

Validation of SMAP Soil Moisture at Terrestrial National Ecological Observatory Network (NEON) Sites Show Potential for Soil Moisture Retrieval in Forested Areas

Edward Ayres , Andreas Colliander , *Senior Member, IEEE*, Michael H. Cosh, *Senior Member, IEEE*, Joshua A. Roberti, Sam Simkin, and Melissa A. Genazzio

Abstract—Soil moisture influences forest health, fire occurrence and extent, and insect and pathogen impacts, creating a need for regular, globally extensive soil moisture measurements that can only be achieved by satellite-based sensors, such as NASA’s soil moisture active passive (SMAP). However, SMAP data for forested regions, which account for ~20% of land cover globally, are flagged as unreliable due to interference from vegetation water content, and forests were underrepresented in previous validation efforts, preventing an assessment of measurement accuracy in these biomes. Here we compare over twelve thousand SMAP soil moisture measurements, representing 88 site-years, to *in situ* soil moisture measurements from forty National Ecological Observatory Network (NEON) sites throughout the US, half of which are forested. At unforested NEON sites, agreement with SMAP soil moisture (unbiased RMSD: $0.046 \text{ m}^3 \text{ m}^{-3}$) was similar to previous sparse network validations (which include inflation of the metric due to spatial representativeness errors). For the forested sites, SMAP achieved a reasonable level of accuracy (unbiased RMSD: $0.06 \text{ m}^3 \text{ m}^{-3}$ or $0.053 \text{ m}^3 \text{ m}^{-3}$ after accounting for random representativeness errors) indicating SMAP is sensitive to changes in soil moisture in forest ecosystems. Moreover, we identified that both an index of vegetation water content and canopy height were related to mean difference (MD), which incorporates measurement bias and representativeness bias, and suggests a potential approach to improve SMAP algorithm parameterization for forested regions. In addition, expanding the number and extent of soil moisture measurements at forested validation sites would likely further reduce MD by minimizing representativeness errors.

Index Terms—*In situ* satellite validation, National Ecological Observatory Network (NEON), soil moisture active passive (SMAP), soil water content.

Manuscript received May 26, 2021; accepted October 7, 2021. Date of publication October 19, 2021; date of current version November 9, 2021. This work was supported by the U.S. Department of Agriculture, Agricultural Research Service and Battelle IRAD100543. (*Corresponding author: Edward Ayres.*)

Edward Ayres, Joshua A. Roberti, Sam Simkin, and Melissa A. Genazzio are with the Battelle, National Ecological Observatory Network, Boulder, CO 80301 USA (e-mail: eayres@battelleecology.org; jroberti@battelleecology.org; sssimkin@battelleecology.org; mgenazzio@battelleecology.org).

Andreas Colliander is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 USA (e-mail: andreas.colliander@jpl.nasa.gov).

Michael H. Cosh is with the USDA Agricultural Research Service, Beltsville Agricultural Research Center, Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705 USA (e-mail: michael.cosh@usda.gov).

Digital Object Identifier 10.1109/JSTARS.2021.3121206

I. INTRODUCTION

SATELLITE calibration and validation efforts have been well documented in recent years with the validation of the AMSR-E [1], SMOS [2], and SMAP missions [3]. This has led to the development of protocols [4] and practices [5] for satellite validation of soil moisture from a variety of sources. Many of the validation activities for passive soil moisture remote sensing are focused on densely instrumented watersheds, but there are also many sparse networks which can provide a validation source, often over a variety of landscapes [6]. Unfortunately, there are limitations to the accuracy of these networks due to the limited number of samples at these footprints [7]–[9]. A 33 km microwave satellite footprint would ideally have many representative measurement sites, but so few networks exist with these large numbers [10], that it is necessary to use large soil sparse networks with a greater variety of landscapes to aid in the interpretation of landscape on remote sensing signals.

The National Ecological Observatory Network (NEON)¹ was recently established by the National Science Foundation to monitor drivers of, and responses to, change in US ecosystems over decadal timescales. It is a distributed network consisting of 81 aquatic (34) and terrestrial (47) sites throughout the US, including Alaska, Hawaii, and Puerto Rico, where co-located measurements are made using standardized protocols and sensors with all data made freely available (see Fig. 1). Across the Observatory over 170 unique data products are generated via three collection systems: remote sensing aircraft, automated sensors (including *in situ* soil moisture sensors), and observational sampling by field crews. NEON measurements span a wide range of taxa, including plants, birds, beetles, and microbes, as well as many aspects of biogeochemistry, ecohydrology, meteorology, and soil ecology. In addition to providing data, NEON archives approximately one hundred thousand physical samples annually which are available for community use, provides physical infrastructure to support external research projects, and generates outreach material to enhance learning and diversity in ecology and related disciplines.

¹[Online]. Available: <https://www.neonscience.org>

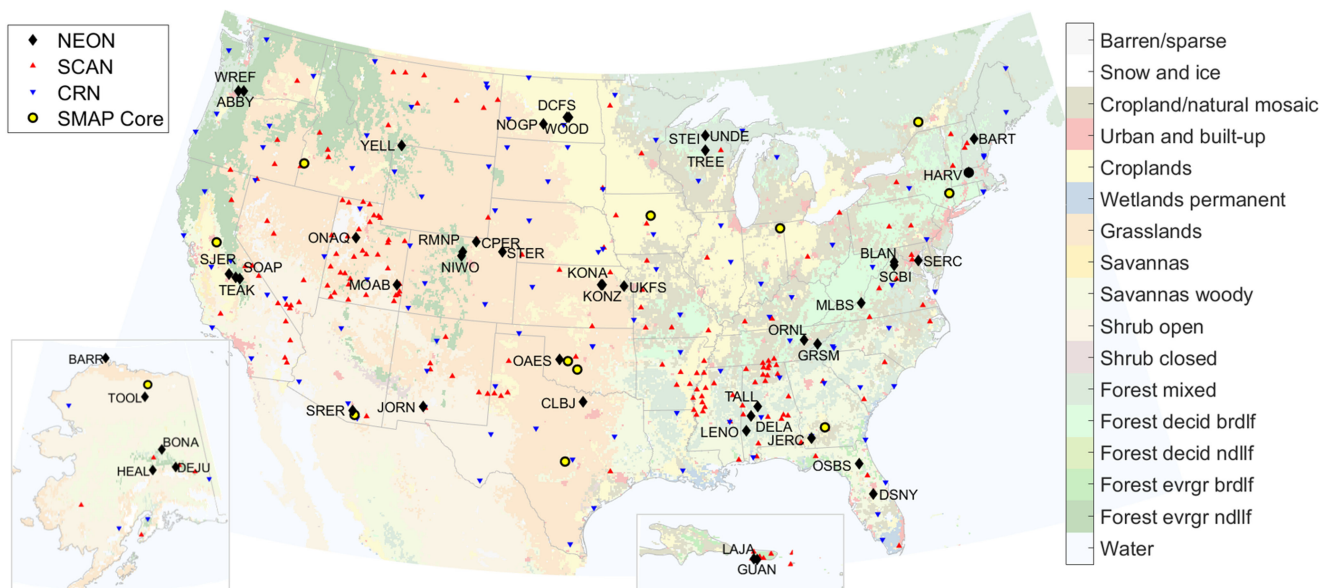


Fig. 1. NEON terrestrial sites along with SMAP core and candidate validation sites and two large networks that comprise part of the sparse validation network (SCAN and CRN) over MODIS-based IGBP land cover.

The NASA SMAP mission was launched in late January 2015 [11]. Soil moisture products are available from the start of science measurements on March 31, 2015 to the present [12]–[17]. The mission carries *L*-band radiometer and radar instruments. The radar ceased operation in July 2015, but the radiometer continues to function as expected. The capability of the mission to measure soil moisture relies on the sensitivity of *L*-band brightness temperature to soil moisture and its relative insensitivity to surface roughness and vegetation [18].

For determining the accuracy and improving the performance of the products, the SMAP mission developed a calibration and validation plan, which employs a suite of complementary methodologies to achieve a robust global assessment [19], [3]. These methodologies include the utilization of core validation sites (e.g., [20]), sparse networks (e.g., [6]), other satellite data products (e.g., [21]), model-based data products (e.g., [22]), and field campaigns [23], [24], [25]. The core validation sites have soil moisture measurement networks that consist of multiple stations within the SMAP resolution cell, which allows the estimation of the area-average soil moisture matching that measured by SMAP [20]. The inherently relatively large spatial variability of soil moisture makes it more problematic to use single location measurements (sparse networks) to estimate the area-average soil moisture [7]. Therefore, the mission requirements are tested with respect to the mean of the unbiased root-mean-square difference (ubRMSD) determined with respect to the core validation sites [11]. The sparse networks play an important role in the SMAP validation, however, because the core validation sites cover only a limited set of land covers and climate conditions whereas sparse networks expand over larger areas. In particular, the estimation of absolute soil moisture of an area requires a significant number of measurements. The ubRMSD and correlation are less susceptible to the representativeness error although they too suffer from them to some degree [6], [7]. In particular, in

tracking the consistency and changes between different products and algorithm versions sparse networks are useful. Furthermore, despite the representativeness errors, the data are useful for exploring the performance in the conditions that are not covered by core validation sites and the conditions outside of the baseline retrieval domain (dense vegetation, mountainous and snow- and ice-covered areas).

The sparse network observations employed by the SMAP validation include soil climate analysis network (SCAN by US Department of Agriculture) in the US, climate reference network (CRN by National Oceanic and Atmospheric Administration) in the US, Oklahoma Mesonet, [Comisión Nacional de Actividades Espaciales (CONAE) network on Pampas in Argentina, SMOS-Mania in France, a network in Mongolian grasslands [26], plate boundary observatory (PBO)] H₂O network in the Western US using GPS reflectometry, and COSMOS network using cosmic ray sensors. These networks have over 400 stations combined covering large areas but focus mostly on low to moderately vegetated regions. Therefore, their utility for exploring the expansion of the SMAP validation domain to high biomass areas is limited. Another common feature is that the sites have relatively little information on the vegetation surrounding the station. Finally, most sites come with only one sensor per depth (CRN has a triplet of sensors installed within meters from each other) providing no additional information on the soil moisture variability at the site.

The NEON soil moisture data have several attributes that make them particularly suitable for satellite ground validation. First, the sites span a wide range of ecosystems, including many forested sites, which were underrepresented in previous SMAP ground validation assessments [12], [27]. Second, a wide geographic distribution of sites spanning most of the USA increases confidence that the findings from the ground validation may be broadly applicable. Third, multiple within-site soil moisture measurement locations spaced tens of meters apart allowing

some local-scale variation in water content to be captured, albeit at a scale substantially smaller than the SMAP footprint. Lastly, a wide range of open-access co-located data generated using standardized protocols, including data relevant to factors that influence SMAP's data quality, such as aboveground plant biomass and vegetation water content, as well as many soil physical and chemical properties that influence soil water content.

Here we compare multiyear NEON and SMAP soil moisture data across a wide range of sites throughout the US and calculate correlation parameters to assess the suitability of incorporating NEON soil moisture data in the SMAP validation network. This was done for forty sites, as well as separately for forested (19) and nonforested (21) sites, with the expectation that correlations would be better in non-forested sites, but that SMAP soil moisture measurements in forested regions (which are currently flagged as unreliable due to high vegetation water content) still contain useful information. Finally, we explore trends in the correlation parameters in relation to ecosystem properties that are expected to influence the quality of SMAP data, namely aboveground biomass, an index of vegetation water content, and canopy height, as well as soil properties that can influence soil moisture.

