Validation of SMOS, SMAP, and ESA CCI Soil Moisture Over a Humid Region

Xiaoyong Xu and Steven K. Frey

***Abstract*—With recent advances in satellite microwave soil moisture estimation, there is a demand for up-to-date validation of satellite soil moisture products. This article presents a sparse network validation over a humid region within the Laurentian Great Lakes basin for five state-of-the-art satellite soil moisture datasets, including the Soil Moisture and Ocean Salinity (SMOS) Level 2 Soil Moisture User Data Product (MIR_SMUDP2) V650, the Soil Moisture Active Passive (SMAP) Enhanced Level 3 Radiometer Soil Moisture (SPL3SMP_E) Version 4, and the European Space Agency Climate Change Initiative (CCI) Soil Moisture v05.2 (containing the Active, Passive, and Combined sets). Unsurprisingly, the five sets of soil moisture products performed differently. With respect to the unbiased root-mean-squared error (ubRMSE), the CCI Combined product performed best (an average ubRMSE of about 0.04 m^3 m^{-3}), whereas the CCI Passive had the lowest performance with an average ubRMSE exceeding 0.10 m^3 m^{-3}. Overall, in terms of correlation measure, the SMAP and CCI Combined performed better than other products, with the lowest skill from the SMOS product. The SMAP product performed best in the context of the soil moisture anomaly detection, whereas the SMOS and CCI Passive showed the lowest anomaly correlation with the *in situ* observations. The validation results provide an important guidance for hydrological and meteorological applications involving satellite soil moisture datasets in the study region or other similar areas.**

***Index Terms*—European Space Agency (ESA) Climate Change Initiative (CCI), satellite, soil moisture, Soil Moisture Active Passive (SMAP), Soil Moisture and Ocean Salinity (SMOS).**

I. Introduction

ACCURATE estimation of soil water content, which affects the partitioning of energy and water at the land surface, is critically important to understanding variability in the hydrological cycle and in water resource availability. Microwave remote sensing holds the ability to provide spatially distributed surface soil moisture information at multiple scales [1]–[7]. Over the past decade, significant progress has been made in satellite microwave soil moisture detection and estimation. The highlighted advances include launches of the European Space Agency's (ESA) Soil Moisture and Ocean Salinity (SMOS) mission [8], [9], the NASA Soil Moisture Active Passive (SMAP) mission [10], [11], and the ESA Climate Change Initiative (CCI) Soil Moisture project [12]–[15]. Meanwhile, an intensive global research effort has been made to evaluate the soil moisture products derived from those missions or projects [16]–[24]. The evaluation studies have provided important information for the application of satellite soil moisture datasets by the hydro-climatological community. Satellite soil moisture products also undergo continuous updating, and hence there is a demand for up-to-date validation studies. Furthermore, until recently, most of the satellite soil moisture assessment studies were conducted over the arid, semiarid, or cold climate regions, e.g., [16]–[24]. As such, the performances of satellite soil moisture products in other climate regimes (e.g., humid continental climate) were typically under-represented.

The humid regions are typically characterized by complex and substantial soil moisture temporal variations, which were not sufficiently captured in many of the previous validation studies. The primary aim of this study is to advance understanding of the strengths and weaknesses of state-of-the-art satellite soil moisture products in detecting surface soil moisture variability under humid climate conditions, thus providing an important guidance for hydrological and meteorological applications involving satellite soil moisture datasets in humid regions. In [25], the SMAP Enhanced Level 3 Radiometer Soil Moisture, along with the ESA CCI combined and passive soil moisture datasets were evaluated within the Laurentian Great Lakes basin (GLB), which is a typical humid region. The present study continues the evaluation of satellite soil moisture products in the GLB with the following updated components.

1) The SMOS and CCI Active products are additionally included.
2) The latest versions of the satellite soil moisture products are investigated.
3) A longer study period that includes anomaly detection is examined.
4) The validation is conducted using a more vigorous experimental design.

The rest of this article is organized as follows. In Section II, the satellite soil moisture and *in situ* datasets are described. Section III presents the validation results for the satellite soil moisture products. A discussion is provided in Section IV. Finally, Section V concludes this article.

Manuscript received April 27, 2021; revised July 11, 2021; accepted October 18, 2021. Date of publication October 21, 2021; date of current version November 3, 2021. This work was supported by the Canadian Space Agency under Grant 21SUESMVAS. *(Corresponding author: Xiaoyong Xu.)*

Xiaoyong Xu is with the Department of Chemical and Physical Sciences, University of Toronto Mississauga, Mississauga, ON L5L 1C6, Canada (e-mail: xiaoyong.xu@utoronto.ca).

Steven K. Frey is with Aquanty, Waterloo, ON N2L 5C6, Canada, and the Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sfrey@aquanty.com).

Digital Object Identifier 10.1109/JSTARS.2021.3122068

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

II. Data and Methods

A. Satellite Soil Moisture Products

In this work, we assess five satellite soil moisture products, including the SMOS Level 2 Soil Moisture User Data Product (MIR_SMUDP2) V650 [26], [27], the SMAP Enhanced Level 3 Radiometer Soil Moisture (SPL3SMP_E), Version 4 [28], and the ESA CCI Soil Moisture v05.2 Active, Passive, and Combined products [29], [30].

The MIR_SMUDP2 V650, released in November 2017, represents the latest version of SMOS L2 soil moisture retrievals [26]. The product provides the SMOS swath-based soil moisture retrievals and auxiliary information on the Icosahedral Snyder equal area Earth grid (the nodes are equally spaced at about 15 km) for the ascending (6:00 A.M. LST) and descending (6:00 P.M. LST) orbits separately. The retrieved soil moisture was derived from the SMOS L1C brightness temperatures using an iterative algorithm [26], [27].

The SPL3SMP_E Version 4 product is the latest SMAP enhanced Level 3 radiometer passive soil moisture estimation, providing daily composite estimates of surface soil moisture derived from the SMAP A.M. (i.e., from 6:00 A.M. descending half orbits) and P.M. (i.e., from 6:00 P.M. ascending half orbits) brightness temperature datasets, respectively, at the Earth-fixed, global cylindrical 9-km Equal-Area Scalable Earth Grid 2.0 [28]. The 9 km resolution of the SPL3SMP_E product was based upon the native SMAP radiometer footprint of 36 km using the Backus–Gilbert optimal interpolation [28].

The ESA CCI soil moisture datasets are the merged soil moisture retrievals from a series of microwave sensor systems [13], [29]. The level 2 soil moisture retrievals from the active microwave sensors (ERS-1/2 Scatterometer and Metop Advanced Scatterometers) were merged to generate the CCI Active soil moisture product, whereas the retrieved soil moisture values from the passive microwave sensors, including SMMR, SSM/I, TRMM, AMSR-E, AMSR2, Windsat, SMOS, and SMAP in v05.2, were merged to generate the CCI Passive product. The CCI Combined soil moisture dataset was produced by merging all input soil moisture retrieved from both the active and passive microwave sensors.

All input active microwave sensor soil moisture retrievals were taken directly from external sources and then rescaled to the climatology of Advanced Scatterometer (ASCAT) surface soil moisture using the cumulative distribution function (CDF) matching method prior to merging into the CCI Active. All input passive microwave sensor soil moisture retrievals were obtained from the passive sensor brightness temperatures with the Land Parameter Retrieval Model (LPRM) within the CCI Soil Moisture Production System [29]. The passive microwave sensor soil moisture retrievals were then rescaled to the soil moisture climatology from the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) (or the rescaled AMSR2 climatology, which was used only for the rescaling of SMAP soil moisture) before merging into the CCI Passive product. In contrast, in the CCI Combined dataset, the input soil moisture retrievals from both the active and passive microwave sensors, ahead of the merging, were all rescaled to the soil moisture climatology from the Global Land Data Assimilation System (GLDAS).

