draper et al. 2013) and 2) despite the removal of temporal means (30-day or long-term) from the raw time series before calculation of ubRMSE, the anomaly time series still contains significant first-order autocorrelation. To obtain the confidence interval of the classical sampling theory-derived ubRMSE of CSASM, the methods in Mudelsee (2002) and Sherman et al. (1998) were followed to calculate the persistence time and block length parameters of MBB, which is outlined below in this section. The MBB procedure applied to triple collocation was adapted from Ólafsdóttir and Mudelsee's (2014) bivariate correlation problem, which is described in Chen et al. (2018).

MBB draws blocks of data from the original time series (i.e., soil moisture anomalies here) to form resamples that preserve the characteristic temporal dependence of the original dataset (Künsch 1989; Liu and Singh 1992). Block length was determined from the autocorrelation coefficient, which is calculated from the persistence time τ of the soil moisture anomaly time series [i.e., CSASM, or θ_{WA} in (7)]. Persistence time was estimated by minimizing the sum of squares:

$$S(\tau_{\theta_{\mathrm{WA}}}) = \sum_{i=2}^{n} \left\{ \theta_{\mathrm{WA},i} - \exp\left[-(t_i - t_{i-1})/\tau_{\theta_{\mathrm{WA}}} \right] \theta_{\mathrm{WA},i-1} \right\}^2, \tag{10}$$

where n is the length of the time series, $\theta_{\mathrm{WA},i}$ is the ith data point, and t_i is the linear time point (in unit of day) with potentially uneven intervals. The equivalent first-order autoregressive process [AR(1)] autocorrelation coefficient is given by $a = \exp(-d/\tau)$, where $d = (t_n - t_1)/(n-1)$ is the average time spacing. The autocorrelation coefficient was then bias-corrected to approximate the AR(1) process with an even time spacing:

$$a' = [a(n-1) + 1]/(n-4). \tag{11}$$

The optimal block length was then estimated as

$$l_{\mathrm{opt}} = \mathrm{NINT}\{[\sqrt{6}\, a'/(1-a'^2)]^{2/3} n^{1/3}\}, \tag{12}$$

where NINT denotes rounding to the nearest integer. Overlapping blocks of data with the length of l_{opt} were then randomly drawn with replacement to form a new sample of the original data length. The resampling procedure was repeated 1000 times in each grid pixel. Estimated 95% confidence intervals for ubRMSE were obtained from the 2.5th and 97.5th percentiles of the MBB-sampled distribution.

3. Data

a. CSASM at SMAP core validation sites

A set of CVSs, operated by independent Cal/Val Partners, containing calibrated in situ measurements were chosen by the SMAP mission to determine the quality of the SMAP data products (Colliander et al. 2017a). Table 1 lists the CVS that provide average soil moisture at the 33-km scale reference pixels to validate the SMAP L2_SM_P_E product (Chan et al. 2018). Locations of these CVS are shown in Fig. 1a along with examples of their internal soil moisture networks. At each site, scaling functions have been established and calibrated to upscale point-based observations within each CVS to the grid cell averages. Specifically, the SMAP Cal/Val team used a Voronoi diagram approach to determine the relative weighting between individual measurement stations within most CVS (Colliander et al. 2017a). The number of sensors N listed in the table corresponds to the subset of stations used for computing CSASM for the 33-km reference pixel. Additional stations located too far outside the 33-km pixel or having significant data gaps were excluded.

Temporary, denser in situ networks were installed for short-term soil moisture field experiments at four SMAP CVS: WG (23 June–27 October 2015, 30 temporary sensors; Colliander et al. 2017b), SF (18 May–17 August 2016, 20 temporary sensors; Colliander et al. 2019), LR (5 May–6 October 2017, 19 temporary sensors), and CM (17 May–18 August 2016, 34 sensors; Bhuiyan et al. 2018). These temporary networks, including both temporary and permanent soil moisture sensors, provided much denser spatial sampling, and thus more accurate assessments of spatial variability, within the CVS reference pixels—albeit for relatively short time periods.

b. Satellite soil moisture retrievals

1) SMAP LEVEL 2 ENHANCED SOIL MOISTURE

NASA's SMAP satellite observes global soil moisture with the onboard L-band (1.41 GHz) radiometer since January 2015 (Entekhabi et al. 2010). The SMAP Enhanced Passive Soil Moisture (L2_SM_P_E; Chan et al. 2018) is retrieved on the 9-km global EASE Grid 2.0 with a contributing spatial domain of 33 km × 33 km. SMAP L2_SM_P_E data (version 4, CRID T15570) used in the analysis were acquired from both ascending (1800 LST) and descending (0600 LST) orbits between 31 March 2015 and 28 February 2018.

TABLE 1. SMAP soil moisture core validation sites in this analysis. In the rightmost column, N is the number of sensors contributing to the CSASM in each corresponding 33-km reference pixel, and N_{eff} is the effective sample size determined by the sensor weights applied in the upscaling function (see section 2a).

	Code	PI	Location	Climate	IGBP land cover	Reference	N (N_{eff})
Walnut Gulch	WG	D. C. Goodrich	United States (Arizona)	Arid	Shrub open	Keefer et al. (2008)	18 (11)
Reynolds Creek	RC	M. Seyfried	United States (Idaho)	Arid	Grasslands	Seyfried et al. (2001)	7 (5)
TxSON	TX	T. Caldwell	United States (Texas)	Temperate	Grasslands	Caldwell et al. (2018)	28 (12)
Fort Cobb	FC	P. J. Starks	United States (Oklahoma)	Temperate	Grasslands		11 (7)
Little Washita	LW	P. J. Starks	United States (Oklahoma)	Temperate	Grasslands	Cosh et al. (2006)	12 (9)
South Fork	SF	M. H. Cosh/J. Prueger	United States (Iowa)	Cold	Croplands	Coopersmith et al. (2015)	19 (15)
Little River	LR	D. Bosch	United States (Georgia)	Temperate	Cropland/natural mosaic	Bosch et al. (2007)	19 (9)
Kenaston	KN	A. Berg	Canada	Cold	Croplands	Tetlock et al. (2019)	30 (12)
Carman	CM	H. McNairn	Canada	Cold	Croplands	McNairn et al. (2015)	20 (17)
Monte Buey	MB	M. Thibeault	Argentina	Arid	Croplands		12 (5)
REMEDHUS	RD	J. Martínez-Fernández	Spain	Temperate	Croplands	Martinez-Fernandez and Ceballos (2005)	15 (13)
Twente	TW	Z. Su	The Netherlands	Temperate	Cropland/natural mosaic	Dente et al. (2012)	5 (5)
HOBE	HB	K. Jensen	Denmark	Temperate	Cropland/natural mosaic	Bircher et al. (2012)	15 (13)
Mongolian Grasslands	MH	J. Asanuma	Mongolia	Cold	Grasslands	Wen et al. (2014)	5 (5)
Yanco	YC	J. Walker	Australia	Semiarid	Croplands/grasslands	Smith et al. (2012)	23 (8)

2) ASCAT SOIL MOISTURE

The European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) Advanced Scatterometer (ASCAT) sensor on board the *Meteorological Operational-B* (*MetOp-B*) satellite measures C-band (5.3 GHz) radar backscatter since September 2012, with 25–34 km spatial resolution and equatorial ascending and descending overpasses at 2130 and 0930 LST, respectively. The ASCAT level 2 soil moisture index product in a 12.5-km swath grid utilized here is based on the change-detection algorithm developed by Vienna University of Technology (Wagner et al. 1999; Naeimi et al. 2009). Here, grid cells were masked if probabilities of snow cover, frozen soil moisture, and estimated retrieval error are greater than 50%, and if rainfall probability is greater than 10%. The ASCAT soil moisture index data were converted to volumetric units ($m^3\,m^{-3}$) and resampled to the 33-km SMAP reference pixels by spatial averaging of 12.5-km grid cells that fell within a 33-km reference pixel whenever a minimum of four ASCAT grid cells were present.