II. SOIL WATER CONTENT DATA PRODUCTS

A. National Ecological Observatory Network

Soil moisture is measured by sensors in five instrumented soil plots located near the NEON tower at each of the 47 terrestrial sites. The location of the NEON soil plot closest to the tower (typically ~ 10 – 30 m from the tower to access to power and communication) was selected by NEON staff to be in the locally dominant soil type immediately surrounding the tower and within, or as close as possible, to the tower-based eddy covariance airshed, which is determined by the predominant wind direction at each site (e.g., [28]). Human biases in the siting of the remaining soil plots were minimized by positioning them along a transect with a direction close to the middle of the tower airshed direction and with the distance between plots based in part on an assessment of spatial variability of soil temperature and soil moisture at each site to maximize spatial independence [29]. Where necessary, soil plot locations were adjusted to avoid obstacles (e.g., footpaths, boulders, streams) and a more compact layout was used if it was constrained by the size of the permitted area. To further minimize the potential for human biases, the location of the soil moisture measurements within each soil plot was determined using a predefined layout with an ordered set of installation locations [30]. A single vertical profile of soil moisture is measured in each of the five soil plots at depths down to 2 m or restrictive feature if the soil is shallower. Soil moisture is typically measured at eight depths (fewer if constrained by soil depth) with the three shallowest sensors installed at nominal depths of 6, 16, and 26 cm below the soil surface, while installation depths for deeper sensors vary among sites based on soil horizon thicknesses. Measurements are made every 10 s using Sentek EnviroSCAN TriSCAN sensors, a capacitance sensor that integrates its measurements over a volume of soil that extends ± 5 cm vertically and 14 cm horizontally from the

sensor's midpoint, with the majority of the signal determined by soil within a few centimeters of the sensor [31]. The sensors come with a set of default calibration coefficients; however, the manufacturer recommends the sensors be calibrated to local soil type to mitigate measurement uncertainties. A rigorous soil moisture calibration framework was developed for the entire Observatory to complete this task [32] with the goal of limiting measurement uncertainties to $\leq 0.03 \text{ m}^3 \text{ m}^{-3}$. Sensor data from all sites are processed using standardized algorithms that calculate soil water content, perform quality assurance and quality control (QA/QC) tests, and generate 1- and 30-min averages that are made freely available for download on a site-month basis from the NEON data portal with a latency of 15–45 days (NEON data product ID: DP1.00094.001).² Each datum is published with a final quality flag based on the QA/QC tests to summarize its status (0 = datum considered trustworthy and 1 = potentially suspect), with additional detailed data quality metrics available in the “expanded” data package.

B. SMAP

The SMAP brightness temperature (TB) measurements have a 38-km resolution (defined by the half-power footprint on the Earth's surface of the radiometer antenna pattern), and the measurements filter out radio frequency interference [33], [34], [35]. The mission introduced an enhanced TB processing after the radar malfunction that uses the Backus–Gilbert approach to sample the measurements on the 9-km EASEv2 grid [36]. The soil moisture product based on the enhanced TB product is also sampled onto the 9-km grid [13]. Since the spatial resolution of the TB measurement is considerably larger than the 9-km spacing of the sampling grid, the TB from a given 9-km grid cell is inverted into a soil moisture estimate using ancillary data and parameters for a 33-km aggregation domain centered on the 9-km grid cell, thereby approximating the spatial resolution of the TB measurement. The radiometer-based product provides the soil moisture from three algorithms: single channel vertical polarization (SCA-V), single channel horizontal polarization (SCA-H) and dual channel algorithms (DCAs) [37].

III. METHODS

Soil moisture sensor data (NEON data product ID: DP1.00094.001) were downloaded from the NEON data portal using the `neonUtilities` R package [38] between March 24, and April 8, 2021. Only data from the shallowest soil moisture sensor (nominal depth 6 cm) were used because the SMAP level 2 enhanced radiometer-based product (L2SMPE) used here measures surface soil moisture (0–5 cm) [37]. Since the sensors were initially installed with the manufacturer's default calibration coefficients prior to being updated with soil-specific calibration coefficients, raw sensor measurements were back-calculated prior to applying a soil-specific calibration that prevented implausible soil moistures (i.e., $> 0.6 \text{ m}^3 \text{ m}^{-3}$) to all data to generate a consistent time series at each measurement location

²[Online]. Available: <https://data.neonscience.org/data-products/DP1.00094.001>

TABLE I
SITE INFORMATION FROM <https://www.neonscience.org/Field-Sites/Explore-Field-Sites>

Site ID	Latitude (°)	Longitude (°)	Elevation (m)	MAT (°C) ^a	MAP (mm) ^a	Soil subgroup
ABBY	45.76244	-122.33	365	10	2451	Andic Humudepts
BART	44.06389	-71.2874	274	6.2	1325	Aquic Haplorthods
BLAN	39.0337	-78.0418	183	12.1	983	Ultic Hapludalfs
CLBJ	33.40123	-97.57	272	17.5	926	Udic Paleustalfs
CPER	40.81554	-104.746	1654	8.6	344	Aridic Argiustolls
DCFS	47.16165	-99.1066	575	4.9	490	Typic Haplustolls
DEJU	63.88112	-145.751	517	-3	305	Typic Haplocryepts
DELA	32.54173	-87.8039	25	17.6	1372	Aquic Paleudults
DSNY	28.12505	-81.4362	20	22.5	1216	Aeric Alaquods
GRSM	35.68896	-83.502	575	13.1	1375	Typic Humudepts
HARV	42.53691	-72.1727	348	7.4	1199	Oxyaquic Dystrudepts
JERC	31.19484	-84.4686	47	19.2	1308	Arenic Kandiodults
JORN	32.59069	-106.843	1324	15.7	271	Typic Petrocalcids
KONA	39.11045	-96.6129	323	12.7	850	Pachic Vertic Argiudolls
KONZ	39.10077	-96.5631	414	12.4	870	Pachic Udertic Argiustolls
LENO	31.85386	-88.1612	13	18.1	1386	Vertic Epiaquepts
MLBS	37.37831	-80.5248	1170	8.8	1227	Fluvaquents
MOAB	38.24828	-109.388	1799	10.1	319	Ustic Haplocalcids
NIWO	40.05425	-105.582	3490	0.3	1005	Typic Haplocryolls
NOGP	46.76972	-100.915	589	5.9	457	Typic Argiustolls
OAES	35.4106	-99.0588	519	15.5	779	Lithic Haplustepts
ONAQ	40.1776	-112.452	1662	9	288	Xeric Haplocalcids
ORNL	35.96413	-84.2826	344	14.4	1340	Typic Paleudults
OSBS	29.68928	-81.9934	46	20.9	1302	Typic Quartzipsamments
RMNP	40.2759	-105.546	2742	2.9	731	Ustic Haplocryolls
SCBI	38.89293	-78.1395	352	11.6	1126	Ultic Hapludalfs
SERC	38.89013	-76.56	33	13.6	1075	Aquic Hapludults
SJER	37.10878	-119.732	400	16.4	540	Psammentic Haploxerolls
SOAP	37.03337	-119.262	1210	13.4	900	Ultic Haploxeralfs
SRER	31.91068	-110.835	997	19.3	346	Typic Torrifluvents
STEI	45.50894	-89.5864	476	4.8	797	Alfic Epiaquods
STER	40.46189	-103.029	1365	9.7	433	Pachic Argiustolls
TALL	32.95047	-87.3933	166	17.2	1383	Typic Hapludults
TEAK	37.00583	-119.006	2149	8	1223	Pachic Humixerepts
TREE	45.49369	-89.5857	467	4.8	797	Alfic Haplorthods
UKFS	39.04043	-95.1922	322	12.7	990	Pachic Argiudolls
UNDE	46.23391	-89.5373	521	4.3	802	Argic Fraguaquods
WOOD	47.1282	-99.2413	591	4.9	494	Typic Haplustolls
WREF	45.82049	-121.952	351	9.2	2225	Typic Hapludands
YELL	44.95348	-110.539	2133	3.4	493	Pachic Argiustolls

^aMAT and MAP represent mean annual temperature and mean annual precipitation, respectively.

[32]. A final quality flag was assigned to each datum based on the data product quality metrics and only data with a final quality flag of 0 (i.e., considered trustworthy) were used in the comparison to SMAP soil moisture data. From the SMAP L2SMPE product the 9-km grid cell closest to each of the NEON sites was selected. The SMAP quality flags were extracted along with the data, but not applied automatically to allow investigation of areas where they are typically triggered (e.g., forested areas).

Some sites were excluded from analysis due to limited availability of high-quality data, including ocean impact on SMAP measurements (NEON sites GUAN and LAJA in Puerto Rico), insufficient coincidental SMAP and NEON data (BARR, BONA and HEAL in Alaska), unavailability of NEON soil moisture data (PUUM), or frost heaving causing in-situ sensor depths to change throughout the time series (TOOL, HEAL, BARR, and BONA in Alaska). This left 40 of the 47 NEON terrestrial sites for further analysis. The remaining sites were diverse (see Table I), spanning a wide range of mean annual temperatures (-3 to 23°C), mean annual precipitations (271–2451 mm), elevations

(13–3490 m), latitudes (28 – 64°N), and nine soil orders (Alfisols, Andisols, Aridisols, Entisols, Gelisols, Inceptisols, Mollisols, Spodosols, Ultisols; [39]).³ The NEON soil moisture measurements from the different soil plots were averaged to determine a site average, similar to previous SMAP comparisons to sites with more than one measurement location [6]. To avoid the introduction of noise in the averaged NEON time series caused by measurements from soil plots being included/excluded at any given time point based on their final quality flag, average NEON soil moisture values were only produced when data from all of the soil plots with the greatest data availability were available. The list of soil plots used to create the average NEON time series are shown in Table II.

The NEON measurements closest to the SMAP overpass times were picked for establishing matchup pairs with NEON and SMAP soil moisture records. The following metrics were

³[Online]. Available: <https://www.neonscience.org/field-sites/explore-field-sites>

TABLE II
NEON SOIL PLOTS THAT WERE INCLUDED IN THE MEAN SOIL MOISTURE
CALCULATION BASED ON DATA AVAILABILITY

Site ID	No. stations	NEON soil plot number(s)
ABBY	3	3,4,5
BART	2	2,4
BLAN	5	1,2,3,4,5
CLBJ	3	3,4,5
CPER	5	1,2,3,4,5
DCFS	4	2,3,4,5
DEJU	3	1,2,3
DELA	4	1,2,4,5
DSNY	5	1,2,3,4,5
GRSM	3	1,2,5
HARV	3	1,4,5
JERC	3	1,2,3
JORN	5	1,2,3,4,5
KONA	4	1,2,4,5
KONZ	1	2
LENO	2	4,5
MLBS	2	1,5
MOAB	3	1,3,5
NIWO	3	2,4,5
NOGP	4	2,3,4,5
OAES	5	1,2,3,4,5
ONAQ	2	4,5
ORNL	3	1,3,5
OSBS	5	1,2,3,4,5
RMNP	5	1,2,3,4,5
SCBI	3	1,2,4
SERC	2	1,2
SJER	2	1,3
SOAP	4	1,2,3,5
SRER	4	2,3,4,5
STEI	4	1,2,3,5
STER	3	1,4,5
TALL	3	1,4,5
TEAK	2	2,5
TREE	2	3,5
UKFS	3	1,3,5
UNDE	2	3,4
WOOD	4	2,3,4,5
WREF	3	3,4,5
YELL	2	3,5

computed for each NEON site as defined in [41]: RMSD, unbiased RMSD (ubRMSD), mean difference (MD), and Pearson correlation (r). RMSD is defined as

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (1)$$

where x and y represent the SMAP and NEON measurement samples for a given site, and N is the number of coincidental samples for that site; ubRMSD is defined as

$$\text{ubRMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N ((x_i - \bar{x}) - (y_i - \bar{y}))^2} \quad (2)$$

where \bar{x} and \bar{y} represent the mean of the SMAP and NEON measurements for the site; MD is defined as

$$\text{MD} = \frac{1}{N} \sum_{i=1}^N (x_i - y_i) \quad (3)$$

and r is defined as

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (4)$$

The absolute MD value (i.e., removing the sign) was also determined and the slope of the least squares fit was noted for further analysis.