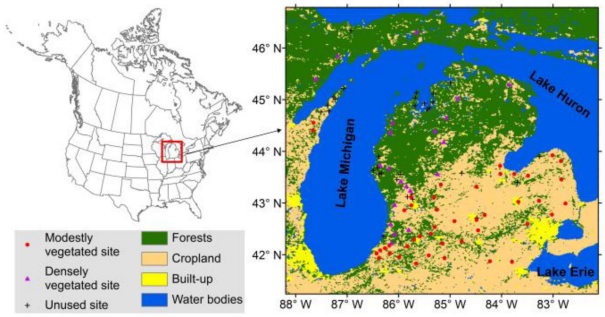

Fig. 1. Study domain and location of the *in situ* soil moisture observational sites from MAWN. The dots and triangles denote the MAWN stations located within the modestly vegetated (mainly cropland) and densely vegetated (forests) regions. The MAWN stations marked with the crosses are not used in this study.

The ESA CCI soil moisture was provided at a spatial resolution of 0.25° in both latitude and longitude and a temporal resolution of 1 day centered at 0:00 UTC [29]. Relative to previous versions, the highlighted update made to the ESA CCI soil moisture v05.2 is the inclusion of SMAP radiometer soil moisture. As in the SMOS and SMAP products, the soil moisture values were provided in the unit of volumetric water content (i.e., m^3 m^{-3}) in the CCI Passive and Combined. The CCI Active soil moisture retrievals, which were expressed in degree of saturation (%), are converted to the volumetric water content values (using a soil porosity map provided by the CCI project) for the validation herein.

In this study, the validation is based upon nearly five years (March 31, 2015–December 31, 2019) of daily time series. The study period is the longest common time coverage (at the time of this analysis) for all five satellite soil moisture products of interest here. Within each dataset, the ancillary information and quality assessment flags were used to exclude the corresponding soil moisture estimates that were contaminated by open water, frozen surface, snow, rain, or radio frequency interference, etc. To be consistent with the temporal resampling of the daily CCI products [29], the SMOS and SMAP daily composite soil moisture estimates were constructed (or reconstructed) with respect to the 24-h window centered at 0:00 UTC using their observations from both the ascending and descending modes.

B. In Situ *Soil Moisture Observations*

Our study domain is in the U.S. state of Michigan, within the GLB. The satellite soil moisture products were evaluated using the *in situ* soil moisture data observed by Michigan State University's Enviro-weather Automated Weather Station Network (MAWN).[1] Fig. 1 shows the study domain and location of the MAWN observational stations.

The study domain is characterized by a humid continental climate. The land cover types mainly include cropland, forests, built-up (urban), and water bodies (see Fig. 1), which facilitates the assessment of land cover impacts on satellite soil moisture

[1][Online]. Available: https://mawn.geo.msu.edu

estimation. The MAWN network is currently comprised of 100 stations, which provide hourly surface and root zone soil moisture measurements. The MAWN *in situ* soil moisture measurements from the top 30 cm layer were used to validate the satellite soil moisture products. It is acknowledged that the measuring soil depths for the *in situ* soil moisture and satellite retrievals are not ideally matched; however, the soil depth discrepancy was shown to have negligible impact on the validation in this region [31], [32]. The *in situ* daily soil moisture time series are constructed with respect to the 24-h windows centered at 0:00 UTC (as used in the daily-spaced satellite products) using all hours of data. In the *in situ* time series data, any spikes or temporal inhomogeneities that could not be properly associated with physical processes were removed. The stations with less than 100 valid daily observations (over the study period of March 31, 2015–December 31, 2019) were excluded from the validation. Also, those stations that experienced the surface soil moisture measuring depth changes within the study period were excluded from the validation. In addition, the MAWN stations within 10 km of water bodies were not used. After the filtering, the 44 modestly vegetated (red dots in Fig. 1) and 21 densely vegetated (pink triangles in Fig. 1) MAWN sites were utilized for satellite soil moisture product validation in this study.

C. Performance Metrics

The validation was conducted at point-scale (i.e., the sparse network validation) for all five satellite soil moisture products. At the location of each *in situ* station, the unbiased root-mean-squared error (ubRMSE), correlation coefficient (R), and anomaly R are computed based upon the daily time series using the following equations:

$$\text{ubRMSE} = \sqrt{E\left[\left(\left(\theta_s - E\left[\theta_s\right]\right) - \left(\theta_i - E\left[\theta_i\right]\right)\right)^2\right]} \quad (1)$$

$$R = E\left[\left(\theta_s - E\left[\theta_s\right]\right)\left(\theta_i - E\left[\theta_i\right]\right)\right]\left(\sigma_s \sigma_i\right)^{-1} \quad (2)$$

$$\text{Anomaly } R = E[\left(\theta_s' - E\left[\theta_s'\right]\right)\left(\theta_i' - E\left[\theta_i'\right]\right)](\sigma_s' \sigma_i')^{-1} \quad (3)$$

where E [•] is the expectation operator. θ_s and θ_i indicate the daily time sequences of satellite soil moisture and *in situ* observations, respectively. θ_s' and θ_i' are the anomaly equivalents of θ_s and θ_i, respectively. The soil moisture anomalies are defined as departures of raw soil moisture values from their monthly climatology. σ_s, σ_i, σ_s', and σ_i' denote the standard deviations (stdev) of θ_s, θ_i, θ_s', and θ_i', respectively.

When the above performance metrics are used, the key information is the soil moisture variability rather than the absolute value of soil moisture. Although the sparse network validation unavoidably suffers from the spatial discrepancy between point measurements and satellite products, the point measurements from MAWN proved to be spatially representative in detecting the soil moisture variability [32].

III. Results

A. Surface Soil Moisture Variability

We first compare the surface soil moisture variability between satellite estimation and *in situ* measurements. Due to the large number of stations, it is not possible to display the time series comparison between satellite products and all *in situ* measurements. Hence, Fig. 2 presents the daily time series comparison between satellite soil moisture and *in situ* measurements at a single representative validation site.

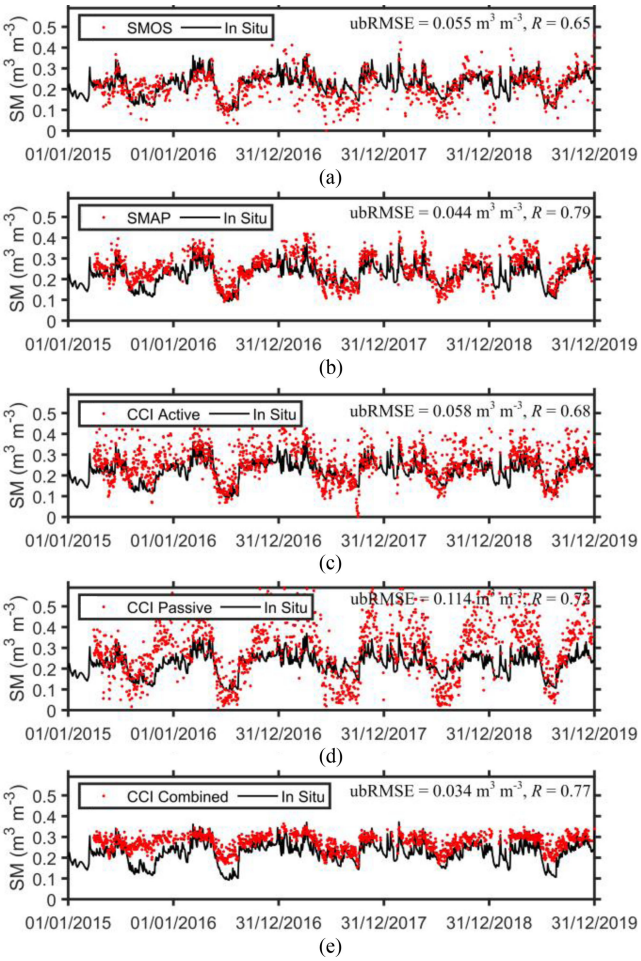

Fig. 2. Comparison of daily surface soil moisture (SM) time series: (a) *In situ* versus SMOS, (b) *In situ* versus SMAP, (c) *In situ* versus CCI Active, (d) *In situ* versus CCI Passive, and (e) *In situ* versus CCI Combined at a single station (42.65 °N, –84.92 °E).