3) SOIL MOISTURE OCEAN SALINITY

The Soil Moisture Ocean Salinity (SMOS) satellite was launched in 2009 by the European Space Agency (ESA; Kerr et al. 2001) and measures L-band microwave emission (1.400–1.427 GHz) with equatorial ascending and descending overpasses at 0600 and 1800 LST, respectively. The SMOS Level 2 v650_001 soil moisture data used in this analysis are acquired from both ascending and descending overpasses in an equal-area ISEA 4H9 15-km grid (Carr et al. 1997). Retrievals with Data Quality indeX (DQX) of greater than $0.07\,m^3\,m^{-3}$ or covered by snow or ice were removed. Remaining grid cells with valid data were resampled to the 33-km SMAP reference pixels using the same approach described above for ASCAT.

d. Land surface modeling soil moisture

1) SMAP NATURE RUN

The SMAP Nature Run, version 3 (NRv3 or NR), surface soil moisture field was generated with the NASA Catchment land surface model with an early version of

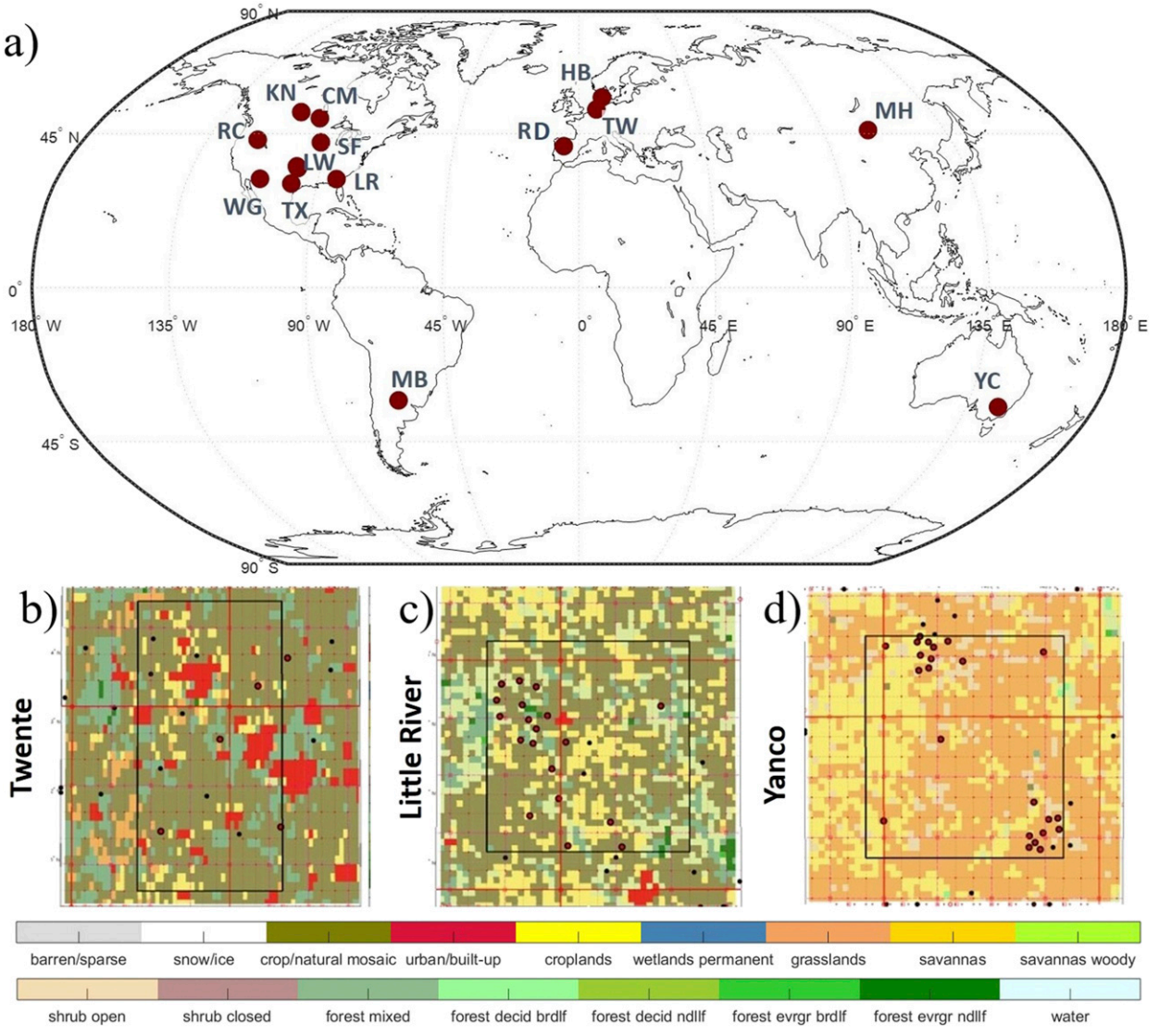

FIG. 1. (a) Location of SMAP soil moisture core validation sites for the L2_SM_P_E product. Examples of soil moisture sensor networks (red circles) are shown for the (b) Twente (TW), (c) Little River (LR), and (d) Yanco (YC) 33-km reference pixels (bounded by black boxes).

the SMAP level 4 surface and root zone soil moisture algorithm by the NASA Goddard Space Flight Center (GSFC) Global Modeling and Assimilation Office (Reichle et al. 2014). It was generated in a model-only configuration using a single ensemble member without perturbations or assimilation of SMAP brightness temperature observations. NRv3 is available at a 3-hourly interval on the 9-km EASEv2 grid. These values were aggregated to match 33-km SMAP CVS reference pixels via spatial averaging.

2) ECMWF

The land surface model soil moisture product used in the triple collocation analysis was the operational analysis layer-1 (0–7 cm) volumetric soil moisture field acquired from the European Centre for Medium-Range Weather Forecasts (ECMWF) Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land (H-TESSEL) land surface scheme (Balsamo et al. 2009). The operational soil moisture analysis product data are produced by ECMWF's Land Data Assimilation System via the assimilation of 2-m air temperature and relative humidity observations into the H-TESSEL land surface model (Drusch et al. 2009; de Rosnay et al. 2014). The ECMWF soil moisture analysis data are available at 0000, 0600, 1200, and 1800 UTC and on a N640 reduced Gaussian grid. ECMWF soil moisture data were resampled to the SMAP 33-km reference pixels using a nearest-neighbor approach.

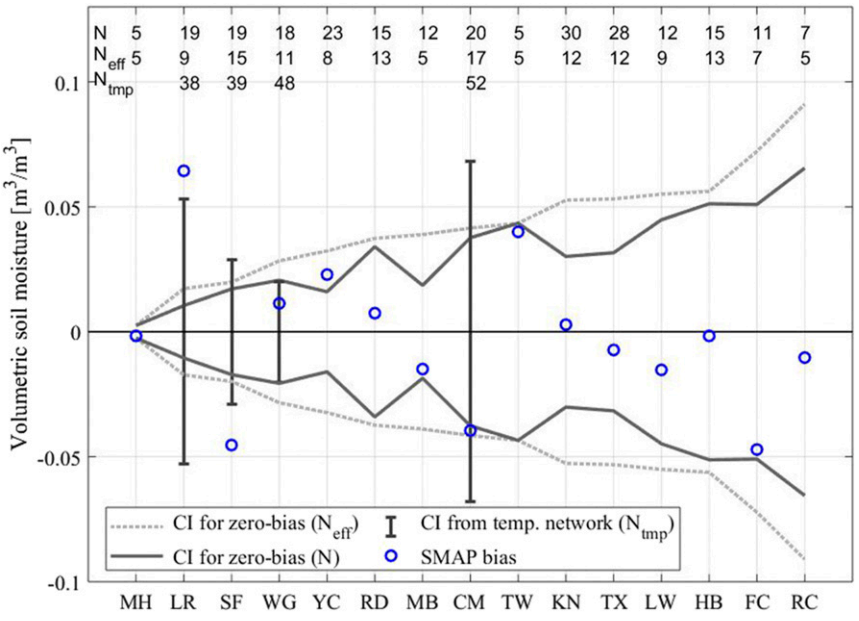

FIG. 2. The 95% confidence intervals (CI) for zero bias and observed SMAP L2SM_P_E soil moisture bias at SMAP CVS (see Table 1 for site abbreviations).