A critical aspect of the interpretation of the results is understanding the spatial scale differences between the SMAP measurements and the NEON *in situ* measurements. A SMAP measurement represents an area-average soil moisture for a grid pixel of approximately 33-km on the side (see Section II-B), while the NEON sensor measurements represent soil moisture in the immediate surrounding of the sensor (see Section II-A). Soil moisture has a high spatial variability naturally, which have been studied extensively in the past (e.g., [41]–[44]). In order to measure the area-average soil moisture with point measurements, the area needs to be sampled with an adequate number of sampling locations distributed over the area (e.g., [45]). However, the relative temporal trends of a single location are usually strongly correlated with the trends over several km [7]. Therefore, for determining the bias-insensitive metrics (ubRMSD and r), the sampling requirements are lower than for determining the bias-sensitive metrics (RMSD and MD) [46].

Assuming the representativeness errors of the *in situ* measurements are independent, an averaging of the metrics over multiple measurement locations diminish their impact and allows comparisons of the bias-insensitive metrics between different products, or even subsets of locations (if the representation errors are similar in the two subsets). However, the actual ubRMSD and r values degrade due to the random and multiplicative errors [47]. The random error in the ubRMSD can be expressed as

$$\text{ubRMSD} = \sqrt{\text{ubRMSD}_T^2 + \sigma_\epsilon^2} \quad (5)$$

where ubRMSD_T is the ubRMSD computed with respect to the true soil moisture and σ_ϵ is the random error. The impact of the multiplicative error component is hard to estimate because it would require information on the distribution of the site-specific multiplicative error and the temporal variation of the true area-average soil moisture at the site. For the NEON-SMAP comparisons, the five soil moisture stations deployed at each site reduces the representativeness errors, which lessens the inflation of the metrics compared to single-location networks. However, the averaging combines also the errors in the retrieval product, which in reality are not the same for each location; the aggregate value represents the combined performance with limited traceability of the distribution of the real errors at the single-pixel scale.

The bias-sensitive errors also include a bias error that, in principle, tends to diminish with number of sites. The 95% confidence interval for the bias component of the representativeness errors can be computed as (e.g., [46])

$$\text{CI}_B = t_{0.025, N_{\text{sites}}-1} \frac{\sigma_B}{\sqrt{N_{\text{sites}}}} \quad (6)$$

where $t_{0.025, N_{\text{sites}}-1}$ is the critical value at 0.025 for the t -distribution with $N_{\text{sites}}-1$ degrees of freedom; σ_B is the site-to-site (spatial) standard deviation of the bias error, and N_{sites} is the number of sites in the network. The equation shows that the bias component of the representativeness errors does not straightforwardly diminish with the addition of measurement locations (N_{sites}), because the soil moisture variability (σ_B) tends to increase with increasing area (up to a limit) (e.g., [45]). As mentioned before, the combined performance represents the aggregate area of the SMAP pixels and does not give performance estimate at the pixel scale as the distribution of the real errors (random, multiplicative and bias in the case of bias-sensitive metrics) from site to site is unknown.

Accounting for the caveats above, we compared the performance metrics to those obtained with other sparse networks and core validation sites and also between forested and non-forested NEON sites. Since the SMAP soil moisture retrievals are affected by vegetation water content, we also gathered three independent NEON datasets that were expected to be closely related to vegetation water content. These were remotely sensed canopy water content (normalized difference infrared index, NDII; DP3.30019.001), remotely sensed plant canopy height (DP3.30015.001), and field-measured aboveground vegetation biomass (derived from the NEON ‘Herbaceous clip harvest’ data product [DP1.10023.001], the ‘Non-herbaceous perennial vegetation structure’ data product [DP1.10045.001], and the ‘Woody vegetation structure’ data product [DP1.10098.001]). NDII and plant canopy height were downloaded from the NEON API as 1 km² TIFF files for each site-year they were available. These data, which had a resolution of 1 m, were averaged to calculate the annual mean NDII and canopy height for both the 1 km² tile that contained NEON soil plot 3 and the entire extent of the remote sensing flights (~ 200 km² site⁻¹). NEON soil plot 3 is the central soil plot at most NEON sites. The annual means were averaged across all years to generate a 1 km² and site mean. In two cases, the NDII and canopy height data spanned two adjacent sites (TREE-STEI and KONZ-KONA); therefore, the site-level means are identical for each site pair, but the 1 km² means are different because soil plot 3 for each site was located within a different 1 km² tile.

Aboveground biomass was calculated by summing: dried live, aboveground biomass from herbaceous clip-harvests; dry woody biomass allometrically estimated from diameter at breast height measurements; and dry biomass allometrically estimated for other non-woody growth forms (e.g., ferns, palms, ocotillo, etc.). Cactus biomass was excluded pending more accurate allometric estimates, but this is not expected to significantly impact the results due to the absence or scarcity of cacti at most sites. Allometric equations for woody biomass tree growth forms came from [48], and wood density data required to select appropriate allometries according to [48] are from [49] and [50]. Woody biomass of shrub growth forms was estimated using taxon-specific allometries for the most common taxa (e.g., [51]–[54]), and [48] tree allometries were applied for the remaining species. Nonwoody biomass was estimated using taxon-specific allometries for ferns [55], palms [56], and ocotillo [57]; when taxon-specific allometries were not available, morphological substitutions were made, similar to the approach described by

[58]. The aboveground biomass sampling plots were chosen to be representative of the area sampled by the NEON field crews at each site (min: 3 km²; max: 215 km²; mean \pm sd: 34 \pm 34 km²).

Soil properties from the Megapit soil pit (NEON data product ID: DP1.00096.001), which was sited to be representative of the in-situ soil moisture sensor locations, were used to evaluate potential relationships to the NEON-SMAP correlation parameters. Data were accessed on April 26, 2021. Since SMAP measures soil moisture in the 0–5 cm layer, weighted averages of soil bulk density, organic carbon (C) content, 2–20 mm size rock content (data on larger rock sizes were unavailable), and sand, silt and clay content were calculated for the 0–5 cm layer based on soil horizon thicknesses. Where organic soil horizons were present, they were assigned sand, silt, and clay contents of 0% since organic soils are comprised of plant residues, rather than minerals. Soil organic C data were log₁₀ transformed while rock content data were log₁₀ transformed after adding 0.1 to eliminate values of 0 g kg⁻¹ rock content.

The SMAP L2 enhanced soil moisture retrieval algorithm uses a MODIS land cover product (MOD44) to categorize the land cover within the 33-km aggregation domain of a grid pixel for the soil moisture retrieval (see the data section) [13]. The pixel is assigned the most common International Geosphere-Biosphere Programme (IGBP) land cover class found at a 1-km resolution within the aggregation domain. The dominant land cover class used by the SMAP product (and available within the product) was applied in this article to delineate the sites in different groups.

An assessment of the spatial representativeness of the NEON sites for the surrounding regions was conducted. A summertime Landsat 8 scene was selected from a year between 2015 and 2020 as an example of the spatial distribution of vegetation. The Normalized Difference Infrared Index (Landsat NDII) was computed for a 1 km and 33 km diameter circle surrounding the NEON site. If the Landsat NDII was similar for 1 and 33 km diameter circles, it is inferred that the vegetation is homogenous across that scale.

Summary statistics for the NEON-SMAP correlation parameters were calculated among the sites. Linear regressions were fitted to explore relationships between the correlation parameters and ecosystems properties (aboveground biomass, vegetation water content index (NDII), canopy height, soil bulk density, soil organic C, rock content, and sand, silt, and clay content) using the `lm()` function in R. Kruskal-Wallis rank sum tests were used to determine if the correlation parameters differed among the landcover types. When significant differences among landcover types were identified, pairwise Wilcoxon rank sum tests with [59] false discovery corrections were used to identify pairwise differences. Links to, and information about, the data used in these analyses are in Table III.

IV. RESULTS

To investigate whether the area where NEON’s soil moisture measurements were made was representative of the area approximating the SMAP footprint scale we compared NDII of the 1 km² tile containing NEON soil plot 3 (and in most cases all the NEON soil plots) with the NDII for the entire NEON site.

TABLE III
NEON DATA PRODUCTS USED IN THE ANALYSES

NEON data product ID	NEON data product name	DOI	Date range
DP1.00094.001	Soil water content and water salinity		2014-12-01 – 2021-03-01
DP3.30019.001	Canopy water indices – mosaic (containing NDII) ^a	https://doi.org/10.6084/m9.figshare.14599968.v1	2013-06-01 – 2020-10-01
DP3.30015.001	Ecosystem structure (canopy height) ^a		2013-06-01 – 2020-10-01
DP1.10023.001	Herbaceous clip harvest	https://doi.org/10.48443/xjxw-2p18	2013-07-01 – 2019-12-31
DP1.10045.001	Non-herbaceous perennial vegetation structure	https://doi.org/10.48443/pgdv-wd42	2014-01-01 – 2019-12-31
DP1.10098.001	Woody vegetation structure	https://doi.org/10.48443/e3qn-xw47	2014-01-01 – 2019-12-31
DP1.00096.001	Soil physical and chemical properties, Megapit	https://doi.org/10.48443/rfmw-p030	2012-06-01 – 2018-07-01

a.NDII and canopy height data were not available for the ORNL site.

TABLE IV
OTHER DATA PRODUCTS USED IN THE ANALYSES

NEON data product name	DOI	Date range
SMAP Enhanced L2 Radiometer Half-Orbit 9 km EASE-Grid Soil Moisture, Version 4	https://doi.org/10.1016/j.rse.2017.08.025	2015-03-31 – 2021-03-01
Landsat 8 Surface Reflectance Data	https://doi.org/10.3133/fs20153034	2013-04-01 – 2021-03-01

TABLE V
SMAP SCA-V AND NEON SOIL MOISTURE CORRELATION PARAMETERS ACROSS ALL SITES, NONFORESTED SITES, AND FORESTED SITES (TOTALS OR MEAN \pm SD)

	All sites (n = 40)	Non-forest sites (n = 21)	Forest sites (n = 19)
Total coincidental measurements	12,881	7428	5453
Coincidental measurements per site	322 \pm 120	354 \pm 118	287 \pm 114
RMSD ($\text{m}^3 \text{m}^{-3}$)	0.112 \pm 0.065	0.078 \pm 0.035	0.150 \pm 0.069
MD ($\text{m}^3 \text{m}^{-3}$)	0.027 \pm 0.115	0.006 \pm 0.072	0.050 \pm 0.148
Absolute MD ($\text{m}^3 \text{m}^{-3}$)	0.092 \pm 0.073	0.057 \pm 0.042	0.131 \pm 0.081
ubRMSD ($\text{m}^3 \text{m}^{-3}$)	0.053 \pm 0.015	0.046 \pm 0.013	0.060 \pm 0.014
Pearson correlation coefficient, r	0.58 \pm 0.25	0.73 \pm 0.09	0.42 \pm 0.28
Slope	0.70 \pm 0.51	0.79 \pm 0.48	0.60 \pm 0.53

These data had unflagged NDII measurements covering $0.99 \pm 0.03 \text{ km}^2$ and $191 \pm 73 \text{ km}^2$ (mean \pm sd), respectively, across the sites and were strongly positively correlated [$p < 0.001$, $r^2 = 0.89$; Fig. 2(a)]. A similar analysis based on canopy height also showed that the 1 km^2 data was positively correlated with site-scale data [$p < 0.001$, $r^2 = 0.82$; Fig. 2(b)]. Since the NDII and canopy height data for the entire NEON site covered an area ($191 \pm 73 \text{ km}^2$) that was smaller than the SMAP measurement footprint (1089 km^2) we also used Landsat NDII data as they were available at the SMAP footprint scale. Landsat NDII for 1 and 33 km diameter areas surrounding the NEON sensors were also strongly positively correlated [$p < 0.001$, $r^2 = 0.89$; Fig. 2(c)]. In addition to these correlations, the dominant landcover within the SMAP footprint at each NEON site accounted for a large proportion of the total footprint (mean \pm sd: $71 \pm 21\%$), which is indicative of a relatively homogeneous area. Collectively these findings indicate that the NEON measurement locations were broadly representative of the corresponding SMAP footprints.