There is a clear seasonal variability of surface soil moisture at the site, which is characterized by dry soils in summer/early autumn and saturated soils in late winter and spring (see Fig. 2). The dry period is consistent with strong incoming solar radiation (insolation) and clear skies, leading to strong evaporation. The soil moisture stays near field capacity from around November to April because evaporation is very weak during this period (due to the weak net radiative forcing at the surface). In addition, the winter evapotranspiration is also very limited because most vegetation is dormant. Early summer (May to June) typically undergoes a transitional drying phase. This is because growing vegetation depletes the soil water content. In addition, during this transition period, as the weather warms, more water evaporates from the soils, which also increased the soil drying.

All five sets of satellite soil moisture products showed reasonable agreement with the *in situ* measurements at the site in terms of the temporal pattern of soil moisture (see Fig. 2).

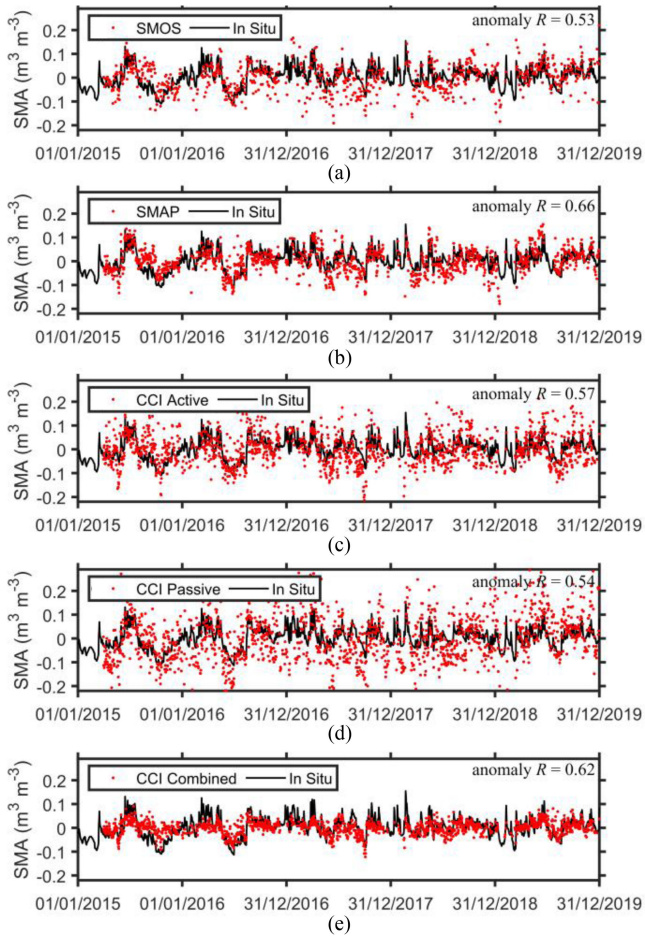

Fig. 3. Comparison of daily surface soil moisture anomaly (SMA) time series: (a) *In situ* versus SMOS, (b) *In situ* versus SMAP, (c) *In situ* versus CCI Active, (d) *In situ* versus CCI Passive, and (e) *In situ* versus CCI Combined at the station as in Fig. 2.

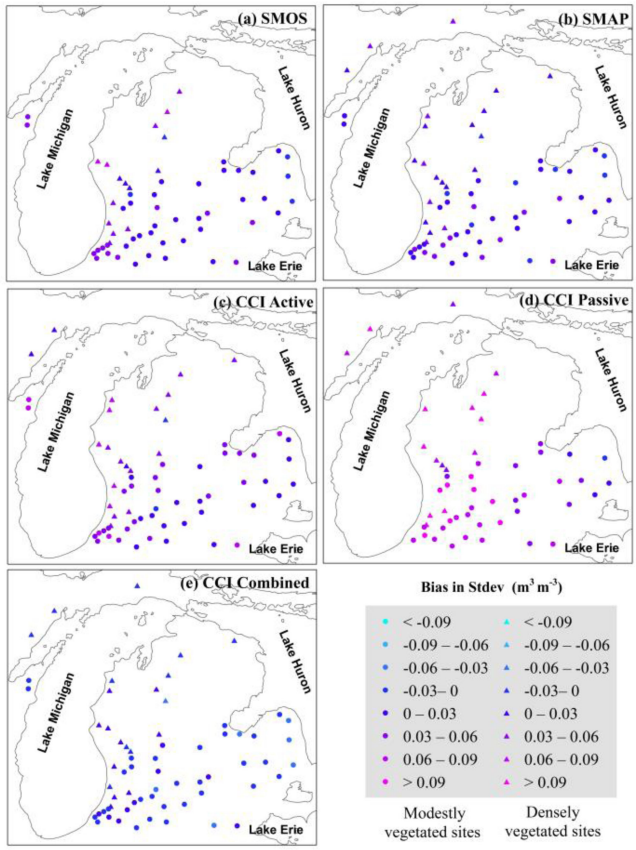

Fig. 4. Daily time series stdev difference between satellite soil moisture and *in situ* measurements across the validation stations: (a) SMOS minus *In situ*, (b) SMAP minus *In situ*, (c) CCI Active minus *In situ*, (d) CCI Passive minus *In situ*, and (e) CCI Combined minus *In situ*. The shape of symbols is the same as in Fig. 1.

Unsurprisingly, the five satellite products are different in the absolute soil moisture magnitude. This reflects their differences in sensor type and source, as well as the soil moisture retrieval and processing algorithm. The CCI Combined product [see Fig. 2(e)] shows the smallest surface soil moisture variability, which is only slightly weaker than the *in situ* soil moisture variability. In contrast, the CCI Passive product [see Fig. 2(d)] provides the largest surface soil moisture variability, characterized by the remarkedly overestimated (wetter) winter/spring soil moisture and consistently underestimated (drier) summertime soil moisture. The CCI Active [see Fig. 2(c)] and SMOS [see Fig. 2(d)] also overestimate the surface soil moisture variability. Overall, the SMAP product shows the best agreement with the *in situ* data in terms of the amplitude of the surface soil moisture fluctuations at the site [see Fig. 2(b)].

The anomaly time series equivalent of Fig. 2 is provided in Fig. 3. The soil moisture anomalies were obtained by removing the monthly climatology from the daily soil moisture values for all datasets. After the soil moisture seasonal cycle is subtracted from the raw time series, the five sets of satellite products are still able to capture the temporal variability of surface soil moisture accurately or at least reasonably at the site (see Fig. 3). This partially demonstrates the potential of the satellite soil moisture products in climate forecasting. Again, the variability of the CCI Combined soil moisture anomalies is weakest, and only slightly weaker than that of the *in situ* anomaly [see Fig. 3(e)], whereas the CCI Passive product provides the strongest anomaly soil moisture variability [see Fig. 3(d)]. The SMAP and *in situ* soil moisture anomalies are also in close agreement [see Fig. 3(b)], suggesting a strong potential of anomaly information detection for the SMAP product.