4. Results and discussion

a. Uncertainty bounds on time-averaged CSASM

As discussed in section 2a, the uncertainty of temporal-average CSASM $\overline{\theta_{WA}}$ were estimated using both effective (N_{eff}) and raw (N) sample sizes (Fig. 2). The 95% confidence intervals for θ_{WA} obtained with N_{eff} (dashed lines) range from 0.0025 $m^3\,m^{-3}$ (MH) to 0.091 $m^3\,m^{-3}$ (RC) and are, on average, 25.7% (standard deviation: 16.5%) of the long-term CSASM mean across all 15 CVS. The N-based confidence intervals (solid lines) are relatively tighter than N_{eff}-based numbers due to the higher standard error of the CSASM obtained through (3). Nevertheless, even N-based sampling uncertainty was found to be large relative to the time-averaged CSASM values, which translates into substantial relative uncertainty in SMAP bias estimates. In fact, at most sites, even after the collection of 3+ years of SMAP data, the observed SMAP biases (for baseline SMAP L2_SM_P_E retrievals) fall within the 95% confidence interval, suggesting these biases are not statistically significant and can therefore be reasonably attributed to spatial sampling limitations within each individual core site.

Note that the 95% CI was estimated via the Student's t-test procedure that assumes soil moisture is randomly sampled in space within the reference pixel. In reality, networks are constructed from permanent soil moisture measurement locations, some of which chosen for reasons of practicality (e.g., accessibility and compatibility with farm operations). This results in either some degree of spatial clustering—for example, the YC network (Fig. 1d)—or preferential sampling of certain land cover types—for example, croplands by the LR network (Fig. 1c). Such clustering, in turn, can lead to a low bias in sampled spatial variability within the true soil moisture field. Therefore, the confidence intervals in Fig. 2 (even when calculated with N_{eff}) are still likely conservative estimates of the true uncertainty in the CSASM.

In addition to surface soil moisture observations obtained from permanent sampling locations, measurements were also collected from additional temporary sensors installed within the WG, SF, LR, and CM core-site domains during limited field campaign periods (see section 3a above). Due to their large number of sampling sites, these expanded (temporary plus permanent station) networks provide a more robust estimate of the actual spatial variability within each CVS reference pixel. This updated information was also applied with the original sample size (i.e., the number of permanent stations) in (5) to simulate the situation where accurate spatial variance could be observed solely from the permanent network. The results (vertical error bars in Fig. 2) suggest that confidence intervals estimated from permanent network observations are indeed often underestimated (see, e.g., the SF, LR, and CM sites). In particular, permanent sensors at LR are all installed on agricultural land, while a large fraction of the reference pixel (55%) occupied by forests is not sampled (Fig. 1c). The smaller CSASM uncertainty observed from the LR

permanent network is therefore an underestimate of true spatial variability within the entire reference pixel. These observations suggest that the network configurations at some CVS should be adjusted to better capture the subpixel soil moisture average and variance. This shortcoming has been noted previously by high-resolution hydrologic modeling at the LR site (Chaney et al. 2015). This bias further underscores the spatial sampling challenges faced when attempting to estimate bias in SMAP surface soil moisture retrievals within areas of mixed land cover.

One direct way to reduce the bias uncertainty originating from the CSASM is (naturally) through increased network sample density. As suggested by Famiglietti et al. (2008), the effective sample size required to achieve a specific uncertainty level can be calculated by inverting (3) and (5). For example, to achieve a bias uncertainty of under $0.03\,m^3\,m^{-3}$ with the currently sampled spatial variance, some of the SMAP CVS (i.e., TW, RC, FC, LW, RD, HB, and CM) would need to increase their current network density by about a factor of 2 (Table 2). Reasons behind the relatively low sampling power of SMAP CVS (with respect to bias) are discussed further in section 5 below.

As an alternative to the relatively high-cost option of increasing network density, improving scaling functions for CSASM is a potential means of better reflecting the contribution from underrepresented land cover or soil types within the reference pixels. To capture undersampled forested sites and better approximate the true average soil moisture of the 33-km reference pixel at the Little River (LR) CVS, an updated scaling function was derived by linear transformation of the permanent network-based CSASM using observations from the temporary network. While the LR CSASM time series used in this analysis is based on an old (not updated) scaling function, the updated scaling function has been applied to CSASM calculations used in current SMAP routine validation activities and assessment reports (e.g., Chan et al. 2018; Jackson et al. 2018). Updated CSASM based on the improved scaling function results in a significantly reduced SMAP L2_SM_P_E retrieval bias ($\sim 0.06\,m^3\,m^{-3}$, compared to $\sim 0.12\,m^3\,m^{-3}$ using older scaling function) at the LR CVS.

b. Time-varying error in CSASM

In addition to uncertainty bounds on time-average CSASM values, section 2 described two methods to estimate the ubRMSE of CSASM time series ($ubRMSE_{WA,T}$: 1) a classical sampling theory as summarized in (7) and 2) TC as summarized in (9). Figure 3 compares $ubRMSE_{WA,T}$ obtained from both methods at SMAP CVS with associated 95% confidence intervals.

TABLE 2. Estimated effective sampling sizes for SMAP CVS required for 95% confidence intervals of zero bias to fall below $0.03\,m^3\,m^{-3}$. SMAP CVS with uncertainty bounds (already) below $0.03\,m^3\,m^{-3}$ are excluded from the table.

	Current N	Current N_{eff}	N_{eff} for CI $\leq 0.03\,m^3\,m^{-3}$
TW	5	5	8
RC	7	5	15
FC	11	7	16
LW	12	9	17
RD	15	13	17
HB	15	13	34
CM	20	17	26
TX	28	12	14
KN	30	12	13

TC analysis was performed with six different data triplets, each containing CSASM and one of the following remote sensing–LSM soil moisture data pairs: SMAP–ECMWF, SMAP–NR, SMOS–ECMWF, SMOS–NR, ASCAT–ECMWF, and ASCAT–NR. Each of these data triplets should theoretically yield the same TC-based estimate of $ubRMSE_{WA,T}$. TC estimates shown in Fig. 3 were averaged from analyses from six data triplets (without resampling). The 95% confidence intervals for TC-derived $ubRMSE_{WA,T}$ were generated from a 6000-member moving-block bootstrap resampling analysis (i.e., 1000 samples from each triplet). Confidence intervals for classical sampling theory-derived $ubRMSE_{WA,T}$ were generated from a 1000-member moving-block bootstrap resampling analysis. To isolate the impact of random error at different time scales, $ubRMSE_{WA,T}$ were calculated for the case of anomalies calculation against both a single, long-term mean and a 30-day moving-window average.

Overall, both methods (sampling theory and TC) estimate $ubRMSE_{WA,T}$ values that (generally) fall within the range from 0.01 to $0.02\,m^3\,m^{-3}$. While this level of error is well below the $0.04\,m^3\,m^{-3}$ ubRMSE target accuracy for both SMAP and SMOS, users of CSASM should be aware of nontrivial CSASM ubRMSE that will impact remote sensing (and/or model) soil moisture ubRMSE assessments calculated from direct CSASM comparisons. When seasonal signals are removed from the observations, so is much of the temporally persistent spatial variation within the pixel. This explains the smaller ubRMSE values in Fig. 3b compared to Fig. 3a.