In total, 12 881 coincidental SMAP and NEON soil water content measurements were compared across 40 sites (mean \pm sd: 322 ± 120 measurements per site), which corresponds to over two years of SMAP data per site given its ~ 2.5 -day return interval (88 site-years collectively). Averaging correlation parameters across all sites showed NEON and SMAP soil moistures were positively correlated, with SMAP soil moisture generally wetter (positive MD) and less dynamic (slope < 1) than NEON (see Table V). The ubRMSD for the SCA-V algorithm was $0.053 \pm 0.015 \text{ m}^3 \text{m}^{-3}$ and r was 0.58 ± 0.25 (see Fig. 3), which includes a reduced amount of representation errors through the averaging, but the values are inflated with respect to the true ubRMSD performance by the random errors [see (5) in Section III]. The RMSD was substantially larger than the ubRMSD ($0.11 \pm 0.06 \text{ m}^3 \text{m}^{-3}$), because of the large site-specific

MD values, reflected by the average absolute MD (see Table V). It is noteworthy that while the average MD is close to zero, its large distribution contributes to the large RMSD; this is consistent with the fact that the representativeness errors for the bias-sensitive metrics (MD and RMSD) are particularly sensitive to the number of measurement locations [see (6) in Section III]. The relatively low average MD conceals the large site-specific values, which are due to the combination of representativeness errors and actual SMAP biases. To the degree that the representativeness errors become averaged out, the average MD value represents the SMAP performance at the 40-pixel aggregate level, but not at the SMAP pixel scale (see Section III).

Given that SMAP currently flags soil moisture data from sites with a vegetation water content $> 5 \text{ kg m}^{-2}$ ($> 50 \text{ Mg ha}^{-1}$), we recalculated these summary statistics separately for nonforested (i.e., dominant landcover class: cropland/natural mosaic, grasslands, croplands, woody savannas, open shrublands, and savannas) and forested sites (i.e., evergreen needleleaf forest, mixed forests, and deciduous broadleaf forest) since forests typically exceed this threshold while the other vegetation classes do not [62]. The bias-insensitive metrics (ubRMSD and r) were markedly better for non-forested sites as expected based on the more challenging retrieval environment. Fig. 3 shows the histograms for each metric separated in the nonforested and forested groups. There is a large difference in the distribution of r : for all nonforested sites the values are grouped close to the mean value (0.73), suggesting reasonable skill over all non-forested sites, whereas for the forested sites the values have

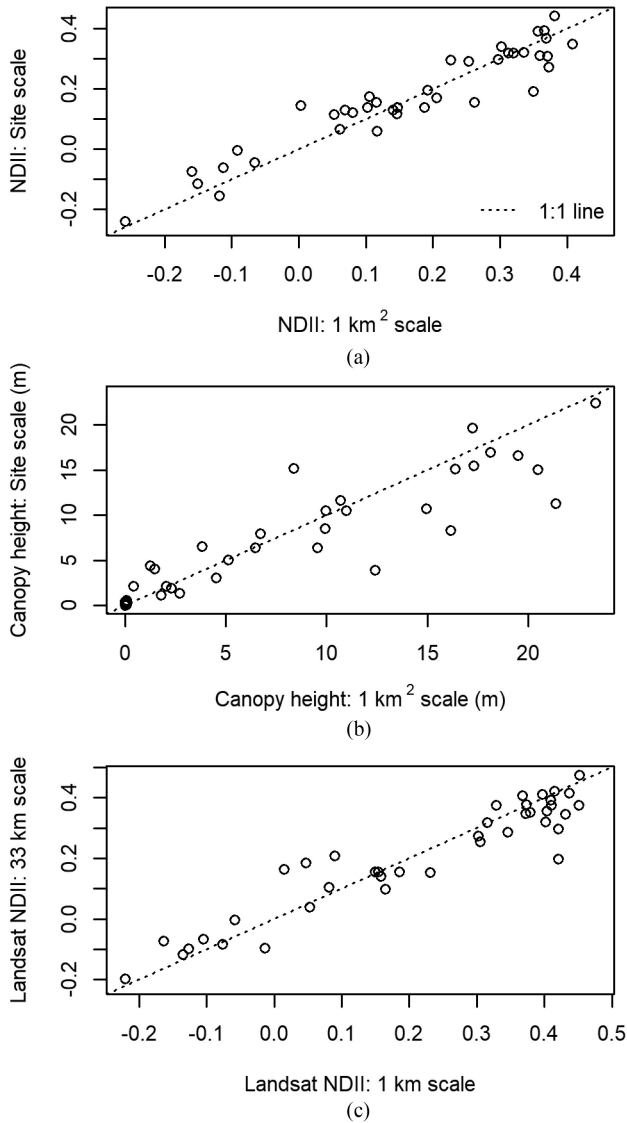


Fig. 2. Comparison of NEON NDII (a) NEON canopy height and (b) Landsat NDII (c) at the NEON sensor local scale (x -axis) and larger scale (y -axis). NDII (a) and canopy height (b) are shown at 1 km² and NEON remote sensing site scale (~ 190 km²), while Landsat NDII (c) are shown for a 1 km and 33 km (i.e., SMAP scale) diameter circle surrounding the NEON site.

a wide range with some sites having a reasonable r and a few sites close to zero, and even less than zero. As discussed in Section III, the representativeness errors for individual sites may be large; therefore, too far-reaching conclusions cannot be drawn from these results, but as such they suggest that the current SMAP soil moisture algorithm has reasonable retrieval skill over the majority of the forested sites, but it is not able to retrieve soil moisture changes over some of them. For the SCA-V algorithm, the average ubRMSD for the non-forested sites is considerably low at $0.046 \text{ m}^3 \text{ m}^{-3}$; the high ubRMSD value ($0.060 \text{ m}^3 \text{ m}^{-3}$) for the forested sites is consistent with the low r values.

The bias-sensitive metrics have also clear differences between the non-forested and forested cases. In the non-forested case, the MD values converge around the average value suggesting

somewhat reliable result for the 21-pixel aggregate MD. However, in the forested case, the MD values are relatively evenly spread between -0.21 and $0.32 \text{ m}^3 \text{ m}^{-3}$ without a convergence toward the average value suggesting that the average value does not represent an expected value. The distribution of MD affects the RMSD values correspondingly. The SCA-V and DCA have similar metrics, while the SCA-H has somewhat poorer values. These differences will be discussed more in the following Section.

Aboveground biomass, vegetation water content index (NDII), and plant canopy height are all expected to be indicative of vegetation water content, which interferes with SMAP retrievals, and exhibited several of the hypothesized impacts on the NEON-SMAP SCA-V soil moisture correlation parameters (see Fig. 4). Specifically, all three properties were positively correlated with ubRMSD [see Fig. 4(a)–(c)] and negatively correlated with r [see Fig. 4(m)–(o)], while both NDII and canopy height were also positively correlated with MD [see Fig. 4(e) and (f)] and RMSD [see Fig. 4(k) and (l)]. Absolute MD and slope were unrelated to aboveground biomass, NDII, or canopy height. The slope of the correlation between aboveground biomass and the ubRMSD indicated that for every 100 Mg ha^{-1} increase in aboveground biomass there was a $0.008 \pm 0.002 \text{ m}^3 \text{ m}^{-3}$ increase in soil moisture ubRMSD, similarly, for every 10 m increase in canopy height ubRMSD increased by $0.011 \pm 0.004 \text{ m}^3 \text{ m}^{-3}$.

Since soil properties can influence soil moisture, we examined relationships between the correlation parameters and bulk density, soil organic carbon (C) content, 2–20 mm size rocks, and soil sand, silt and clay content measured near the *in situ* soil moisture sensors. Variation in soil bulk density and organic C content was significantly related to ubRMSD and r , with ubRMSD decreasing with increasing bulk density and decreasing organic C content and r increasing with increasing bulk density and decreasing organic C content (see Fig. 5). None of the other soil properties were significantly ($p < 0.05$) related to the correlation parameters.

The dominant landcover across the sites spanned nine vegetation classes, however, only three landcovers dominated more than four sites (mixed forest: $n = 9$; grassland: $n = 8$; and evergreen needleleaf forest: $n = 7$), which limits the ability to detect differences among landcover types. Although there were significant relationships between landcover and all the correlation parameters ($p \leq 0.05$), almost all pairwise comparisons were not significant, except for mixed forests having a larger absolute MD than grasslands and a larger RMSD than both grasslands and croplands (see Fig. 6).

V. DISCUSSION

The NEON sites offer soil moisture in a diverse set of conditions, which can be exploited in the validation of the SMAP soil moisture products. The performance metrics over the non-forested sites (ubRMSD of $0.046 \pm 0.013 \text{ m}^3 \text{ m}^{-3}$ and r of 0.73 ± 0.09) were slightly better than the level achieved with the other sparse network grassland stations currently used for the SMAP validation (ubRMSD of $0.050 \text{ m}^3 \text{ m}^{-3}$ and r of 0.68; [3]). While it is difficult to confirm that such small differences in

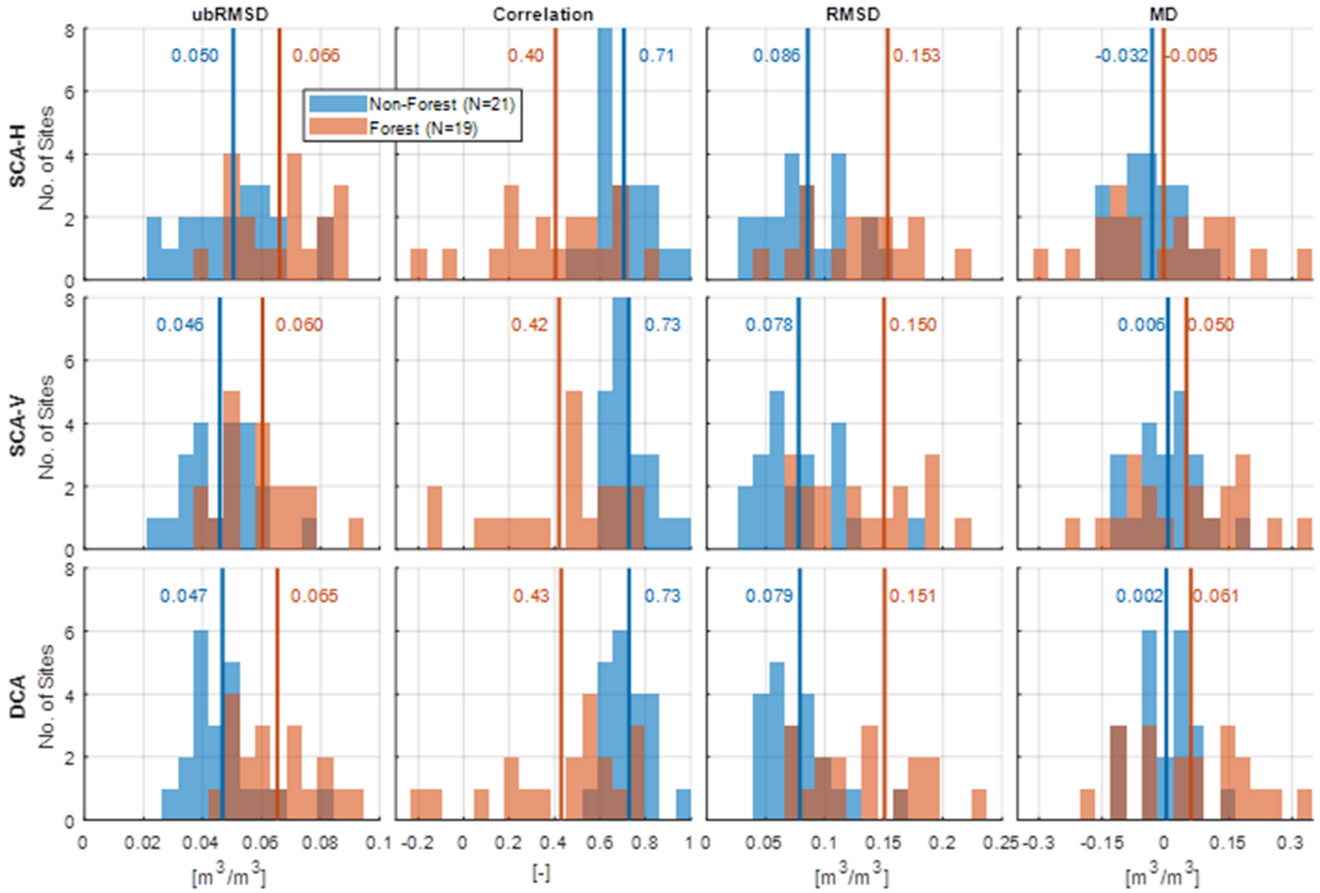


Fig. 3. Histograms of the ubRMSD, Pearson correlation (r), RMSD and MD computed for the nonforested and forested sites. The vertical lines indicate the average value which is also shown next to the lines.