The time series stdev is a measure of the magnitude of soil moisture variability. Fig. 4 presents the stdev differences between each of the satellite soil moisture products and *in situ* measurements (i.e., bias in stdev) across all validation sites. Among the five sets of satellite products, the CCI Passive [see Fig. 4(d)] provides the largest surface soil moisture stdev, which typically exceeds the *in situ* soil moisture stdev by more than 0.06–0.09 m^3 m^{-3} for the majority of stations. The CCI Active soil moisture product [see Fig. 4(c)] typically has the second-highest amplitude of fluctuations with a stdev of 0.03 to 0.06 m^3 m^{-3} higher than the *in situ* stdev at most sites [Fig. 4(c)], followed by the SMOS product with the stdev bias typically also exceeding 0.03–0.06 m^3 m^{-3}, especially for the sites near

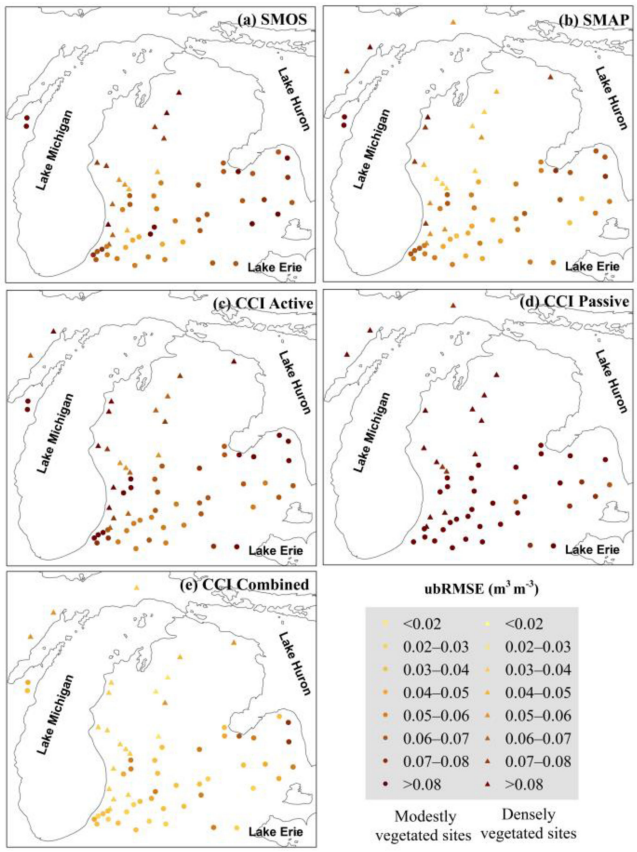

Fig. 5. ubRMSE values at the individual validation stations for (a) SMOS, (b) SMAP, (c) CCI Active, (d) CCI Passive, and (e) CCI Combined. The shape of symbols is the same as in Fig. 1.

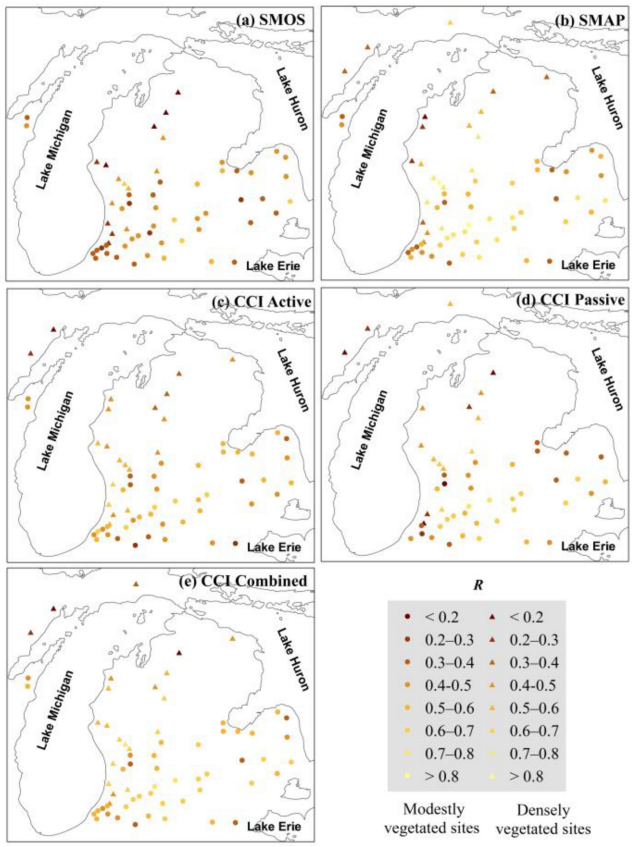

Fig. 6. Same as Fig. 5, but for the time series correlation coefficient *R* metric.

Lake Michigan [Fig. 4(a)]. The CCI Combined product generally shows the lowest soil moisture stdev values, which agree well with those of the *in situ* soil moisture (a stdev bias of within $\pm$ 0.02 m^3 m^{-3}) for the majority of sites [see Fig. 4(e)]. The time series stdev of the SMAP soil moisture is also close to the *in situ* stdev, with a positive stdev bias within 0.03 m^3 m^{-3} for the majority of stations, although a moderate stdev bias of 0.03–0.06 m^3 m^{-3} was observed over a small subset of the validation sites [see Fig. 4(b)].

B. Performance Metrics Across the Validation Sites

Next, we present the sparse network performance metrics for the satellite soil moisture products. The metrics were calculated based upon the daily soil moisture time series for March 31, 2015 to December 31, 2019. The metrics were not computed when there were less than 100 valid daily data points for an individual site.

Fig. 5 shows the ubRMSE errors across the validation stations. The ubRMSE values of the CCI Combined product were typically the lowest among the five datasets, with an ubRMSE within 0.04 m^3 m^{-3} for most validation sites [see Fig. 5(e)]. Overall, the SMAP soil moisture had the second-lowest ubRMSE errors with an ubRMSE typically ranging from 0.03 to 0.06 m^3 m^{-3} [see Fig. 5(b)]. The ubRMSE values are typically larger in the lake coastal areas than over inland sites [see Fig. 5(a) and (b)], which relates to the impact of open water on satellite soil moisture estimation.

The CCI Passive product shows very large ubRMSE (typically exceeding 0.08 m^3 m^{-3}) across the network sites [see Fig. 5(d)]. This feature is highly related to the significantly overestimated soil moisture variability from the product [see Fig. 4(d)] since the ubRMSE must be equal to or higher than the absolute value of the bias in stdev [33]. The CCI Active product typically has an ubRMSE ranging from 0.05 to 0.08 m^3 m^{-3} over the inland areas and greater than 0.08 m^3 m^{-3} for the coastal sites [see Fig. 5(c)]. As compared to the CCI Active, the SMOS product performed slightly better, showing an ubRMSE of about 0.04–0.07 m^3 m^{-3} for most sites [see Fig. 5(a)].