Both cases examined in Fig. 3 indicated that TC ubRMSE estimates are generally larger than comparable estimates based on sampling theory. This difference is pronounced in the case of anomalies relative to 30-day moving averages (Fig. 3b). However, there is a clear reduction in scatterplot correlation (from a correlation

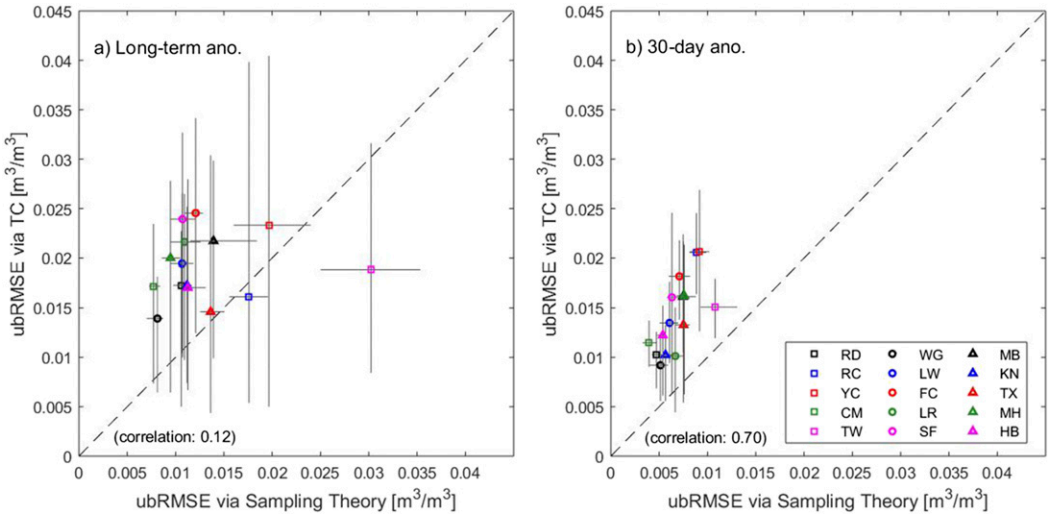

FIG. 3. Comparison of CSASM ubRMSE ($\text{ubRMSE}_{\text{WA},T}$) estimated from both TC (y axis) and classical sampling theory (x axis) based on the removal of (a) a fixed long-term average and (b) 30-day moving-window averages. TC results were averaged from analyses based on six different data triplets (see section 4b) without bootstrap resampling. Error bars (solid gray lines) for sampling theory results reflect 95% confidence intervals obtained from a 1000-member moving-block bootstrap (MBB) sampling. Confidence intervals for TC results were obtained from joint distribution of $\text{ubRMSE}_{\text{WA},T}$ sampled via separate application of MBB to TC analyses with six different soil moisture data triplets (1000 samples each).

of 0.70 in Fig. 3b to 0.12 in Fig. 3a) for the case of anomalies relative to a long-term mean (Fig. 3a), which suggests a partial breakdown in the consistency of the two $\text{ubRMSE}_{\text{WA},T}$ estimates. This lack of correlation in Fig. 3a may result, in part, from the very large sampling error bars associated with the TC $\text{ubRMSE}_{\text{WA},T}$ estimates. See section 5 below for further discussion of these results.

A notable outlier in Fig. 3a is the Twente (TW) CVS site, where $\text{ubRMSE}_{\text{WA},T}$ estimates acquired from sampling theory are significantly higher than from based on TC. A closer look of the TW site was undertaken by considering all available observations within the TW reference pixel (i.e., 12 active soil moisture stations including the 5 currently used to obtain the CSASM via arithmetic averaging). Classical sampling theory was also applied to estimate the ubRMSE of a new CSASM time series calculated as the arithmetic average of all twelve stations. These results (not shown) suggest that 1) soil moisture observed at different locations within the reference pixel show markedly different seasonality and dynamic range, 2) the current CSASM network (five stations) at TW is representative of the spatial variability of the reference pixel, and 3) the estimated $\text{ubRMSE}_{\text{WA},T}$ of $\sim 0.03\,\text{m}^3\,\text{m}^{-3}$ (with removal of long-term mean) is relatively stable and reliable.

These observations are consistent with Dente et al. (2012), which identified land cover and groundwater table height as the main drivers of soil moisture spatial variability at the Twente site. Four of the five stations used in the Twente CSASM calculates are located on grasslands. The one station located in a corn field is found to be drier than the grass stations in spring and summer (Dente et al. 2012). One of the grassland sites has the most variable water table height (lowest depth >120 cm, highest depth at 25–40 cm) that leads to the large dynamic range and disparity between its time series and other sites (and CSASM as well). The large spatial variability in surface soil moisture at the TW site is, therefore, responsible for the relatively high estimates of $\text{ubRMSE}_{\text{WA},T}$ obtained from classical sampling theory. Conversely, lower values of $\text{ubRMSE}_{\text{WA},T}$ ($0.019\,\text{m}^3\,\text{m}^{-3}$) estimated by TC are greatly affected by the extremely low ubRMSE value provided by the CSASM–SMOS–NR triplet (see Fig. 5), which is outside the 95% confidence interval generated from all bootstrapping samples of the other six triplets. Ignoring this outlier, the average TC-based $\text{ubRMSE}_{\text{WA},T}$ estimate increases to $0.024\,\text{m}^3\,\text{m}^{-3}$—which is much closer to the $\text{ubRMSE}_{\text{WA},T}$ estimate obtained from sampling theory. In addition, as discussed below, TC results were generally less robust when only the long-term mean (versus the seasonally varying climatology) was removed from the raw time series, which likely contributes to the TW outlier in Fig. 3a.

5. Discussion

A striking result from Fig. 2 is that, even after 3+ years of data collection, the CSASM time series still lacks

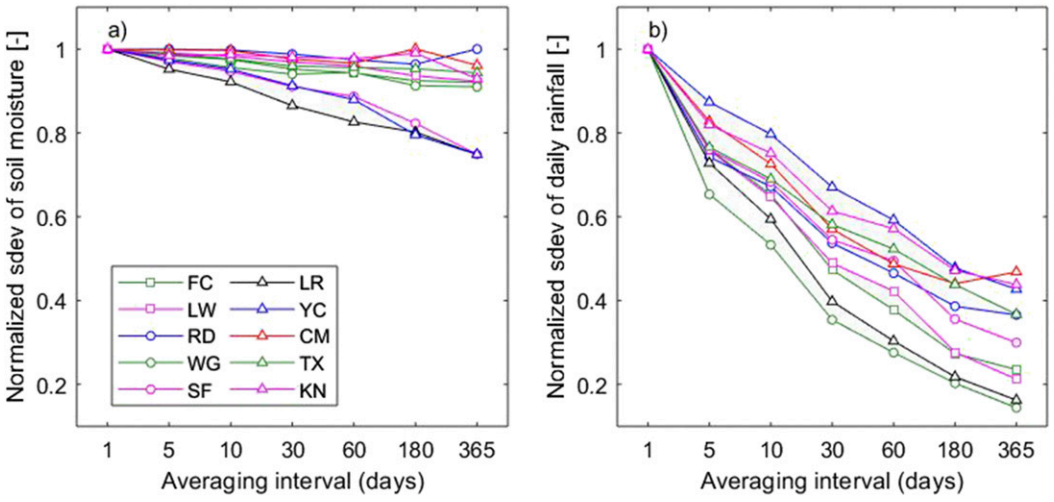

FIG. 4. (a) Normalized spatial standard deviation of temporally averaged soil moisture (within each CVS) for various time-averaging intervals. (b) As in (a), but for rainfall accumulations. Each line represents a different SMAP CVS reference pixel. Mean standard deviations for each time interval are normalized by 1-day standard deviation values.

sufficient sampling power to resolve long-term SMAP biases smaller than $0.05\,m^3\,m^{-3}$ at approximately half of the SMAP core sites. This lack of sampling power has important consequences for the design of future satellite validation strategies. To understand the reasons underlying such results, it is useful to consider the differences between spatial variability expressed in soil moisture versus precipitation data fields. Footprint-scale soil moisture variability is predominantly controlled by precipitation in wet conditions and by soil hydraulic properties and vegetation water content during interstorm periods (Kim and Barros 2002; Oldak et al. 2002). However, when averaged over long periods, the spatial variability of soil moisture and rainfall across the footprint decreases at markedly different rates. Figure 4 compares the changes of spatial standard deviation of soil moisture and precipitation temporal averages for increasing time-averaging intervals for 10 reference pixels with concurrent surface soil moisture and precipitation observations (i.e., the FC, LW, RD, WG, SF, LR, YC, CM, TX, and KN CVS, permanent stations only). The spatial standard deviations associated with each averaging interval are normalized by the 1-day standard deviation values.