ubRMSD are real (rather than noise), it could be the result of NEON's five soil moisture measurement locations spaced tens of meters apart reducing the representativeness errors, in contrast to one measurement location (SCAN and most of SMAP's sparse network sites) or three measurement locations within a few meters of one-another (USCRN). Both ubRMSD and r are subject to inflation of the true error values because of the random errors and multiplicative errors (see Section III). Consistent with these patterns, SMAP's core validation sites, which are also primarily unforested and include an even larger number (mean: 19) of more widely distributed measurement locations within the SMAP cell, had an even lower ubRMSD and higher r ($0.037 \text{ m}^3 \text{ m}^{-3}$ and 0.81 ; [3]). The RMSD for the SMAP core validation sites was lower than the non-forest NEON sites (0.078 versus $0.048 \text{ m}^3 \text{ m}^{-3}$); the difference is likely dominated by the representation errors (random, multiplicative, and bias). The MD for the SMAP core validation sites and the non-forest NEON sites was very similar (0.006 versus $-0.006 \text{ m}^3 \text{ m}^{-3}$, respectively). The core validation site comparison includes 15 sites and for the NEON comparison we included 21 nonforested sites so in both cases the SMAP soil moisture is aggregated over roughly the same number of pixels. The site-to-site representativeness biases should be significantly lower for the core validation sites; therefore, their distribution should be a

more reliable reflection of how the biases are distributed at pixel level. The NEON-based MD varies between -0.114 and $0.168 \text{ m}^3 \text{ m}^{-3}$, while the core validation-based MD varies between -0.0618 and $0.0641 \text{ m}^3 \text{ m}^{-3}$, which likely reflects the larger representation errors in the NEON comparison.

As the representativeness errors at the NEON sites can be assumed to be the same for all SMAP algorithms, the metrics allow their relative comparison despite those errors. However, the inflation of the values through the random and other representativeness errors may conceal some differences depending how large those errors are (see Section III). For SCA-H, over the non-forested sites, the ubRMSD, RMSD and the absolute value of MD are larger, and the r is lower, compared to SCA-V and DCA. This matches the relative performance of the algorithms found using the core validation sites in [3]. Assuming the random errors are on the order of $\sigma_{\varepsilon, CVS} = 0.01 \text{ m}^3 \text{ m}^{-3}$ for the core sites [46], the relationship between the core site and NEON ubRMSD values would suggest a random representativeness error of roughly $\sigma_{\varepsilon, NEON} = 0.029 \text{ m}^3 \text{ m}^{-3}$ by using (4) to solve ubRMSD_T with the core site ubRMSD and σ_{ε} , and then solving σ_{ε} for NEON. The r values are lower for each algorithm while maintaining the relative differences between the algorithms consistent with this level of representativeness error. Therefore, the results suggest that the random representativeness errors for the

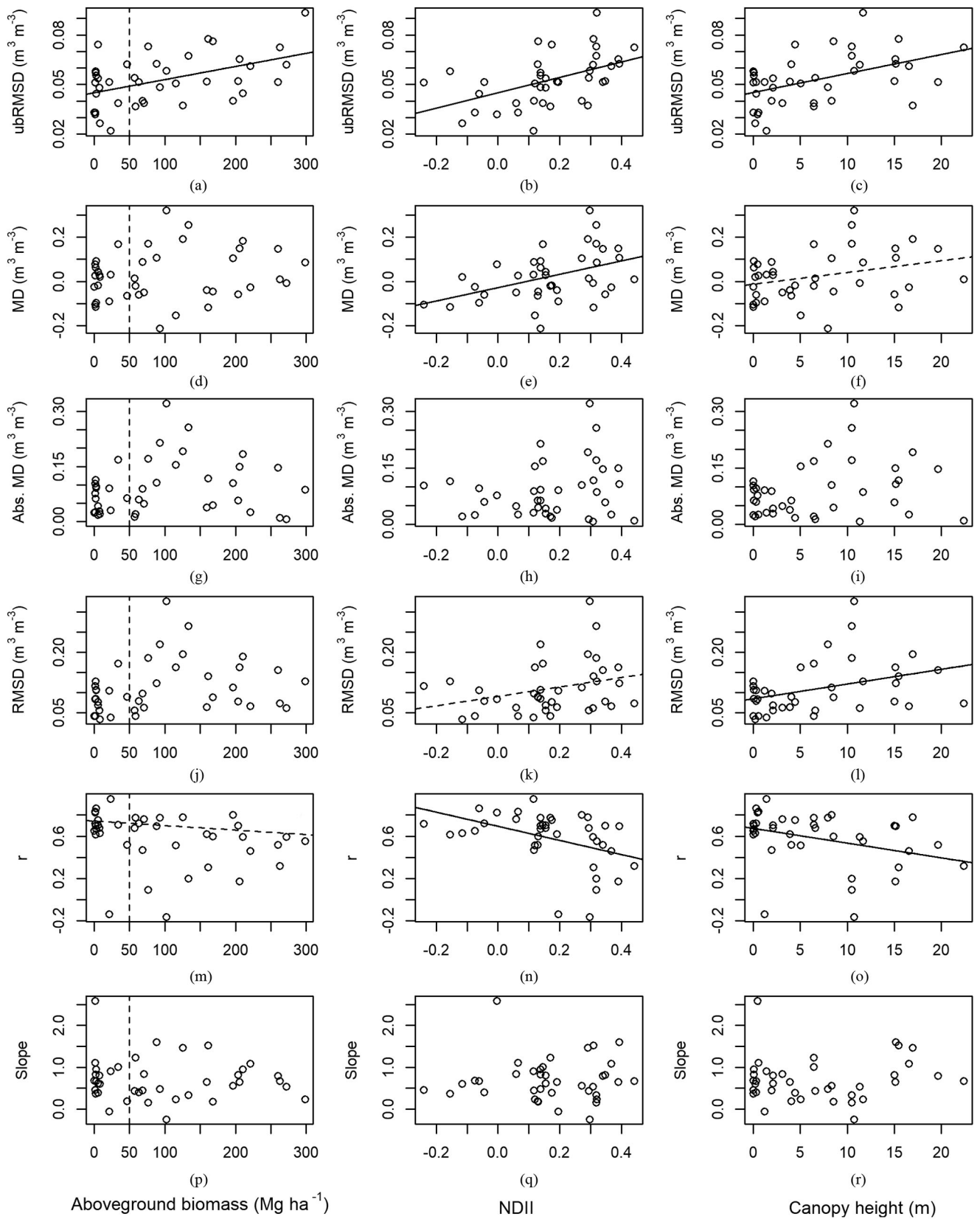


Fig. 4. Relationship between NEON-SMAP SCA-V soil water content correlation parameters and aboveground biomass (left column), canopy water content index (NDII; middle column), and canopy height (right column). Solid regression lines were significant at $p \leq 0.05$, while dashed regressions were significant at $p \leq 0.1$. Dashed vertical lines reflect the approximate upper range of unflagged SMAP data assuming 50% vegetation water content (i.e., aboveground vegetation water content of 50 Mg ha^{-1} or 5 kg m^{-2}).

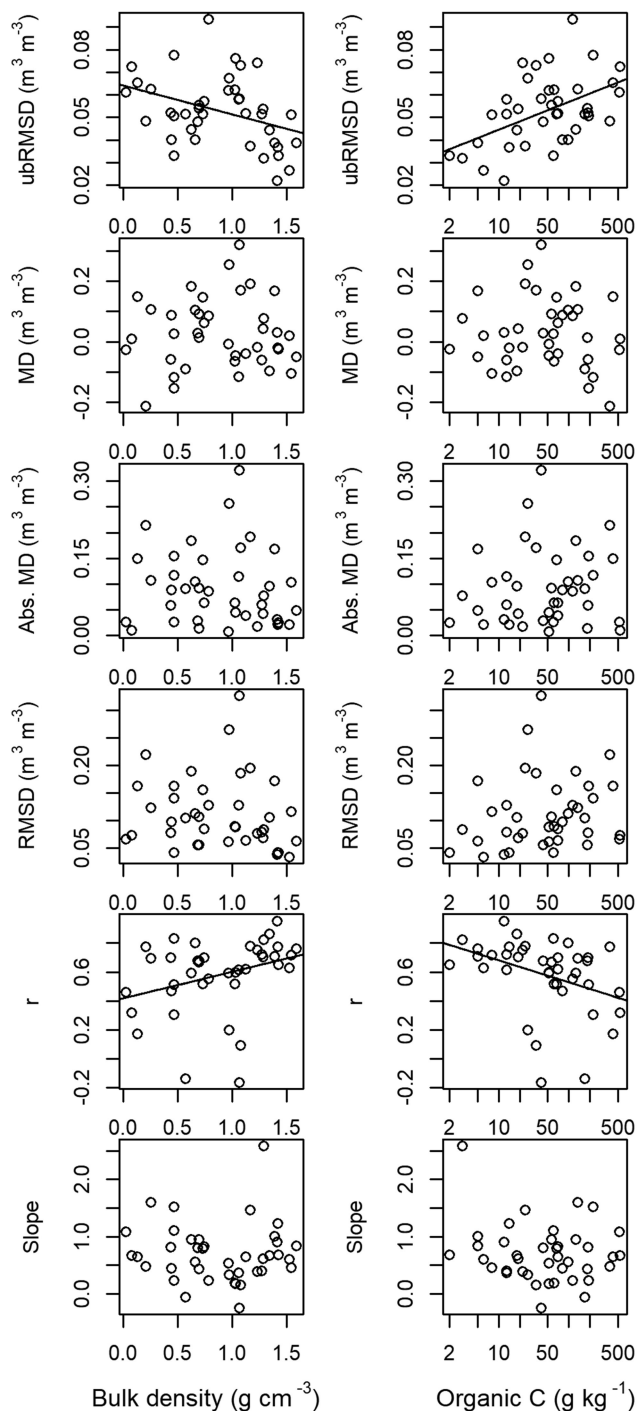


Fig. 5. Relationship between NEON-SMAP SCA-V soil water content correlation parameters and bulk density (left column) and soil organic C content (right column; log scale). Solid regression lines were significant at $p \leq 0.05$.

non-forested NEON sites are not too large to find the ubRMSD performance differences between the current SMAP algorithms.