The time series correlation coefficient *R* values between the satellite soil moisture and *in situ* data are shown in Fig. 6. The *R* values for the CCI Combined [Fig. 6(e)] and SMAP [Fig. 6(b)] typically vary between 0.4 and 0.8 with the exception of a few sparsely distributed sites. Both the CCI Active [Fig. 6(c)] and CCI Passive products [Fig. 6(d)] provide *R* values that vary between 0.4 and 0.7 across most validation sites, although the latter [Fig. 6(d)] has more sites with extremely low *R* values (< 0.3). The SMOS product typically has *R* value of 0.3 to 0.6 over the modestly vegetated sites [Fig. 6(a)]; however, about half of the densely vegetated sites suggested very low (< 0.3) or even negative *R* for the product. By comparing the ubRMSE (Fig. 5) and *R* (Fig. 6) metrics, the sites with higher ubRMSE

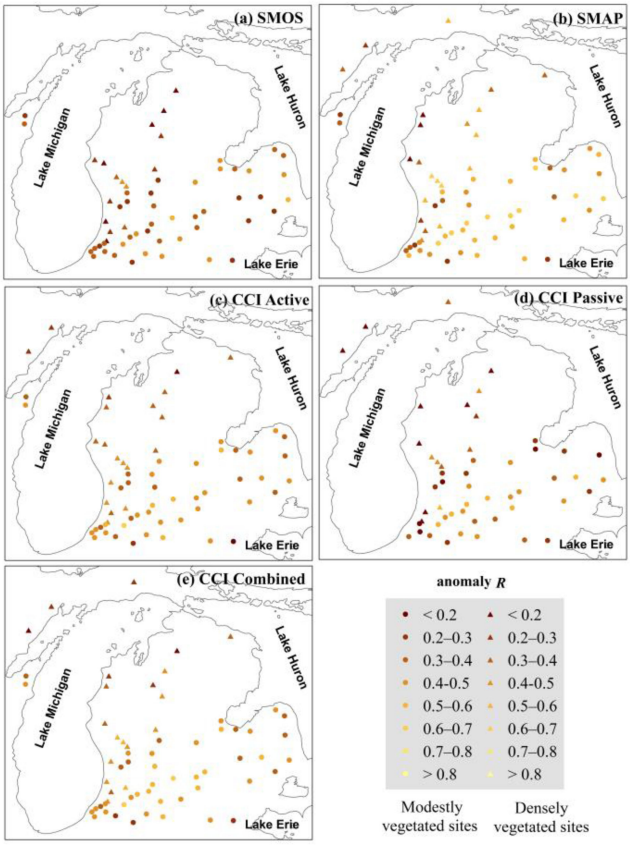

Fig. 7. Same as Fig. 5, but for the anomaly time series correlation coefficient *R* metric.

errors do not necessarily perform worse with respect to the *R* measure. This clearly illustrates that the ubRMSE and *R* metrics provide different perspectives on the performance of the satellite products. As pointed out in [33], the same ubRMSE can correspond to a wide range of *R* values. The relationship between the two metrics (ubRMSE and *R*) depends upon the stdev of satellite and *in situ* time series.

The anomaly *R* values between satellite soil moisture retrievals and *in situ* data are provided in Fig. 7. For each of the satellite soil moisture products, the spatial pattern of the anomaly *R* across the validation sites is very similar to that for the *R* metric, with the exception that the former tends to have lower values because the contribution of the seasonal cycle is deducted. For most of the validation sites, after the seasonal cycle is deducted from the raw time series, the temporal variability of surface soil moisture is still reasonably captured by all five satellite datasets. The anomaly *R* values typically range from 0.4 to 0.7 for the SMAP soil moisture [Fig. 7(b)] and from 0.3 to 0.5 for the CCI Combined [Fig. 7(e)] and Active [Fig. 7(c)] products. The CCI Passive and SMOS provide anomaly *R* values less than 0.4 for most of the modestly vegetated sites and below 0.3 for the majority of densely vegetated sites [see Fig. 7(d) and (a)].

In Figs. 5–7, the metrics for the SMOS and CCI Passive soil moisture products were not computed at a few coastal sites because the two products do not provide sufficient valid soil moisture retrievals for these sites. The corresponding satellite nodes were susceptible to contamination from the open water surface in the two products due to the coarse native footprint of passive sensor (radiometer) systems (typically > 35 km). In contrast, although the native footprint of the SMAP radiometer is also coarse ($\sim$36 km), a water body correction was applied to the SMAP brightness temperature (TB) data for the regions with the presence of open water [34]. The water TB correction was performed at the footprint level based upon the SMAP radiometer antenna gain pattern as described in [35]. The SPL3SMP_E product was derived from the water body corrected TBs, thus leading to a better coverage of the coastal sites. The Active and Combined products of CCI also provide a better coverage of the coastal sites than the SMOS and CCI Passive products. A possible explanation stemmed from the active sensor Metop ASCAT soil moisture data, which were used in the CCI Active and Combined, but not in the CCI Passive. The Active and Combined products of CCI (v05.2) made use of the latest version of ASCAT soil moisture retrievals from the EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H SAF), i.e., the H SAF H115 and H116 Metop ASCAT SSM CDR v5 [36], which have a spatial sampling on a regular 12.5 km grid in orbit geometry with a spatial resolution of about 25 km.

C. Summary Metrics

We now discuss the average verification metrics for the sparse network assessment. Fig. 8 provides the average metrics with 95% confidence intervals from each satellite product, shown separately for three groups: all modestly vegetated (cropland) sites, all densely vegetated (forest) sites, and total validation sites. To estimate the confidence intervals, a large number ($N = 5000$) of ensembles were generated. Each ensemble contained random subsamples from the single-station metric values. The confidence intervals were then estimated based upon the distribution of the *N* ensemble means. As discussed above, the valid reference sites are slightly less for the SMOS and CCI Passive products than for other datasets (see Figs. 5–7); however, this deviation will not result in a biased comparison since the relative mean skill keeps consistent across the five products.

The CCI Combined soil moisture provides the lowest average ubRMSE at about 0.04 m^3 m^{-3} for all three groups of sites (see Fig. 8(a), black). The SMAP soil moisture has the second-lowest average ubRMSE at about 0.056 m^3 m^{-3} for all three site groups (see Fig. 8(a), blue). The SMOS (Fig. 8(a), red) and CCI Active (Fig. 8(a), pink) provide average ubRMSE of about 0.07 m^3 m^{-3} across the network, with a slightly larger error for the forest group than for the cropland group. The largest error came from the CCI Passive product (Fig. 8(a), green), showing an average ubRMSE of about 0.10 $m^3 m^{-3}$ for the group of cropland sites and about 0.12 $m^3 m^{-3}$ for the forest group.

Theoretically, the remotely sensed surface soil moisture is less accurate in densely vegetated areas because of the vegetation attenuation effects on the soil's microwave emission [4], [5]. As shown in Fig. 8(a), in the CCI Combined and SMAP results, the average ubRMSE from the cropland group was not statistically different from that from the forest group. One possible reason