Figure 4 illustrates the contrasting behavior of spatial variability in soil moisture and rainfall accumulation. While soil moisture spatial variability is relatively persistent over longer temporal averaging periods, spatial variability in rainfall accumulation decreases sharply over time. This observation is consistent with scale differences in the known drivers of soil moisture and rainfall. Spatial patterns of rainfall are mainly affected by (spatially and temporally variable) mesoscale atmospheric conditions, while soil moisture spatial variability is strongly linked to static distributions of soil texture and land cover that control the evapotranspiration rate at finer scales. As a result, soil moisture spatial errors are more autocorrelated in time, and sampling approaches that attempt to reduce spatial sampling uncertainty via the use of expanded temporal sampling are much less effective for soil moisture than for rainfall. Put another way, Fig. 4 reflects that spatial patterns in rainfall accumulation vary widely from event to event (and thus cancel out in a mean sense when averaged over long periods) while patterns in soil moisture variability persist over time (and are therefore resistant to removal via simple temporal averaging). Since the bias confidence intervals calculated from (5) are based on the spatial variability present within each CVS after temporal averaging, this persistence leads directly to the larger soil moisture uncertainty bounds seen in Fig. 2. Therefore, in contrast to rainfall, specific aspects of soil moisture time–space variability make it particularly difficult to accurately sample soil moisture biases via comparison with the CSASM.

The generalization of soil moisture time–space dynamics (i.e., dry stays dry and wet stays wet) underlying this result is obviously a simplification of the complex time–space variability in soil moisture fields [see, e.g., Molero et al. (2018) for a more complete analysis]. However, it is consistent with the assumptions underlying the application of temporal stability approaches for upscaling point-scale soil moisture (Cosh et al. 2006, 2008), as well as a study that showed a large fraction of

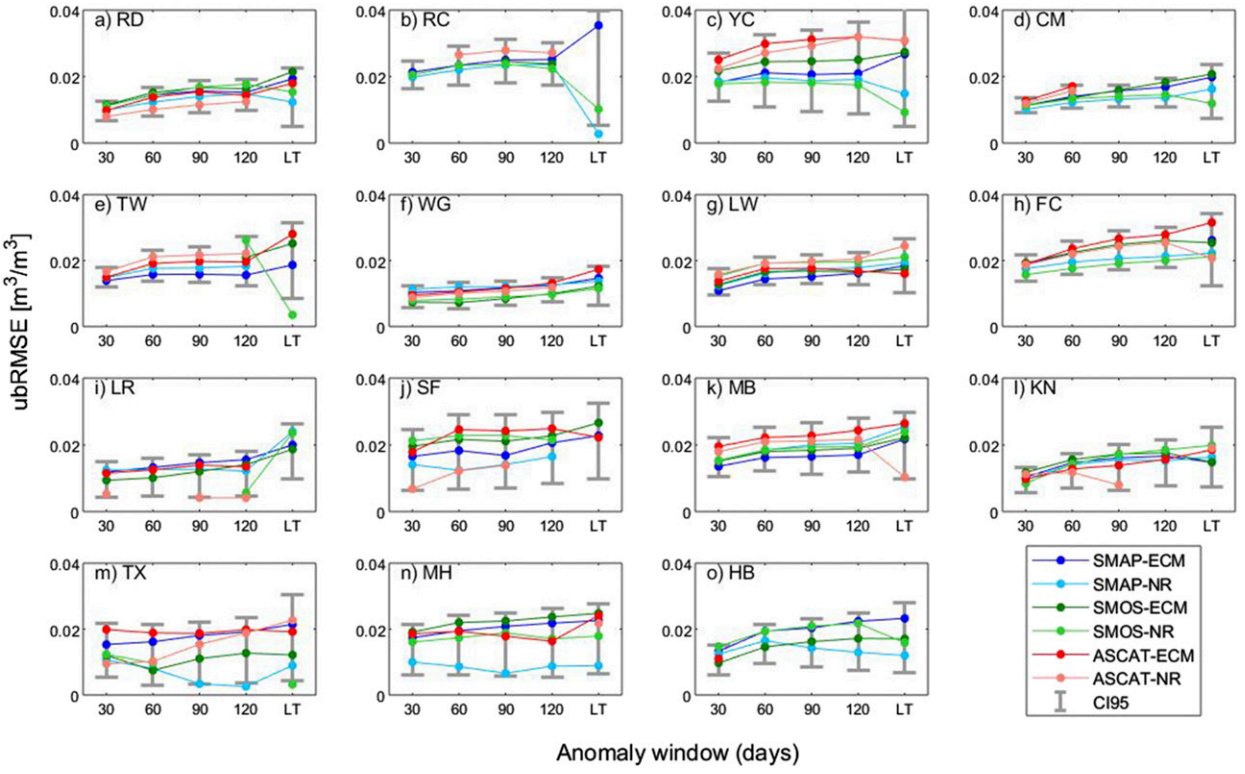

FIG. 5. Estimated ubRMSE for the CSASM temporal anomalies ($\text{ubRMSE}_{\text{WA},T}$) acquired from TC using six different satellite–LSM combinations (combined with CSASM to form a data triplet) as a function of various anomaly window lengths. "LT" refers to anomalies relative to a single, long-term mean value. Plotted 95% confidence intervals (CI95) are calculated using the same procedure as those in Fig. 3.

soil moisture time–space variability can be explained using a small set of empirical orthogonal functions (Joshi and Mohanty 2010). The importance of temporally fixed land surface properties for generating soil moisture spatial variability is also highlighted by a recent study in Oklahoma, United States, which revealed a higher correlation between mesoscale (1–100 km) soil moisture spatial patterns and sand content than that sampled against antecedent precipitation index (Dong and Ochsner 2018).

Two key points also stand out in Fig. 3, regarding estimated CSASM ubRMSE ($\text{ubRMSE}_{\text{WA},T}$). First, as noted above, TC-based $\text{ubRMSE}_{\text{WA},T}$ estimates are clearly biased high relative to comparable sampling theory results. This bias is expected since TC-based $\text{ubRMSE}_{\text{WA},T}$ includes other sources of error—in addition to the random sampling error (the only error source captured by sampling theory). In addition, nonrandom sampling causes the $\text{ubRMSE}_{\text{WA},T}$ estimates from classical sampling to be biased low. The only apparent exception to this trend is the TW site in Fig. 3a, which can be explained by the extremely low $\text{ubRMSE}_{\text{WA},T}$ estimate provided by one of the potential TC triplets (see discussion in section 4).