The MD values for all algorithms are very comparable with the values obtained with the core validation sites. The differences are $0.005 \text{ m}^3 \text{ m}^{-3}$, $0.012 \text{ m}^3 \text{ m}^{-3}$ and $0.011 \text{ m}^3 \text{ m}^{-3}$ for DCA, SCA-V and SCA-H, respectively. This suggests that at the aggregate scale, the assessment based on the NEON sites is very consistent

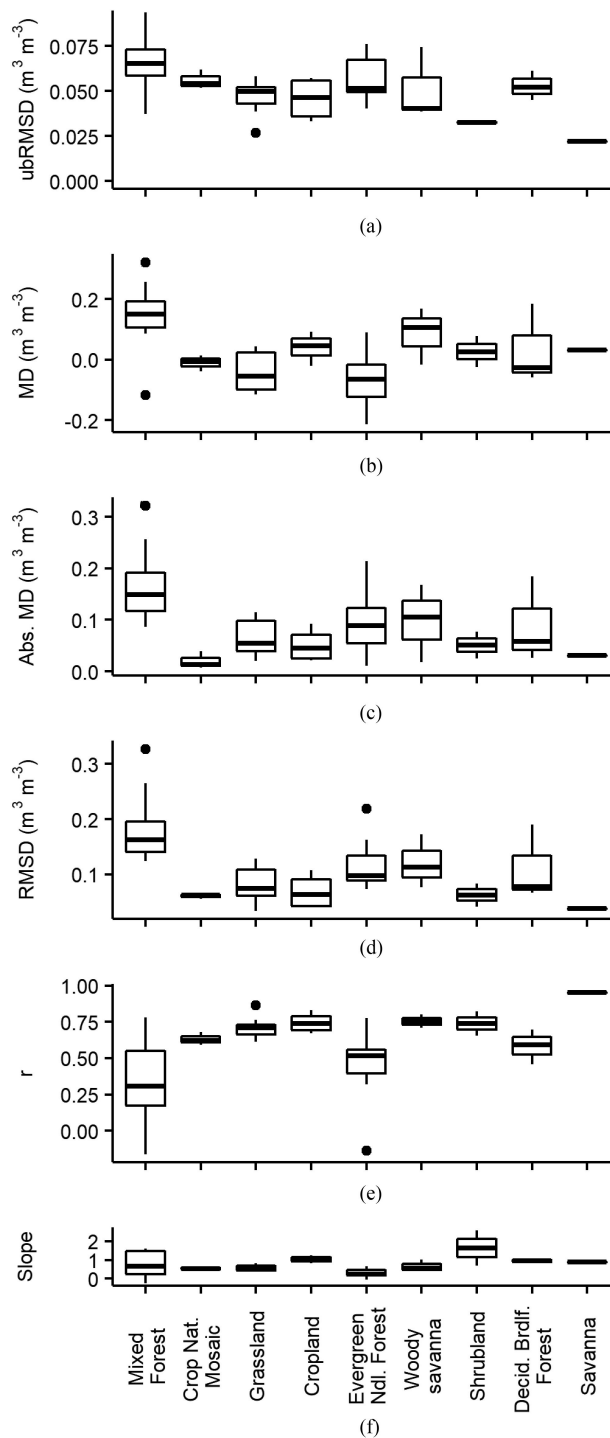


Fig. 6. Boxplots of NEON-SMAP SCA-V correlation parameters for different landcovers (bold line: median; box: interquartile range; whiskers: 1.5 times interquartile range). Number of sites within each category were: Mixed forest: 9; Crop. Nat Mosaic: 3; Grassland: 8; Cropland: 4; Evergreen Ndl. Forest: 7; Woody Savanna: 3; Shrubland: 2; Decid. Brdfl. Forest: 3; and Savanna: 1.

with the core site assessment, which further demonstrates the value of the network in comparative studies.

SMAP currently adds a flag to all soil moisture data from regions with a canopy vegetation water content $>5 \text{ kg m}^{-2}$, which includes forested areas that cover 21% of global land [61].

Given that water availability influences forest ecosystems in many ways, including tree mortality, carbon cycling, fire occurrence and extent, and the impacts of insects and pathogens [62], [63], [64], addressing this current gap could further enhance the applicability of SMAP data. While the correlation parameters were consistently worse for forested than nonforested NEON sites, the ubRMSD was still relatively small ($0.06 \text{ m}^3 \text{ m}^{-3}$) demonstrating that SMAP was still able to detect real changes in soil water content (or something correlated with it). If we assume the representativeness error is roughly the same for the forested sites as for the non-forested sites ($\sigma_{\epsilon, \text{NEON}} = 0.029 \text{ m}^3 \text{ m}^{-3}$) the true ubRMSD would be $0.053 \text{ m}^3 \text{ m}^{-3}$, see (4), which is encouraging considering that the current SMAP algorithms (version R17000) were not designed to account for forest vegetation (furthermore, it can be speculated that for the forested sites the random errors maybe larger due to the complexity of the land cover and soil composition). There were five forested sites with particularly weak correlations ($r < 0.3$; UNDE, STEI, TREE, BART, and DEJU). At three of these sites (UNDE, STEI, and TREE) the SMAP soil moisture measurements regularly saturated, which implies suboptimal parameterizations in the SMAP algorithm, while at the remaining sites there were large differences in the NEON soil moisture measurements among the soil plots, possibly representing poor sensor-soil contact in some plots and/or particularly large spatial variation in soil moisture. Addressing these issues may further improve our assessment of SMAP's ability to detect changes in soil moisture in forested regions. Our findings are consistent with [65], which also showed that SMAP does have sensitivity to soil moisture under forest canopy.

It is important to note that the RMSD for the forested sites ($0.15 \text{ m}^3 \text{ m}^{-3}$) was much larger than the ubRMSD, indicating that MD were responsible for most of the total uncertainty; the source of the large MD would need to be investigated in order to understand the SMAP performance in forested regions. Encouragingly, MD was positively correlated with both an index of aboveground vegetation water content (NDII) and, albeit at $p < 0.1$, canopy height, suggesting that at least some of the MD may be reduced with improved vegetation parameterization in the SMAP retrieval algorithm in the future. Moreover, global maps of canopy water content indices and canopy height data are already available (Landsat NDII and Global Forest Canopy Height; [65]) facilitating their potential incorporation into an updated SMAP soil moisture data product for forested regions. In any case, the low ubRMSD indicates that SMAP data could be relatively reliably used to assess temporal trends in soil moisture in forested regions, for example by calculating soil moisture percentiles to monitor drought and flood evolution and risk (e.g., [66]).

Other factors besides canopy water content indices and height likely also contribute to the MD that was observed at forested and unforested NEON sites (besides the representativeness errors) such as, soil clay content, soil surface roughness, organic content in the surface soil layers, and physical temperature biases in the modeled temperature used by SMAP, several of which could be at least partially mitigated with additional studies. MD represented a lower proportion of total RMSD

at the primarily unforested and densely instrumented SMAP core validation sites ($\sim 54\%$ for SCA-V [absolute MD: $0.026 \text{ m}^3 \text{ m}^{-3}$; RMSD: $0.048 \text{ m}^3 \text{ m}^{-3}$]; [3]) than for the sparsely instrumented sites ($\sim 80\%$ [absolute MD $0.065 \text{ m}^3 \text{ m}^{-3}$; RMSD: $0.081 \text{ m}^3 \text{ m}^{-3}$]; [12]), indicating the benefits of having a larger number of measurement locations throughout the SMAP footprint. In addition, MD at SMAP's core validation sites, which are primarily unforested, was recently shown to be negatively correlated with soil clay content [3], suggesting another possible pathway to account for biases. Although in this case we did not observe this relationship at the NEON validation sites, perhaps suggesting that either the relationship does not extend to forested sites or that due to its small magnitude it can only be observed with smaller representativeness errors. The SMAPVEX19-22 [67] and SMAPVEX22-Boreal field experiments conducted in three forested locations at the SMAP footprint scale will shed more light on the effects of the forest vegetation; spatial variability of soil moisture in forested areas at the SMAP footprint scale; soil surface roughness magnitude and variability, and the impact of the organic layer on the surface soil moisture dynamics and brightness temperature response.

It is interesting to note that ubRMSD increased linearly with aboveground biomass, NDII, canopy height, and \log_{10} (soil organic C content), and decreased linearly with bulk density, indicating that any approach to processing SMAP soil moisture data across the diverse ecosystem types in this article could be applied universally, rather than having unique processing pipelines for forested and unforested sites. That all these ecosystem properties were related to ubRMSD is not surprising, given that these properties are themselves correlated with one another: forests often have high aboveground biomass, NDII, canopy height, soil organic C and relatively low 0–5 cm bulk density, while the opposite is true of many grasslands and deserts. This makes it difficult to disentangle the underlying cause of the relationship; however, it is likely to be at least partly due to the vegetation water content, which is known to interfere with SMAP retrievals [68]. In addition, since soil porosity, which controls the upper limit of soil moisture, is closely related to soil bulk density and organic C content, it is conceivable that higher ubRMSDs and lower r at sites with low bulk density and high organic C content (i.e., high soil porosity) are simply due to the large range of soil moistures that occur at those sites allowing larger possible differences between NEON and SMAP soil moistures, while at high bulk density and low organic C (i.e., low soil porosity) sites the possible difference in NEON and SMAP soil moistures is more constrained. The relationships that we identified between the correlation parameters and various ecosystem properties also highlight the advantages of having a wide range of co-located standardized data, including in-situ sensor data, field crew generated data, and remotely sensed data, in satellite validation networks.

Based on the relationships between ubRMSD and the ecosystem properties, the minimum ubRMSD was on average $\sim 0.04 \text{ m}^3 \text{ m}^{-3}$ in low biomass, low canopy water content, short-stature, low soil organic C, and high bulk density ecosystems, and increased to $\sim 0.07 \text{ m}^3 \text{ m}^{-3}$ in high biomass, high canopy

water content, tall, high soil organic C, and low bulk density ecosystems. This trend is expected because the sensitivity of the SMAP measurements to soil moisture changes decreases with increasing vegetation water content [37]. The current SMAP algorithms have been designed for the low-biomass end, but they have not been designed for the high-biomass conditions, which presumably results in additional degradation in the high end. If we assume the absolute MD corresponds to the level found for SMAP's core validation sites (i.e., $0.026 \text{ m}^3 \text{ m}^{-3}$; [3]) these values will slightly increase to RMSDs of ~ 0.05 and $\sim 0.07 \text{ m}^3 \text{ m}^{-3}$, respectively. Notably, the low end of these estimates are similar to values observed at SMAP's core validation sites, which are primarily low biomass, low canopy water content, and short-stature ecosystems (ubRMSD: $\sim 0.037 \text{ m}^3 \text{ m}^{-3}$ and RMSD: $\sim 0.048 \text{ m}^3 \text{ m}^{-3}$ for SCA-V; [3]). At the high end, it is foreseen that future algorithm developments for high-biomass areas, partially enabled by the data provided by the NEON forest sites, will bring down the ubRMSD component and the RMSD.