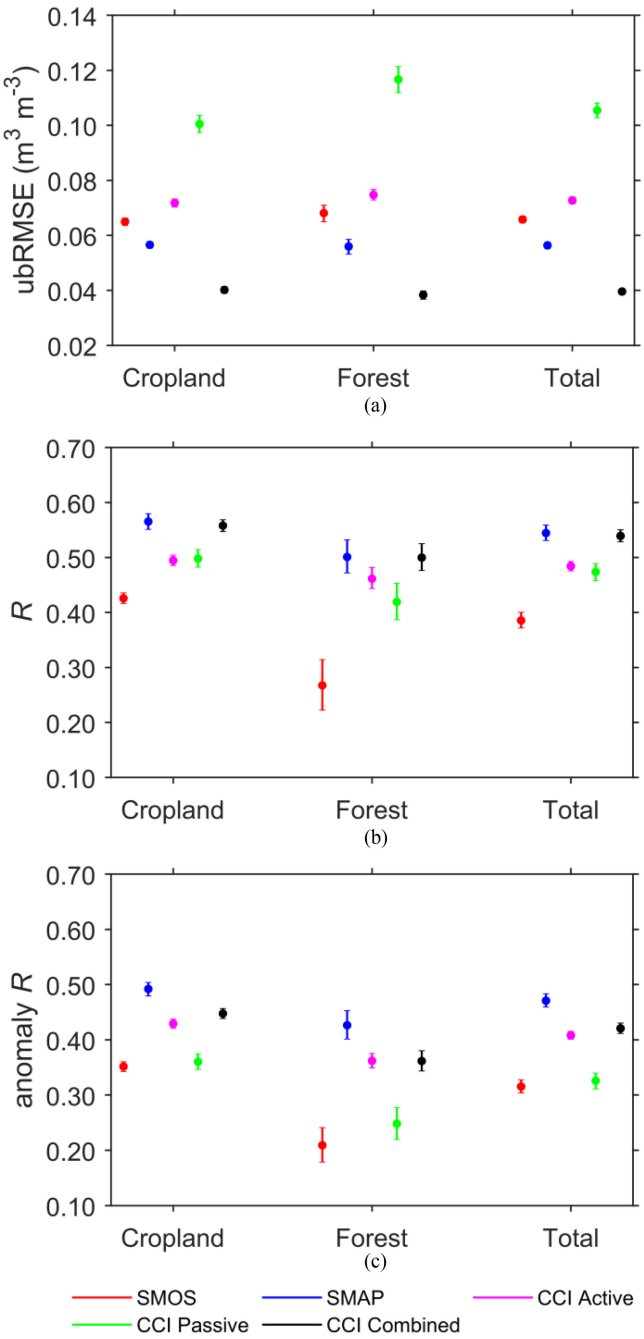

Fig. 8. Average metrics: (a) ubRMSE, (b) *R*, and (c) anomaly *R*, from each of the five satellite soil moisture products (SMOS, SMAP, CCI Active, CCI Passive, and CCI Combined), across all modestly vegetated (cropland) sites, all densely vegetated (forest) sites, and total (both cropland and forest) sites, respectively. Error bars denote 95% confidence intervals.

is that a few of the cropland sites near the lakes held substantial ubRMSE values [see Fig. 5(e) and (b)] due to the impact of open water on the corresponding satellite nodes (but the size of water body is within the acceptable level based upon the product's quality assessment flags), which increased the average ubRMSE for the cropland group. When those coastal cropland sites are partially excluded [e.g., as shown in the CCI Passive product in Fig. 5(d)], the average ubRMSE value would be lower for the cropland group than for the forest group (Fig. 8(a), green), consistent with the effects of vegetation canopy on satellite soil moisture estimation.

In terms of the average *R*, there is no statistically significant difference between the SMAP (Fig. 8(b), blue) and CCI Combined (Fig. 8(b), black) datasets. The two products are the collective best, with a mean *R* of about 0.56 (0.50) for the cropland (forest) group. The CCI Active and Passive products are the collective second best, with a mean *R* close to 0.50 for the cropland sites and slightly below 0.50 for the forest group (Fig. 8(b), pink and green). The SMOS product (Fig. 8(b), red) provided the lowest summary *R*, which is slightly above 0.4 for the cropland group, but as low as about 0.27 for the forest group. The very low average *R* from the SMOS forest group has been previously observed in the study region [32].

For each satellite product, the average *R* from the cropland group was always higher than its forest group counterpart. This reflects the difficulty in capturing the variability and phase of soil moisture in densely vegetated areas from satellite microwave remote sensing. In addition, the relatively large error bars (95% confidence intervals) for the average *R* from the forest group partially illustrate a substantial uncertainty in the capability of satellite microwave remote sensing to detect the soil moisture variability and phase in densely vegetated regions.

A mean anomaly *R* of about 0.50 (0.42) for the cropland (forest) group is observed for the SMAP product, which outperforms other datasets [see Fig. 8(c)]. The CCI Combined and CCI Active are the collective second best with respect to the average anomaly *R*, showing a mean anomaly *R* close to 0.44 (0.35) for the cropland (forest) group (Fig. 8(c), pink and black). The SMOS product and CCI Passive (Fig. 8(c), red and green) collectively provide the lowest average anomaly *R*, which is about 0.35 (below 0.25) for the cropland (forest) group. Again, in each product, the average anomaly *R* from the cropland group is always higher than the forest group counterpart, reflecting the difficulty in capturing the anomaly variability of soil moisture in heavily vegetated areas.

IV. Discussion

In terms of the correlation measure *R*, the overall skill of SMOS soil moisture (MIR_SMUDP2 V650) in this humid study region (a mean *R* of about 0.4 for all sites) is lower than the global average skill of the product based upon more than 100 stations from nine soil moisture networks (a mean *R* above 0.6) [26]. This deviation from the global trend is thought to be partially related to a humid climate and the complexity of land cover in this region (see Fig. 1).

Overall, the ubRMSE of SMAP soil moisture obtained in this region (above 0.05 m^3m^{-3} on average) exceeds that obtained for the arid or semiarid areas (around 0.04 m^3m^{-3} on average) [20], [23]. Earlier versions of SPL3SMP_E have been previously assessed in the same region [25], [31]. The average ubRMSE of 0.056 m^3m^{-3} obtained there for SPL3SMP_E version 4 is equivalent to that for the SPL3SMP_E version 3 product (0.058 m^3 m^{-3} on average) [31]. Meanwhile, the overall skill of SPL3SMP_E version 4 is also similar to SPL3SMP_E version

3 [31] with respect to the *R* and anomaly *R* measures, showing respective average values of about 0.53 and 0.47 for both versions.

Notice that the CCI soil moisture data are the merged products by making use of observations from the existing microwave sensors in the past and nowadays [29]. The CCI Combined and Passive products are not entirely independent of the SMOS and SMAP soil moisture products, as the CCI Combined and Passive products (version v05.2) ingested the SMOS and SMAP measuring information (see Section II-A). However, the CCI soil moisture is not simply the sum of the individual sensor soil moisture products. The SMOS and SMAP soil moisture used in the CCI Combined and Passive products differs from the individual SMOS and SMAP products in the following aspects. First, in the CCI Combined and Passive products, the soil moisture retrievals from passive microwave sensors (including SMAP, SMOS, AMSR2, etc.) were not directly taken from external sources. Instead, they were produced within the CCI Soil Moisture Production System based upon the LPRM algorithm [29], which largely differs from the retrieval algorithms used for generating the individual SMOS [27] and SMAP [34] soil moisture products. Second, the LPRM-based passive sensor soil moisture retrievals were rescaled to the climatology of AMSR-E or AMSR2 soil moisture in the CCI Passive product and to the climatology of GLDAS soil moisture in the CCI Combined product using the CDF method, whereas the rescaling was not applied to the individual SMOS and SMAP products. Third, the SMOS and SMAP observations from only their descending overpasses were used in the CCI products [29], whereas the daily composites using the observations from both the ascending and descending orbits are evaluated for the individual SMAP and SMOS products in this study. Additionally, the SMAP, SMOS, and CCI Active products are totally independent of one another.

The three sets of CCI products (Active, Passive, and Combined) performed differently in the study area. This is mainly due to the different sensor and data sources employed in each dataset. In the CCI Combined product, both active and passive sensor measurements were used, which led to the highest performing soil moisture estimation in this humid region. The CCI Combined product (v03.2) has also shown a good performance in dry regions [24]. In addition, the overall performance of the v05.2 CCI Combined and CCI Passive was slightly improved over their v04.4 versions [25]. This reflects the contribution of SMAP observations, which were not included in the earlier versions of CCI Passive and CCI Combined products.