Second, the correspondence between TC and classical sampling cases is much weaker for the case of anomalies calculated against a long-term mean (Fig. 3a). This is likely due to a breakdown in the reliability of TC estimates in the presence of significant soil moisture seasonality. As discussed earlier, TC estimates include additional error sources not considered in the sampling theory-based ubRMSE [section 2b(2)]. However, nonstationary components (e.g., seasonality) were removed from the datasets prior to TC analysis, and therefore errors beyond the temporal scale of normalizing were filtered from the resulting ubRMSE. Figure 5 directly examines this issue by showing TC-based $\text{ubRMSE}_{\text{WA},T}$ estimates acquired from the application of different temporal normalization lengths to calculate soil moisture anomalies. At most CVS, higher $\text{ubRMSE}_{\text{WA},T}$ estimates are indeed associated with larger normalizing window lengths.

In addition, if all TC assumptions are met, all six data triplets (represented in Fig. 5 with different colored lines) should produce the same $\text{ubRMSE}_{\text{WA},T}$ estimate for a certain anomaly treatment. However, progressively larger discrepancies between triplets are observed as the normalizing window was made longer—indicating a

deterioration in TC stability due to violation of TC assumptions when low-frequency variability is not filtered. This tendency is most obvious at the RD, TW, WG, LW, FC, LR, and MB CVS sites. Since TC results are the most robust when variations between the six lines are minimized, ubRMSE calculated using anomalies against a single, long-term average is, in general, the least-reliable TC approach. This effect causes the reduced consistency between TC and sampling theory results demonstrated in Fig. 3a.

Therefore, while the TC error measures can be considered more complete than those based on sampling theory (since they consider errors sources arising from a wider variety of sources), their reliability is limited to cases where low-frequency soil moisture variability has been masked or is negligible. This is a serious limitation given the typically large seasonal variability present in soil moisture time series and the need to accurately represent this seasonality for many applications.

6. Summary

Reflecting current practices in the validation of satellite-based surface soil moisture products, SMAP soil moisture retrievals have been validated via direct comparison of core site average soil moisture (CSASM) obtained via the weighted averaging of point-scale measurements from ground networks established within each core validation site. Here, errors in the CSASM arising from limited spatial sampling within these networks are examined. Accurate statistical representations of these errors are highly useful when interpreting validation metrics acquired via comparisons between SMAP soil moisture retrievals and CSASM time series.

Sampling error in estimates of temporally average CSASM are found to be relatively large, translating into large uncertainty bounds for SMAP bias estimates obtained from the CSASM comparisons. At most CVS, observed SMAP bias falls within the 95% confidence intervals of zero bias and therefore is statistically insignificant (i.e., reasonably attributable to spatial sampling limitations within each CVS). In addition, observations from temporary networks with better sampling density suggest that spatial variability is generally undersampled by the CVS permanent networks. This implies that classical sampling theory is, if anything, understating the magnitude of the bias sampling uncertainty. Moreover, temporal persistence in soil moisture spatial variability suggests that such spatial uncertainty in CSASM cannot be removed even after temporal averaging over a long period (Fig. 4). The magnitude of this uncertainty must be considered when attempting to use CSASM values to estimate biases in coarse-scale satellite or modeled soil moisture products.

In general, the reliability of the currently available CSASM as a reference to estimate bias of a soil moisture product range from 0.01 to 0.09 $m^3 m^{-3}$ depending on the geophysical characteristics of the site and layout of the network. This uncertainty can be reduced by better characterizing the soil moisture spatial variability of the reference pixel through measures such as 1) increasing the network density, 2) calibrating the upscaling function (against a more reliable ground truth obtained through field campaigns of intensive sampling; e.g., in the LR case discussed earlier), and 3) optimizing sampling locations. For example, an estimated average of about seven additional in situ probes are required to reduce the current bias uncertainty to below 0.03 $m^3 m^{-3}$ across seven SMAP CVS listed in Table 2.

Unbiased-RMSE of the CSASM is estimated via two methods—classical sampling theory and triple collocation. As expected, due to their sensitivity to a wider variety of error sources, TC-derived ubRMSE estimates were generally larger than those estimated from sampling theory. In addition, correlation between the two methods was degraded when anomalies are calculated relative to a long-term mean (rather than a 30-day moving-average climatology). This is due to reduced robustness in TC estimates acquired when soil moisture seasonality is present (Fig. 5). Despite these differences, both methods estimated the ubRMSE of CSASM at most CVS to be 0.01–0.02 $m^3 m^{-3}$—well below the 0.04 $m^3 m^{-3}$ target accuracy for SMAP and SMOS missions. This modest CSASM ubRMSE obtained at 33-km reference pixels is the result of the relatively coarse spatial scale of dynamic forcings (i.e., mesoscale precipitation events) principally driving the temporal dynamics of soil moisture fields.

Overall, results point out a contrast in the ability of the CSASM to be used for ubRMSE versus bias assessments. Given that the ubRMSE of CSASM is generally in the range 0.01–0.02 $m^3 m^{-3}$, and assuming an ubRMSE (against unknown truth) of 0.04 $m^3 m^{-3}$ for SMAP retrievals, estimates of SMAP ubRMSE obtained via direct comparisons against the CSASM will be on the order of 0.041–0.045 $m^3 m^{-3}$—following the root-sum-square adding procedure described in (6). While this inflation of ubRMSE is not trivial, it does not seriously degrade the ability of the CSASM to assess the temporal precision of SMAP soil moisture retrievals.

In contrast, the impact of CSASM spatial sampling errors on bias are more profound. The temporally stable nature of the soil moisture spatial variability leads to highly autocorrelated CSASM spatial sampling errors that cannot be effectively eliminated by ergodic sampling approaches (which attempt to remedy spatial sampling shortcoming by first averaging over long time periods). Fully addressing this issue will likely require

either enhanced short-term spatial sampling to improve the upscaling functions or a permanent increase of the spatial density of CVS ground sampling networks (e.g., see Table 2). Therefore, the spatial sampling requirements for soil moisture *bias* assessments are much more stringent than those required for temporal *precision* assessments (via metrics like temporal correlation or ubRMSE). This tendency should be considered in future soil moisture calibration/validation efforts.

Acknowledgments. This study is supported by Dr. Wade Crow's membership on the NASA Soil Moisture Active Passive mission. A partial contribution to this work was made at Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. USDA is an equal opportunity provider and employer. The REMEDHUS network data was provided by the University of Salamanca team supported by the Spanish Ministry of Economy and Competitiveness with the project ESP2017- 89463-C3-3-R, and the European Regional Development Fund (ERDF). The Twente network was supported by the Netherlands Organization for Scientific Research (Project No. ALW-GO/14-29) and the ESA MOST Dragon IV program [Monitoring Water and Energy Cycles at Climate Scale in the Third Pole Environment (CLIMATE-TPE)]. The authors are grateful to ESA for offering the ELBARA-III radiometer (ZS). Erica Tetlock from Environment and Climate Change Canada and Tracy Rowlandson are acknowledged for work with the Kenaston Network with funding from Environment and Climate Change Canada, the Canadian Space Agency, and the Natural Sciences and Engineering Research Council of Canada. The VILLUM Foundation provided funding for the HOBE observatory.

REFERENCES

Balsamo, G., P. Viterbo, A. C. M. Beljaars, B. J. J. M. van den Hurk, M. Hirschi, A. K. Betts, and K. Scipal, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the ECMWF-IFS. *J. Hydrometeor.*, **10**, 623–643, https://doi.org/10.1175/2008JHM1068.1.

Bhuiyan, H. A., and Coauthors, 2018: Assessing SMAP soil moisture scaling and retrieval in the Carman (Canada) study site. *Vadose Zone J.*, **17**, 180132, https://doi.org/10.2136/vzj2018.07.0132.

Bircher, S., N. Skou, K. H. Jensen, J. P. Walker, and L. Rasmussen, 2012: A soil moisture and temperature network for SMOS validation in Western Denmark. *Hydrol. Earth Syst. Sci.*, **16**, 1445–1463, https://doi.org/10.5194/hess-16-1445-2012.