Passive microwave measurements receive radiation from a surface layer of the soil with variable thickness depending on the soil moisture (which dictates the dielectric constant). Based on the penetration depth for dielectric medium the average thickness of this layer is on the order of 5 cm for L-band measurements [37]. Therefore, the validation of the SMAP soil moisture products is focused on using *in situ* sensor measurements within this layer. However, the sensors do not measure exactly the same volume of the soil column that affects the L-band radiation, which results in some inherent mismatch between the measurements (e.g., [69]). In addition to these general limitations in conducting SMAP *in situ* comparisons, there are inherent challenges in observing forested ecosystems. Passive L-band microwaves are impacted by vegetation water content (e.g., [68], [70]). This attenuation of the signal through the forest canopy makes it more difficult to detect the soil moisture signal at the forest floor. Furthermore, often at the forest floor, there is a layer of litter or duff from the forest canopy (needles, leaves, seeds, twigs, etc.) that have fallen and are in some state of decay. Some research has been conducted on the impact of this litter layer on the remote sensing signal (e.g., [71]–[73]). Soil organic matter content at the forest floor tends to be greater and has an impact that may need to be accounted for in the SMAP algorithm [74]. A fundamental limitation to the retrieval accuracy is the uncertainty of the physical temperature knowledge of not only the soil but also the vegetation, because vegetation emission contributes significantly to the received signal over forests (e.g., [67]). Other topics of concern for forest monitoring are the possibility of interception of precipitation and leaf wetness as a temporary store for water within the sensing volume [75]. The SMAP morning overpasses occur during times of high leaf wetness from overnight deposition or dew. Finally, in a heavily managed landscape, such as the continental United States, arable land is often used for agricultural purposes and NEON sites are often, but not universally, located in less intensively managed ecosystems, though there may have been agricultural or management in the past. The propensity for NEON sites to be

non-arable may also be a confounding factor for remote sensing validation.

Forested sites are currently underrepresented in the SMAP validation network. Forests account for 21% of total global land cover, but none of the core validation sites, only 11% (46 out of 416) of the initial sparse validation sites, and 1% (3 out of 389) of the more recently used sparse validation sites are predominantly forest ecosystems [12], [20], [27]. Incorporating the NEON sites increases the number of forested sites currently in the sparse validation network seven-fold, but forests remain underrepresented. Moreover, some forest types are particularly underrepresented such as evergreen broadleaf forests (tropical forests), which account for 10% of land cover but $< 1\%$ of the sparse network sites [12], [28]. Collectively this indicates the need to increase *in situ* soil moisture monitoring in forested ecosystems and their inclusion in satellite validation networks.

VI. CONCLUSION

We found that NEON and SMAP soil moisture measurements agreed with one another to a similar extent as other sparse validation networks at unforested sites. Moreover, at forested sites, which were underrepresented in previous SMAP validations, SMAP soil moisture measurements achieved a reasonable level of sensitivity to soil moisture changes (ubRMSD: 0.06 or $0.053 \text{ m}^3 \text{ m}^{-3}$ after accounting for random representativeness errors). Encouragingly, we identified that both an index of aboveground vegetation water content (NDII) and canopy height were related to the MD magnitude, suggesting a plausible approach to correcting bias. In addition, expanding the number and extent of soil moisture measurements throughout the SMAP footprint at forested validation sites would likely further minimize MD by ensuring that the spatial scale of *in situ* and satellite-based soil moisture measurements were similar, thereby minimizing representativeness errors. These findings were facilitated by the wide array of standardized, open-access, national-scale ecological data products that are uniquely available at NEON sites, demonstrating the suitability of NEON for satellite validation.

ACKNOWLEDGMENT

Mahesh Pun contributed to early discussions of NEON-SMAP soil moisture comparisons. The National Ecological Observatory Network is a program sponsored by the National Science Foundation and operated under cooperative agreement by Battelle. This material is based in part upon work supported by the National Science Foundation through the NEON Program. E. A., J. A. R., M. A. G., and S. S. were supported by Battelle, Columbus OH. Courtney Meier and Katie Jones provided input on the aboveground biomass calculations. A contribution to the work was done at the Jet Propulsion Laboratory, California Institute of Technology under a contract with National Aeronautics and Space Administration. Landsat 8 data is courtesy of the United States Geological Survey. USDA is an equal opportunity provider and employer.

REFERENCES

- [1] T. J. Jackson *et al.*, "Validation of advanced microwave scanning radiometer soil moisture products," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4256–4272, Dec. 2010.
- [2] T. J. Jackson *et al.*, "Validation of soil moisture and ocean salinity (SMOS) soil moisture over watershed networks in the U.S.," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1530–1543, May 2012.
- [3] A. Colliander *et al.*, "Validation of soil moisture data products from the NASA SMAP mission," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. Under Rev.*, to be published.
- [4] C. Montzka *et al.*, "Soil moisture product validation good practices protocol version 1.0," in *Good Practices Satell. Derived Land Product Validation Land Product Validation Subgroup (WGCV/CEOS)*, 2020.
- [5] A. Gruber *et al.*, "Validation practices for satellite soil moisture retrievals: What are (the) errors?," *Remote Sens. Environ.*, vol. 244, 2020, Art. no. 111806.
- [6] F. Chen *et al.*, "Application of triple collocation in ground-based validation of soil moisture active/passive (SMAP) level 2 data products," *J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 489–502, 2017.
- [7] W. T. Crow *et al.*, "Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products," *Rev. Geophys.*, vol. 50, 2012, Art. no. RG2002.
- [8] E. Coopersmith, M. H. Cosh, R. Bindlish, and J. Bell, "Comparing AMSR-E soil moisture estimates to the extended record of the U.S. climate reference network (USCRN)," *Adv. Water Resour.*, vol. 85, pp. 79–85, 2016.
- [9] E. Coopersmith, M. H. Cosh, J. Bell, and W. T. Crow, "Multi-profile analysis of soil moisture within the U.S. climate reference network," *Vadose Zone J.*, vol. 15, no. 1, pp. 1–8, 2016, doi: [10.2136/vzj2015.01.0016](https://doi.org/10.2136/vzj2015.01.0016).
- [10] T. Ochsner *et al.*, "State of the art in large-scale soil moisture monitoring," *Soil Sci. Soc. Amer. J.*, vol. 77, no. 6, pp. 1888–1919, 2013.
- [11] D. Entekhabi, S. Yueh, P. O'Neill, and K. Kellogg, "SMAP HANDBOOK – soil moisture active passive: Mapping soil moisture and freeze/thaw from space," SMAP Project, Jet Propulsion Lab., Pasadena, CA, USA, 2014.
- [12] S. K. Chan *et al.*, "Assessment of the SMAP passive soil moisture product," *IEEE Trans. Geosci. Rem. Sens.*, vol. 54, no. 8, pp. 4994–5007, Aug. 2016.
- [13] S. K. Chan *et al.*, "Development and assessment of the SMAP enhanced passive soil moisture product," *Remote Sens. Environ.*, vol. 204, pp. 931–941, 2018.
- [14] N. N. Das *et al.*, "The SMAP mission combined active-passive soil moisture product at 9 km and 3 km spatial resolutions," *Remote Sens. Environ.*, vol. 211, pp. 204–217, 2018.
- [15] N. N. Das *et al.*, "The SMAP and copernicus sentinel 1A/B microwave active-passive high resolution surface soil moisture product," *Remote Sens. Environ.*, vol. 233, 2019, Art. no. 111380.
- [16] S. B. Kim *et al.*, "Surface soil moisture retrieval using the L-band synthetic aperture radar onboard the soil moisture active passive (SMAP) satellite and evaluation at core validation sites," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 1897–1914, Apr. 2017.
- [17] R. H. Reichle *et al.*, "Assessment of the SMAP level-4 surface and root-zone soil moisture product using in situ measurements," *J. Hydrometeorol.*, vol. 18, no. 10, pp. 2621–2645, 2017.
- [18] T. J. Schmugge, P. E. O'Neill, and J. R. Wang, "Passive microwave soil moisture research," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-24, no. 1, pp. 12–22, Jan. 1986.
- [19] T. J. Jackson *et al.*, "SMAP science data calibration and validation plan," SMAP Mission, Jet Propulsion Lab., Pasadena, CA, USA, Mar. 14, 2014. [Online]. Available: <http://smap.jpl.nasa.gov/science/validation/>
- [20] A. Colliander *et al.*, "Validation of SMAP surface soil moisture products with core validation sites," *Remote Sens. Environ.*, vol. 191, pp. 215–231, 2017.
- [21] M. S. Burgin *et al.*, "A comparative study of the SMAP passive soil moisture product with existing satellite-based soil moisture products," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2959–2971, May 2017.
- [22] F. Chen *et al.*, "Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation," *Remote Sens. Environ.*, vol. 214, pp. 1–13, 2018.
- [23] A. Colliander *et al.*, "Validation and scaling of soil moisture in a semi-arid environment: SMAP validation experiment 2015 (SMAPVEX15)," *Remote Sens. Environ.*, vol. 196, pp. 101–112, 2017.
- [24] A. Colliander *et al.*, "Comparison of high-resolution airborne soil moisture retrievals to SMAP soil moisture during the SMAP validation experiment 2016 (SMAPVEX16)," *Remote Sens. Environ.*, vol. 227, pp. 137–150, 2019.
- [25] N. Ye *et al.*, "Evaluation of SMAP downscaled brightness temperature using SMAPEX-4/5 airborne observations," *Remote Sens. Environ.*, vol. 221, pp. 363–372, 2019.
- [26] T. J. Jackson, R. Hurksmans, A. Hsu, and M. H. Cosh, "Soil moisture algorithm validation using data from the advanced microwave scanning radiometer (AMSR-E) in Mongolia," *Italian J. Remote Sens.*, vol. 30, no. 31, pp. 23–32, 2005.
- [27] P. O'Neill *et al.*, "Calibration and validation for the L2/3_SM_P version 7 and L2/3_SM_P_E version 4 data products," SMAP Project, JPL D-56297, Jet Propulsion Lab, Pasadena, CA, USA, Aug. 31, 2020.
- [28] E. Ayres, H. Luo, and H. Loescher, "D11 FIU site characterization: Summary," 2015.
- [29] H. Loescher, E. Ayres, P. Duffy, H. Luo, and M. Brunke, "Spatial variation in soil properties among North American ecosystems and guidelines for sampling designs," *PLoS ONE*, vol. 9, no. 1, 2014, Art. no. e83216.
- [30] E. Ayres, "TIS soil plot layout," 2019.
- [31] "Calibration manual for Sentek soil moisture sensors, version 2.0," Sentek Pty., Stepney, SA, Australia, 2011.
- [32] J. A. Roberti *et al.*, "A robust calibration method for continental-scale soil water content measurements," *Vadose Zone J.*, vol. 17, no. 1, pp. 1–9, 2018.
- [33] J. R. Piepmeier *et al.*, "Radio-frequency interference mitigation for the soil moisture active passive microwave radiometer," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 761–775, Jan. 2014.
- [34] J. R. Piepmeier *et al.*, "SMAP L-band microwave radiometer: Instrument design and first year on orbit," *IEEE Trans. Geosci. Rem. Sens.*, vol. 55, no. 4, pp. 1954–1966, Apr. 2017.
- [35] P. N. Mohammed, M. Aksoy, J. R. Piepmeier, J. T. Johnson, and A. Bringer, "SMAP L-band microwave radiometer: RFI mitigation prelaunch analysis and first year on-orbit observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6035–6047, Oct. 2016.
- [36] M. J. Chaubell *et al.*, "Algorithm theoretical basis document (ATBD): SMAP level 1(B/C) enhanced radiometer brightness temperature data product," SMAP Project, Rev. B, May 31, 2018.
- [37] P. E. O'Neill, S. Chan, E. G. Njoku, T. Jackson, R. Bindlish, and J. Chaubell, "SMAP enhanced L2 radiometer half-orbit 9 km EASE-grid soil moisture, version 4," NASA Nat. Snow Ice Data Center Distrib Active Arch. Center, Boulder, CO, USA, 2020, doi: [10.5067/Q8J8E3A89923](https://doi.org/10.5067/Q8J8E3A89923).
- [38] C. K. Lunch and C. M. Laney, "neonUtilities: Utilities for working with NEON data," Nat. Ecol. Observ. Netw., Boulder, CO, USA, 2021, [Online]. Available: <https://github.com/NEONScience/NEON-utilities>
- [39] M. SanClements *et al.*, "Collaborating with NEON," *BioScience*, vol. 70, no. 2, 2020, Art. no. 107.
- [40] 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, 2010.
- [41] I. Rodríguez-Iturbe, G. K. Vogel, R. Rigon, D. Entekhabi, F. Castelli, and A. Rinaldo, "On the spatial organization of soil moisture fields," *Geophys. Res. Lett.*, vol. 22, no. 22, pp. 2757–2760, 1995.
- [42] J. S. Famiglietti, J. W. Rudnicki, and M. Rodell, "Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas," *J. Hydrol.*, vol. 210, no. 1–4, pp. 259–281, 1998.
- [43] W. T. Crow and E. F. Wood, "Multi-scale dynamics of soil moisture variability observed during SGP'97," *Geophys. Res. Lett.*, vol. 26, no. 23, pp. 3485–3488, 1999.
- [44] M. H. Cosh, P. J. Starks, J. A. Guzman, and D. N. Moriasi, "Upper Washita river experimental watersheds: Multiyear stability of soil water content profiles," *J. Environ. Qual.*, vol. 43, no. 4, pp. 1328–1333, 2014.
- [45] J. S. Famiglietti, D. Ryu, A. A. Berg, M. Rodell, and T. J. Jackson, "Field observations of soil moisture variability across scales," *Water Resour. Res.*, vol. 44, no. 1, 2008, Art. no. W01423.
- [46] F. Chen *et al.*, "Uncertainty of reference pixel soil moisture averages sampled at SMAP core validation sites," *J. Hydrometeorol.*, vol. 20, pp. 1553–1568, 2019.
- [47] J. Dong *et al.*, "Comparison of microwave remote sensing and land surface modeling for surface soil moisture climatology estimation," *Remote Sens. Environ.*, vol. 242, 2020, Art. no. 111756.
- [48] D. C. Chojnacky, L. S. Heath, and J. C. Jenkins, "Updated generalized biomass equations for North American tree species," *Forestry*, vol. 87, no. 1, pp. 129–151, 2014.
- [49] J. Chave, D. Coomes, S. Jansen, S. L. Lewis, N. G. Swenson, and A. E. Zanne, "Towards a worldwide wood economics spectrum," *Ecol. Lett.*, vol. 12, no. 4, pp. 351–366, 2009.
- [50] A. E. Zanne *et al.*, "Towards a worldwide wood economics spectrum," *Ecol. Lett.*, vol. 12, pp. 351–366, 2009.