Performance discrepancies are also related to the different rescaling schemes associated with the three CCI datasets. The CCI Passive product provided the largest ubRMSE. This was caused by the remarkably overestimated time series stdev in the product [see Fig. 4(d)] since the ubRMSE metric is highly sensitive to the bias in stdev [33]. In the CCI Passive product, all input passive sensor soil moisture observations, prior to merging, were rescaled to the climatology of AMSR-E with the CDF matching method as used in [37]. The AMSR-E (LPRM-based) soil moisture had a relatively large climatological stdev (0.08 m^3 m^{-3} or higher) in the study region [38], which exaggerated the amplitude of soil moisture fluctuations in the CCI Passive product. In the CCI Combined dataset, the soil moisture retrievals from both passive and active sensors were rescaled to the soil moisture climatology of GLDAS, which was typically characterized by a smaller climatological stdev than that of AMSR-E, e.g., [29]. This led to a weaker amplitude of soil moisture fluctuations, which agreed better with the stdev observed for the *in situ* data [see Fig. 4(e)], and therefore the relatively low ubRMSE for the CCI Combined product. The climatology of ASCAT soil moisture was utilized as the reference for rescaling all input active sensor soil moisture observations in the CCI Active product [29], which may be partially responsible for the intermediate stdev bias [Fig. 4(c)] and ubRMSE [Fig. 5(c)] obtained for the product.

It should be noted that although the ubRMSE for the CCI Passive dataset is typically the largest across the validation sites, the corresponding *R* value is not necessarily the lowest among the five datasets. This is because the correlation measure is not sensitive to bias in stdev [33]. At most validation sites, the *R* values for the CCI Passive product are close to those for the CCI Active product and higher than those from SMOS [see Figs. 6 and 8(b)].

V. Conclusion

Unsurprisingly, the SMOS (MIR_SMUDP2 V650), SMAP (SPL3SMP_E V004) and ESA CCI (Active, Passive, and Combined) soil moisture products performed differently over the study region. In terms of the ubRMSE metric, the CCI Combined product performed best (i.e., with the lowest ubRMSE), whereas the CCI Passive had the lowest performance (i.e, with the largest ubRMSE). Overall, with respect to the R metric, the SMAP and CCI Combined performed better than other products, with the lowest R skill obtained from the SMOS product. The SMAP product performed best in the context of the SMA detection, whereas the SMOS and CCI Passive collectively provided the lowest anomaly *R* skill. As expected, all products performed better in areas with less heavy vegetation than in the more densely vegetated areas.

The results from this study advance our understanding of the strengths and weaknesses associated with the five satellite soil moisture products in regard to their ability to represent regional soil moisture variability under humid climate and vegetation conditions similar to those of the GLB. The quantified limitations and uncertainties inherent to each product provide important guidance for the development and improvement of the corresponding soil moisture retrieval algorithms, such as the rescaling scheme within the CCI Passive product in regions with similar climate and vegetation scenarios. This study also complements previous efforts in satellite soil moisture assimilation over this region, e.g., [32], [38]–[40], and could provide additional evidence-based support for the objective selection of the most up-to-date satellite soil moisture datasets for use in hydro-climatological analyses or computational modeling activities, e.g., [41]–[43].

ACKNOWLEDGMENT

The authors are grateful to Michigan State University and the Enviro-weather Project for the MAWN soil moisture data used in this study. Thanks also go to the ESA Earth Observation Services Helpdesk, the ESA CCI Soil Moisture Team, and the National Snow and Ice Data Center for providing access to their satellite soil moisture products.

REFERENCES

[1] T. J. Jackson, "Soil moisture estimation using special satellite microwave/imager satellite data over a grassland region," *Water Resour. Res.*, vol. 33, no. 6, pp. 1475–1484, Jun. 1997.

[2] E. G. Njoku, T. J. Jackson, V. Lakshimi, Y. K. Chan, and S. V. Nghiem, "Soil moisture retrieval from AMSR-E," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 215–229, Feb. 2003.

[3] R. Bindlish *et al.*, "Soil moisture estimates from TRMM microwave imager observations over the Southern United States," *Remote Sens. Environ.*, vol. 85, no. 4, pp. 507–515, Jun. 2003.

[4] T. J. Jackson, "Estimation of surface soil moisture using microwave sensors," in *Encyclopedia of Hydrological Sciences*, M. G. Anderson, J. J. McDonnell, and T. Gale, Eds. Hoboken, NJ, USA: Wiley, Apr. 2006. [Online]. Available: doi: 10.1002/0470848944.hsa060.

[5] M. Owe, R. A. M. de Jeu, and T. H. R. Holmes, "Multisensor historical climatology of satellite-derived global land surface soil moisture," *J. Geophys. Res.*, vol. 113, no. F1, Mar. 2008, Art. no. F01002.

[6] W. Wagner *et al.*, "The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications," *Meteor. Zeitsch.*, vol. 22, no. 1, pp. 5–33, Feb. 2013.

[7] X. Xu, J. Li, and B. A. Tolson, "Progress in integrating remote sensing data and hydrologic modeling," *Prog. Phys. Geography*, vol. 38, no. 4, pp. 464–498, Aug. 2014.

[8] Y. H. Kerr *et al.*, "The SMOS mission: New tool for monitoring key elements of the global water cycle," *Proc. IEEE*, vol. 98, no. 5, pp. 666–687, May 2010.

[9] Y. H. Kerr *et al.*, "The SMOS soil moisture retrieval algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1384–1403, May 2012.

[10] D. Entekhabi *et al.*, "The soil moisture active and passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010.

[11] J. R. Piepmeier *et al.*, "SMAP L-band microwave radiometer: Instrument design and first year on orbit," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 1954–1966, Apr. 2017.

[12] Y. Y. Liu *et al.*, "Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 2, pp. 425–436, 2011.

[13] W. Dorigo *et al.*, "ESA CCI soil moisture for improved earth system understanding: State-of-the art and future directions," *Remote Sens. Environ.*, vol. 203, pp. 185–215, Jul. 2017.

[14] A. Gruber, W. Dorigo, W. Crow, and W. Wagner, "Triple collocation-based merging of satellite soil moisture retrievals," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6780–6792, Dec. 2017.

[15] A. Gruber, T. Scanlon, R. van der Schalie, W. Wagner, and W. Dorigo, "Evolution of the ESA CCI soil moisture climate data records and their underlying merging methodology," *Earth Syst. Sci. Data*, vol. 11, pp. 717–739, May 2019.

[16] A. Al Bitar *et al.*, "Evaluation of SMOS soil moisture products over continental U.S. using the SCAN/SNOTEL network," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1572–1586, May 2012.

[17] I. Gherboudj, R. Magagi, K. Goita, A. A. Berg, B. Toth, and A. Walker, "Validation of SMOS data over agricultural and Boreal forest areas in Canada," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1623–1635, May 2012.

[18] C. Albergel *et al.*, "Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations," *Remote Sens. Environ.*, vol. 118, pp. 215–226, Mar. 2012.

[19] S. Chan *et al.*, "Assessment of the SMAP passive soil moisture product," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4994–5007, Aug. 2016.

[20] A. Colliander *et al.*, "Validation of smap surface soil moisture products with core validation sites," *Remote Sens. Environ.*, vol. 191, pp. 215–231, Mar. 2017.

[21] Y. Chen *et al.*, "Evaluation of SMAP, SMOS, and AMSR2 soil moisture retrievals against observations from two networks on the Tibetan Plateau," *J. Geophys. Res. Atmos.*, vol. 122, no. 11, pp. 5780–5792, Jun. 2017.

[22] C. Li *et al.*, "The evaluation of SMAP enhanced soil moisture products using high-resolution model simulations and in-situ observations on the Tibetan Plateau," *Remote Sens.*, vol. 10, no. 4, Mar. 2018, Art. no. 535.