Bosch, D. D., V. Lakshmi, T. J. Jackson, M. Choi, and J. M. Jacobs, 2006: Large scale measurements of soil moisture for validation of remotely sensed data: Georgia soil moisture experiment of 2003. *J. Hydrol.*, **323**, 120–137, https://doi.org/10.1016/j.jhydrol.2005.08.024.

——, J. M. Sheridan, R. R. Lowrance, R. K. Hubbard, T. C. Strickland, G. W. Feyereisen, and D. G. Sullivan, 2007: Little River experimental watershed database. *Water Resour. Res.*, **43**, W09470, https://doi.org/10.1029/2006WR005844.

Brocca, L., F. Melone, T. Moramarco, and R. Morbidelli, 2009: Soil moisture temporal stability over experimental areas in Central Italy. *Geoderma*, **148**, 364–374, https://doi.org/10.1016/j.geoderma.2008.11.004.

Caldwell, T. G., T. Bongiovanni, M. H. Cosh, C. Halley, and M. H. Young, 2018: Field and laboratory evaluation of the CS655 soil water content sensor. *Vadose Zone J.*, **17**, 170214, https://doi.org/10.2136/vzj2017.12.0214.

Carr, D., R. Kahn, K. Sahr, and T. Olsen, 1997: ISEA discrete global grids. *Statistical Computing and Statistical Graphics Newsletter*, Vol. 8, No. 2/3, American Statistical Association, Alexandria, VA, 31–39.

Chan, S. K., and Coauthors, 2018: Development and assessment of the SMAP enhanced passive soil moisture product. *Remote Sens. Environ.*, **204**, 931–941, https://doi.org/10.1016/j.rse.2017.08.025.

Chaney, N. W., J. K. Roundy, J. E. Herrera-Estrada, and E. F. Wood, 2015: High-resolution modeling of the spatial heterogeneity of soil moisture: Applications in network design. *Water Resour. Res.*, **51**, 619–638, https://doi.org/10.1002/2013WR014964.

Chen, F., and Coauthors, 2017: Application of triple collocation in ground-based validation of Soil Moisture Active/Passive (SMAP) Level 2 data products. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **10**, 489–502, https://doi.org/10.1109/JSTARS.2016.2569998.

——, W. T. Crow, R. Bindlish, A. Colliander, M. Burgin, J. Asanuma, and K. Aida, 2018: Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. *Remote Sens. Environ.*, **214**, 1–13, https://doi.org/10.1016/j.rse.2018.05.008.

Choi, M., J. M. Jacobs, and M. H. Cosh, 2007: Scaled spatial variability of soil moisture fields. *Geophys. Res. Lett.*, **34**, L01401, https://doi.org/10.1029/2006GL028247.

Clewley, D., and Coauthors, 2017: A method for upscaling in situ soil moisture measurements to satellite footprint scale using random forests. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **10**, 2663–2673, https://doi.org/10.1109/JSTARS.2017.2690220.

Colliander, A., and Coauthors, 2017a: Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.*, **191**, 215–223, https://doi.org/10.1016/j.rse.2017.01.021.

——, and Coauthors, 2017b: Validation and scaling of soil moisture in a semi-arid environment: SMAP Validation Experiment 2015 (SMAPVEX15). *Remote Sens. Environ.*, **196**, 101–112, https://doi.org/10.1016/j.rse.2017.04.022.

——, and Coauthors, 2019: Comparison of High-resolution airborne soil moisture retrievals to SMAP soil moisture during the SMAP validation experiment 2016 (SMAPVEX16). *Remote Sens. Environ.*, **227**, 137–150, https://doi.org/10.1016/j.rse.2019.04.004.

Coopersmith, E. J., M. H. Cosh, W. A. Petersen, J. Prueger, and J. J. Niemeier, 2015: Soil moisture model calibration and validation: An ARS watershed on the South Fork Iowa River. *J. Hydrometeor.*, **16**, 1087–1101, https://doi.org/10.1175/JHM-D-14-0145.1.

Cosh, M. H., T. J. Jackson, P. J. Starks, and G. C. 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, https://doi.org/10.1016/j.jhydrol.2005.08.020.

——, ——, S. Moran, and R. Bindlish, 2008: Temporal persistence and stability of surface soil moisture in a semi-arid watershed.

Remote Sens. Environ., **112**, 304–313, https://doi.org/10.1016/j.rse.2007.07.001.

——, ——, C. Smith, B. Toth, and A. A. Berg, 2013: Validating the BERMS in situ soil water content data record with a large scale temporary network. *Vadose Zone J.*, **12**, https://doi.org/10.2136/vzj2012.0151.

Crow, W. T., D. Ryu, and J. S. Famiglietti, 2005: Upscaling of field-scale soil moisture measurements using distributed land surface modeling. *Adv. Water Resour.*, **28**, 1–14, https://doi.org/10.1016/j.advwatres.2004.10.004.

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

de Rosnay, P., G. Balsamo, C. Albergel, J. Muñoz-Sabater, and L. Isaksen, 2014: Initialisation of land surface variables for numerical weather prediction. *Surv. Geophys.*, **35**, 607–621, https://doi.org/10.1007/s10712-012-9207-x.

Dente, L., Z. Su, and J. Wen, 2012: Validation of SMOS soil moisture products over the Maqu and Twente regions. *Sensors*, **12**, 9965–9986, https://doi.org/10.3390/s120809965.

Dong, J., and T. E. Ochsner, 2018: Soil texture often exerts a stronger influence than precipitation on mesoscale soil moisture patterns. *Water Resour. Res.*, **54**, 2199–2211, https://doi.org/10.1002/2017WR021692.

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

Drusch, M., P. de Rosnay, G. Balsamo, E. Andersson, P. Bougeault, and P. Viterbo, 2009: Towards a Kalman filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System. *Geophys. Res. Lett.*, **36**, L10401, https://doi.org/10.1029/2009GL037716.

Entekhabi, D., and Coauthors, 2010: The Soil Moisture Active Passive (SMAP) mission. *Proc. IEEE*, **98**, 704–716, https://doi.org/10.1109/JPROC.2010.2043918.

Faber, N. M., 1999: Estimating the uncertainty in estimates of root mean square error of prediction: Application to determining the size of an adequate test set in multivariate calibration. *Chemom. Intell. Lab. Syst.*, **49**, 79–89, https://doi.org/10.1016/S0169-7439(99)00027-1.

Famiglietti, J. S., and Coauthors, 1999: Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. *Water Resour. Res.*, **35**, 1839–1851, https://doi.org/10.1029/1999WR900047.

——, D. Ryu, A. A. Berg, M. Rodell, and T. J. Jackson, 2008: Field observations of soil moisture variability across scales. *Water Resour. Res.*, **44**, W01423, https://doi.org/10.1029/2006WR005804.

Gruber, A., C.-H. Su, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner, 2016: Recent advances in (soil moisture) triple collocation analysis. *Int. J. Appl. Earth Obs. Geoinf.*, **45B**, 200–211, https://doi.org/10.1016/j.jag.2015.09.002.

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

——, and Coauthors, 2012: Validation of Soil Moisture and Ocean Salinity (SOMS) soil moisture over watershed networks in the U.S. *IEEE Trans. Geosci. Remote Sens.*, **50**, 1530–1543, https://doi.org/10.1109/TGRS.2011.2168533.

——, and Coauthors, 2018: Soil Moisture Active Passive (SMAP) Project: Calibration and Validation for the L2/3_SM_P Version 5 and L2/3_SM_P_E Version 2 Data Products, Doc. JPL D-56297, Jet Propulsion Laboratory, 44 pp., https://nsidc.org/sites/nsidc.org/files/technical-references/L2SMPE_Asmt_Rpt_EOPM_v5c_Jun2018.pdf.