- [51] M. B. Cleary, E. Pendall, and B. E. Ewers, "Testing sagebrush allometric relationships across three fire chronosequences in Wyoming, USA," *J. Arid Environ.*, vol. 72, no. 4, pp. 285–301, 2008.
- [52] L. F. Huenneke, D. Clason, and E. Muldavin, "Spatial heterogeneity in Chihuahuan desert vegetation: Implications for sampling methods in semi-arid ecosystems," *J. Arid Environ.*, vol. 47, no. 3, pp. 257–270, 2001.
- [53] J. A. Lutz *et al.*, "Community composition and allometry of leucothoe davisiae, cornus sericea, and chrysolepis sempervirens," *Can. J. Forest Res.*, vol. 44, no. 6, pp. 677–683, 2014.
- [54] C. Zhang, J. Wang, X. Zhao, F. Xia, and K. V. Gadow, "Sexual dimorphism in reproductive and vegetative allometry for two dioecious rhamnus plants in north-eastern China," *Eur. J. Forest Res.*, vol. 131, no. 5, pp. 1287–1296, 2012.
- [55] H. L. Gholz, C. C. Grier, A. G. Campbell, and A. T. Brown, "Equations for estimating biomass and leaf area of plants in the pacific northwest," Research Paper 41. Forest Res. Lab., School Forestry, Oregon State Univ., Corvallis, OR, USA, 1979.
- [56] H. L. Gholz, D. N. Guerin, and W. P. Cropper, "Phenology and productivity of saw palmetto (*Serenoa repens*) in a North Florida slash pine plantation," *Can. J. Forest Res.*, vol. 29, no. 8, pp. 1248–1253, 1999.
- [57] E. G. Bobich and T. E. Huxman, "Dry mass partitioning and gas exchange for young ocotillo (*Fouquieria splendens*) in the Sonoran desert," *Int. J. Plant Sci.*, vol. 170, no. 3, pp. 283–289, 2009.
- [58] C. B. Halpern and J. A. Lutz, "Canopy closure exerts weak controls on understory dynamics: A 30-year study of overstory-understory interactions," *Ecol. Monographs*, vol. 83, no. 2, pp. 221–237, 2013.
- [59] Y. Benjamini and Y. Hochberg "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc., Ser. B (Methodol.)*, vol. 57, no. 1, pp. 289–300, 1995.
- [60] S. Chan, R. Bindlish, R. Hunt, T. Jackson, and J. Kimball, "Soil moisture active passive (SMAP) ancillary data report: Vegetation water content," *Preliminary*, vol. 1, no. 47, pp. 1–15, 2013.
- [61] S. Kim, "Soil moisture active passive (SMAP) ancillary data report: Landcover classification," 2013.
- [62] H. Jactel *et al.*, "Drought effects on damage by forest insects and pathogens: A meta-analysis," *Global Change Biol.*, vol. 18, no. 1, pp. 267–276, 2012.
- [63] P. Meir *et al.*, "Threshold responses to soil moisture deficit by trees and soil in tropical rain forests: Insights from field experiments," *BioScience*, vol. 65, no. 9, pp. 882–892, 2015.
- [64] D. Jensen, J. T. Reager, B. Zajic, N. Rousseau, M. Rodell, and M. Hinkley "The sensitivity of US wildfire occurrence to pre-season soil moisture conditions across ecosystems," *Environ. Res. Lett.*, vol. 13, no. 1, 2018, Art. no. 014021
- [65] P. Potapov *et al.*, "Mapping global forest canopy height through integration of GEDI and landsat data," *Remote Sens. Environ.*, vol. 253, 2020, Art. no. 112165.
- [66] C. Zhao, S. M. Quiring, S. Yuan, D. B. McRoberts, N. Zhang, and Z. Leasor, "Developing and evaluating national soil moisture percentile maps," *Soil Sci. Soc. Amer. J.*, vol. 84, no. 2, pp. 443–460, 2020.
- [67] A. Colliander *et al.*, "SMAP detects soil moisture under temperate forest canopies," *Geophys. Res. Lett.*, vol. 47, no. 19, pp. e2020GL089697, 2020.
- [68] T. J. Jackson and T. J. Schmugge, "Vegetation effects on the microwave emission of soils," *Remote Sens. Environ.*, vol. 36, pp. 203–212, 1991.
- [69] P. J. Shellito *et al.*, "SMAP soil moisture drying more rapid than observed in situ following rainfall events," *Geophys. Res. Lett.*, vol. 43, pp. 8068–8075, 2016.
- [70] M. Kurum, P. E. O'Neill, R. H. Lang, A. T. Joseph, M. H. Cosh, and T. J. Jackson, "Effective tree scattering and opacity at L-band," *Remote Sens. Environ.*, vol. 118, pp. 1–9, 2012.
- [71] M. Schwank, M. Guglielmetti, C. Matzler, and H. Fluhler, "Testing a new model for the L-band radiation of moist leaf litter," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 1982–1994, Jul. 2008.
- [72] M. Kurum, P. E. O'Neill, R. H. Lang, M. H. Cosh, A. T. Joseph, and T. J. Jackson, "Impact of conifer forest litter on microwave emission at L-band," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1071–1084, Apr. 2012.
- [73] J. Seppänen, J. Kainulainen, J. Heiskanen, J. Praks, and M. Hallikainen, "Measurements of boreal coniferous forest soil and humus with an airborne radiometer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 3219–3228, Jul. 2016.
- [74] H. R. Manns, A. A. Berg, and A. Colliander, "Soil organic carbon as a factor in passive microwave retrievals of soil water content over agricultural croplands," *J. Hydrol.*, vol. 528, pp. 643–651, 2015.
- [75] E. D. Kabela, B. K. Hornbuckle, M. H. Cosh, M. C. Anderson, and M. L. Gleason, "Dew frequency, duration, amount, and distribution in corn and soybean during SMEX05," *Agricultural Forest Meteorol.*, vol. 149, no. 1, pp. 11–24, 2009.



Edward Ayres received the Ph.D. degree in biological sciences (plant-soil ecology) from Lancaster University, Lancaster, U.K. and the Centre for Ecology and Hydrology, Lancaster, U.K. in 2003.

He is currently a Lead Soil Ecologist with the National Ecological Observatory Network (NEON), Boulder, CO, USA. He oversees NEON's soil sensor data products and the Megapit Soil Archive. Previously with NEON, he assisted with selecting the terrestrial sensor measurement locations and characterizing soils throughout the observatory. His research

interests center around plant-soil interactions.



Andreas Colliander (Senior Member, IEEE) received the M.Sc. (Tech.), Lic.Sc. (Tech.), and D.Sc. (Tech.) degrees from the Helsinki University of Technology (TKK; now Aalto University), Espoo, Finland, in 2002, 2005, and 2007, respectively.

He is currently a Research Scientist with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. He is currently leading the calibration and validation of the geophysical retrievals of NASA's SMAP mission. His research interests include development of microwave remote

sensing techniques.



Michael H. Cosh (Senior Member, IEEE) received the Ph.D. degree in civil and environmental engineering from Cornell University, Ithaca, NY, USA, in 2002.

He is currently a Research Hydrologist with the U.S. Department of Agriculture, Agricultural Research Service, Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA. He conducts research on satellite calibration and validation for such missions as the Soil Moisture Active Passive and Soil Moisture Ocean Salinity Missions. His current research interests include the monitoring of soil moisture from both in situ

resources and satellite products.

Dr. Cosh is a Fellow of the American Society of Agronomy.



Joshua A. Roberti received the B.S. degree in 2009 and the M.S. degree in 2012 in meteorology from Plymouth State University, Plymouth, NH, USA.

From 2012 to Mar 2017, he was an Associate Scientist at the National Ecological Observatory Network, where he developed many of NEON's measurement algorithms, QA/QC frameworks, and uncertainty budgets. From March 2017 to September 2021, he was a lead Data Scientist within NEON's Instrumented Systems (IS) group, where he developed interactive dashboards and predictive analytics

to aid in identifying and predicting sensor malfunctions/failures within the Observatory. In September 2021, he left the NEON program and is currently a Data Scientist in the private sector.



Sam Simkin received the B.A. degree in biology from Earlham College, Richmond, IN, USA, in 1995, the M.S. degree in conservation ecology from the University of Georgia, Athens, GA, USA, in 1998 and the Ph.D. degree in natural resources from Cornell University, Ithaca, NY, USA, in 2012.

He was a Research Assistant with the Cary Institute of Ecosystem Studies and a Research Associate with the University of Colorado, Boulder, CO, USA. Since 2017, he has been a Research Scientist with Battelle, National Ecological Observatory Network in Boulder, CO, USA. His research interests include ecosystem ecology, plant ecology, and wetland biogeochemistry.



Melissa A. Genazzio received the M.Sc. degree in marine biology from the University of North Carolina Wilmington, Wilmington, NC, USA, in 2014.

She is currently the Global Environmental Research Infrastructure and International Strategic Development Staff Scientist with Battelle Memorial Institute. She aims to develop interoperability across e-infrastructures and make the informatics interoperable among large-scale environmental science observatories across the globe. She was the Biological Science Coordinator with the NASA Global Change Master Directory and NASA Common Metadata Repository at Goddard Space Flight Center in Greenbelt, MD, USA, and the Science Assistant for the Division of Ocean in the Directorate for Geosciences with The National Science Foundation, Arlington, VA, USA.