[23] C. Cui *et al.*, "Soil moisture mapping from satellites: An intercomparison of SMAP, SMOS, FY3B, AMSR2, and ESA CCI over two dense network regions at different spatial scales," *Remote Sens.*, vol. 10, no. 1, Jan. 2018, Art. no. 33.

[24] C. Cammalleri, J. V. Vogt, B. Bisselink, and A. de Roo, "Comparing soil moisture anomalies from multiple independent sources over different regions across the globe," *Hydrol. Earth Syst. Sci.*, vol. 21, pp. 6329–6343, Dec. 2017.

[25] X. Xu, B. Shew, S. Zaman, J. Lee, and Y. Zhi, "Assessment of SMAP and ESA CCI soil moisture over the Great Lakes basin," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Waikoloa, HI, USA, Oct. 2020, pp. 4590–4593.

[26] "Read-me-first note for the release of the SMOS level 2 soil moisture data products," Expert Support Laboratory (ESL) Level 2 Soil Moisture & Array Systems Computing, Chicago, IL, USA, Nov. 15, 2017. [Online]. Available: https://earth.esa.int/documents/10174/1854503/SMOS-Level-2-Soil-Moisture-v650-release-note Accessed: Mar. 19, 2021.

[27] Algorithm Theoretical Basis Document (ATBD) for the SMOS Level 2 Soil Moisture Processor Development Continuation Project, Array Systems Computing Inc., May 1, 2017. [Online]. Available: https://earth.esa.int/documents/10174/1854519/SMOS_L2_SM_ATBD Accessed: Mar. 19, 2021.

[28] P.E. O'Neill, S. Chan, E. G. Njoku, T. Jackson, R. Bindlish, and J. Chaubell, "SMAP enhanced L3 radiometer global daily 9 km EASE-grid soil moisture, version 4," NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, CO, USA, 2020. [Online]. Available: https://doi.org/10.5067/NJ34TQ2LFE90 Accessed: Feb. 1, 2021.

[29] A. Pasik *et al.*, "ESA climate change initiative plus soil moisture: Algorithm theoretical baseline document (ATBD) supporting product version v05.2," D2.1 Version 1, May 29, 2020. [Online]. Available: https://climate.esa.int/media/documents/ESA_CCI_SM_RD_D2.1_v1_ATBD_v05.2.pdf Accessed: Mar. 29, 2021.

[30] R. van der Schalie, W. Preimesberger, A. Pasik, T. Scanlon, and R. Kidd, "ESA climate change initiative plus soil moisture: Soil moisture ECV product user guide (PUG) supporting product version v05.2," D4.2 Version 1, May 29, 2020. [Online]. Available: https://climate.esa.int/media/documents/ESA_CCI_SM_RD_D4.2_v1_Product_Users_Guide_v05.2.pdf Accessed: Mar. 29, 2021.

[31] X. Xu, "Evaluation of SMAP level 2, 3, and 4 soil moisture datasets over the Great Lakes region," *Remote Sens.*, vol. 12, no. 22, Nov. 2020, Art. no. 3785.

[32] X. Xu *et al.*, "Assimilation of SMOS soil moisture over the Great Lakes basin," *Remote Sens. Environ.*, vol. 169, pp. 163–175, Nov. 2015.

[33] D. Entekhabi, R. H. Reichle, R. D. Koster, and W. T. Crow, "Performance metrics for soil moisture retrievals and application requirements," *J. Hydrometeorol.*, vol. 11, no. 3, pp. 832–840, Jun. 2010.

[34] P. E. O'Neill, R. Bindlish, S. Chan, J. Chaubell, E. G. Njoku, and T. Jackson, "SMAP algorithm theoretical basis document: Level 2 & 3 soil moisture (passive) data products," JPL D-66480, NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, CO, USA, 2020. [Online]. Available: https://nsidc.org/data/smap/technical-references. Accessed on: Feb. 1, 2021.

[35] J. Chaubell *et al.*, "Improving brightness temperature measurements near coastal areas for SMAP," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4578–4588, Nov. 2019.

[36] H SAF, "ASCAT surface soil moisture climate data record v5 12.5 km sampling - Metop," EUMETSAT SAF on Support to Operational Hydrology and Water Management, 2020, doi: 10.15770/EUM_SAF_H_0006.

[37] R. H. Reichle and R. D. Koster, "Bias reduction in short records of satellite soil moisture," *Geophys. Res. Lett.*, vol. 31, no. 19, Oct. 2004, Art. no. L19501.

[38] X. Xu, B. A. Tolson, J. Li, and B. Davison, "Comparison of X-band and L-band soil moisture retrievals for land data assimilation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 3850–3860, Sep. 2017.

[39] X. Xu, B. A. Tolson, J. Li, and B. Davison, "Assimilation of synthetic remotely-sensed soil moisture in environment Canada's MESH model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1317–1327, Apr. 2017.

[40] X. Xu *et al.*, "Assimilation of SMOS soil moisture in the MESH model with the ensemble Kalman filter," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.* Jul. 2014, pp. 3766–3769.
[41] S. Xu *et al.*, "Investigating groundwater-lake interactions in the Laurentian Great Lakes with a fully-integrated surface water-groundwater model," *J. Hydrol.*, vol. 594, Mar. 2021, Art. no. 125911.
[42] A. R. Erler *et al.*, "Evaluating climate change impacts on soil moisture and groundwater resources within a lake-affected region," *Water Resour. Res.*, vol. 55, no. 10, pp. 8142–8163, Oct. 2019.
[43] W. R. Peltier, M. d'Orgeville, A. R. Erler, and F. Xie, "Uncertainty in future summer precipitation in the Laurentian Great Lakes basin: Dynamical downscaling and the influence of continental scale processes on regional climate change," *J. Clim.*, vol. 31, no. 7, pp. 2651–2673, Apr. 2018.

Xiaoyong Xu received the B.Sc. and M.Sc. degrees in atmospheric physics from Nanjing Institute of Meteorology, Nanjing, China, in 1999 and 2002, respectively, and the Ph.D. degree in geography from the University of Waterloo, Waterloo, ON, Canada, in 2015.

From 2006 to 2007, he was a Research Associate with the NOAA National Severe Storms Laboratory and the Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK, USA. From 2007 to 2011, he was a Research Fellow with the University of Saskatchewan, Saskatoon, SK, Canada. From 2015 to 2016, he was a Postdoctoral Fellow with the University of Waterloo. From 2016 to 2018, he was a Postdoctoral Scientist with Aquanty Inc., Waterloo, Canada. Since 2018, he has been an Assistant Professor with the Department of Chemical and Physical Sciences, University of Toronto Mississauga, Mississauga, ON, Canada. His broad research interests lie in hydrological and atmospheric sciences, including hydrological modeling, hydrologic remote sensing, land data assimilation, atmospheric waves and coupling processes, convective-scale data assimilation, and radar meteorology.

Steven K. Frey received the Ph.D. degree in earth sciences from the Department of Earth and Environmental Science, University of Waterloo, Waterloo, ON, Canada, in 2011.

He is currently a Senior Scientist with Aquanty Inc., Waterloo, ON, Canada, and an Adjunct Assistant Professor with the Department of Earth and Environmental Sciences, University of Waterloo. His current research focuses on the development and application of fully integrated groundwater–surface water simulation models for climate change, land use change, and extreme weather impact analysis, as well as for hydrologic risk mitigation strategy evaluation. Much of his research has an agricultural focus or agricultural endpoints associated with it, and he works extensively with agricultural stakeholder groups to communicate how state-of-the-art simulation tools are contributing to society's better understanding of hydrologic functioning in agricultural watersheds. He has ongoing modeling-focused research in all three of Canada's Prairie provinces, as well as Ontario, and as such he is well versed on the hydrologic variability across much of Canada's most productive agricultural regions.