Joshi, C., and B. P. Mohanty, 2010: Physical controls of near-surface soil moisture across varying spatial scales in an agricultural landscape during SMEX02. *Water Resour. Res.*, **46**, W12503, https://doi.org/10.1029/2010WR009152.

Keefer, T. O., M. S. Moran, and G. B. Paige, 2008: Long-term meteorological and soil hydrology database, Walnut Gulch Experimental Watershed, Arizona, United States. *Water Resour. Res.*, **44**, W05S07, https://doi.org/10.1029/2006WR005702.

Kerr, Y. H., P. Waldteufel, J. P. Wigneron, J. M. Martinuzzi, J. Font, and M. Berger, 2001: Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1729–1735, https://doi.org/10.1109/36.942551.

——, and Coauthors, 2016: Overview of SMOS performance in terms of global soil moisture monitoring after six years in operation. *Remote Sens. Environ.*, **180**, 40–63, https://doi.org/10.1016/j.rse.2016.02.042.

Kim, G., and A. P. Barros, 2002: Space–time characterization of soil moisture from passive microwave remotely sensed imagery and ancillary data. *Remote Sens. Environ.*, **81**, 393–403, https://doi.org/10.1016/S0034-4257(02)00014-7.

Kish, L., 1965: *Survey Sampling*. John Wiley and Sons, 643 pp.

Künsch, H., 1989: The jackknife and the bootstrap for general stationary observations. *Ann. Stat.*, **17**, 1217–1241, https://doi.org/10.1214/aos/1176347265.

Liu, R., and K. Singh, 1992: Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the Limits of Bootstrap*, R. Lepage and L. Billard, Eds., John Wiley and Sons, 225–248.

Martinez-Fernandez, J., and A. Ceballos, 2005: Mean soil moisture estimation using temporal stability analysis. *J. Hydrol.*, **312**, 28–38, https://doi.org/10.1016/j.jhydrol.2005.02.007.

McColl, K. A., J. Vogelzang, A. G. Konings, D. Entekhabi, M. Piles, and A. Stoffelen, 2014: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophys. Res. Lett.*, **41**, 6229–6236, https://doi.org/10.1002/2014GL061322.

McNairn, H., and Coauthors, 2015: The Soil Moisture Active Passive validation experiment 2012 (SMAPVEX12): Prelaunch calibration and validation of the SMAP soil moisture algorithms. *IEEE Trans. Geosci. Remote Sens.*, **53**, 2784–2801, https://doi.org/10.1109/TGRS.2014.2364913.

Molero, B., D. J. Leroux, P. Richaume, Y. H. Kerr, O. Merlin, M. H. Cosh, and R. Bindlish, 2018: Multi-timescale analysis of the spatial representativeness of in situ soil moisture data within satellite footprints. *J. Geophys. Res. Atmos.*, **123**, 3–21, https://doi.org/10.1002/2017JD027478.

Mudelsee, M., 2002: TAUEST: A computer program for estimating persistence in unevenly spaced weather/climate time series. *Comput. Geosci.*, **28**, 69–72, https://doi.org/10.1016/S0098-3004(01)00041-3.

——, 2003: Estimating Pearson's correlation coefficient with bootstrap confidence interval from serially dependent time series. *Math. Geol.*, **35**, 651–665, https://doi.org/10.1023/B:MATG.0000002982.52104.02.

——, 2010: *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*. Springer, 454 pp.

Naeimi, V., K. Scipal, Z. Bartalis, and W. Wagner, 2009: An improved soil moisture retrieval algorithm for ERS and METOP

scatterometer observations. *IEEE Trans. Geosci. Remote Sens.*, **47**, 1999–2013, https://doi.org/10.1109/TGRS.2008.2011617.

Ólafsdóttir, K. B., and M. Mudelsee, 2014: More accurate, calibrated bootstrap confidence intervals for estimating the correlation between two time series. *Math. Geosci.*, **46**, 411–427, https://doi.org/10.1007/s11004-014-9523-4.

Oldak, A., Y. Pachepsky, T. J. Jackson, and W. J. Rawls, 2002: Statistical properties of soil moisture images revisited. *J. Hydrol.*, **255**, 12–24, https://doi.org/10.1016/S0022-1694(01)00507-8.

Ran, Y., X. Li, R. Jin, J. Kang, and M. H. Cosh, 2017: Strengths and weaknesses of temporal stability analysis for monitoring and estimating grid-mean soil moisture in a high-intensity irrigated agricultural landscape. *Water Resour. Res.*, **53**, 283–301, https://doi.org/10.1002/2015WR018182.

Reichle, R. H., R. Koster, G. De Lannoy, W. Crow, and J. Kimball, 2014: SMAP level 4 surface and root zone soil moisture (L4_SM) data product, revision A. Algorithm Theoretical Basis Doc., 65 pp., https://nsidc.org/sites/nsidc.org/files/files/data/smap/272_L4_SM_RevA_web.pdf.

Seyfried, M. S., M. D. Murdock, C. L. Hanson, G. N. Flerchinger, and S. Van Vactor, 2001: Long-term soil water content database, Reynolds Creek experimental watershed, Idaho, United States. *Water Resour. Res.*, **37**, 2847–2851, https://doi.org/10.1029/2001WR000419.

Sherman, M., F. M. SpeedJr., and F. M. Speed, 1998: Analysis of tidal data via the blockwise bootstrap. *J. Appl. Stat.*, **25**, 333–340, https://doi.org/10.1080/02664769823061.

Smith, A. B., and Coauthors, 2012: The Murrumbidgee soil moisture monitoring network data set. *Water Resour. Res.*, **48**, W07701, https://doi.org/10.1029/2012WR011976.

Tetlock, E., B. Toth, A. Berg, T. Rowlandson, and J. Ambadan-Thomas, 2019: An 11-year (2007–2017) soil moisture and precipitation dataset. *Earth Syst. Sci. Data*, **11**, 787–796, https://doi.org/10.5194/essd-11-787-2019.

Thiébaux, H. J., and F. W. Zwiers, 1984: The interpretation and estimation of effective sample size. *J. Appl. Meteor. Climatol.*, **23**, 800–811, https://doi.org/10.1175/1520-0450(1984)023<0800:TIAEOE>2.0.CO;2.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.

Wagner, W., G. Lemoine, and H. Rott, 1999: A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sens. Environ.*, **70**, 191–207, https://doi.org/10.1016/S0034-4257(99)00036-X.

Wen, X., H. Lu, C. Li, T. Koike, and I. Kaihotsu, 2014: Intercomparison of soil moisture products from SMOS, AMSR-E, ECMWF and GLDAS over the Mongolia Plateau. *Proc. SPIE*, **9260**, 92600O, https://doi.org/10.1117/12.2068952.

Yee, M. S., J. P. Walker, A. Monerris, C. Rüdiger, and T. J. Jackson, 2016: On the identification of representative in situ soil moisture monitoring stations for the validation of SMAP soil moisture products in Australia. *J. Hydrol.*, **537**, 367–381, https://doi.org/10.1016/j.jhydrol.2016.03.060.

Yilmaz, M. T., and W. T. Crow, 2014: Evaluation of assumptions in soil moisture triple collocation analysis. *J. Hydrometeor.*, **15**, 1293–1302, https://doi.org/10.1175/JHM-D-13-0158.1.

Zwieback, S., K. Scipal, W. Dorigo, and W. Wagner, 2012: Structural and statistical properties of the collocation technique for error characterization. *Nonlinear Processes Geophys.*, **19**, 69–80, https://doi.org/10.5194/npg-19-69-2012.

Zwiers, F. W., 1990: The effect of serial correlation on statistical inferences made with resampling procedures. *J. Climate*, **3**, 1452–1461, https://doi.org/10.1175/1520-0442(1990)003<1452:TEOSCO>2.0.CO;2.