Evaluation of SMAP Core Validation Site Representativeness Errors Using Dense Networks of *In Situ* Sensors and Random Forests

Jane Whitcomb, *Member, IEEE*, Daniel Clewley, Andreas Colliander, *Senior Member, IEEE*, Michael H. Cosh, *Senior Member, IEEE*, Jarrett Powers, Matthew Friesen, Heather McNairn, Aaron A. Berg, David D. Bosch, Alisa Coffin, Chandra Holifield Collins, John H. Prueger, Dara Entekhabi, *Fellow, IEEE*, and Mahta Moghaddam, *Fellow, IEEE*

***Abstract*—In order to validate its soil moisture products, the NASA Soil Moisture Active Passive (SMAP) mission utilizes sites with permanent networks of *in situ* soil moisture sensors maintained by independent calibration and validation partners in a variety of ecosystems around the world. Measurements from each core validation site (CVS) are combined in a weighted average to produce an estimate of soil moisture at a 33-km scale that represents the SMAP's radiometer-based retrievals. Since upscaled estimates produced in this manner are dependent on the weighting scheme applied, an independent method of quantifying their biases is needed. Here, we present one such method that uses soil moisture measurements taken from a dense, but temporary, network of soil moisture sensors deployed at each CVS to train a random forests regression expressing soil moisture in terms of a set of spatial variables. The regression then serves as an independent source of upscaled estimates against which permanent network upscaled estimates can be compared in order to calculate bias statistics. This method, which offers a systematic and unified approach to estimate bias across a variety of validation sites, was applied to estimate biases at four CVSs. The results showed that the magnitude of the uncertainty in the permanent network upscaling bias can sometimes exceed 80% of the upper limit on SMAP's entire allowable unbiased root-mean-square error (ubRMSE). Such large CVS bias uncertainties could make it more difficult to assess biases in soil moisture estimates from SMAP.**

***Index Terms*—Random forests, soil moisture, Soil Moisture Active Passive (SMAP), upscaling.**

Manuscript received February 27, 2020; revised September 9, 2020 and October 7, 2020; accepted October 8, 2020. Date of publication October 26, 2020; date of current version December 4, 2020. This work was supported by the US National Aeronautics and Space Administration (NASA) under NASA Award NNX14AH82G. *(Corresponding author: Jane Whitcomb.)*

Jane Whitcomb and Mahta Moghaddam are with the University of Southern California, Los Angeles, CA 90007 USA (e-mail: jbwhitco@usc.edu; mahta@usc.edu).

Daniel Clewley is with Plymouth Marine Laboratory, Plymouth PL1 3DH, U.K. (e-mail: dac@pml.ac.uk).

Andreas Colliander is with Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: andreas.colliander@jpl.nasa.gov).

Michael H. Cosh is with USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705-2350 USA (e-mail: Michael.Cosh@ars.usda.gov).

Jarrett Powers and Matthew Friesen are with Agriculture and Agri-Food Canada, Science and Technology Branch, Winnipeg, MB R3C 1B2, Canada (e-mail: jarrett.powers@agr.gc.ca; matthew.friesen@agr.gc.ca).

Heather McNairn is with Agriculture and Agri-Food Canada, Science and Technology Branch, Ottawa, ON K1A 0C5, Canada (e-mail: heather.mcnairn@agr.ac.ca).

Aaron A. Berg is with the Department of Geography, Environment and Geomatics, University of Guelph, Guelph, ON N1G 2W1, Canada (e-mail: aberg@uoguelph.ca).

David D. Bosch and Alisa Coffin are with USDA-ARS, Southeast Watershed Research Laboratory, Tifton, GA 31794 USA (e-mail: david.bosch@ars.usda.gov; alisa.coffin@usda.gov).

Chandra Holifield Collins is with USDA-ARS Southwest Watershed Research Laboratory, Tucson, AZ 85719 USA (e-mail: chandra.holifield@ars.usda.gov).

John H. Prueger is with the USDA-ARS National Laboratory for Agriculture and Environment, Ames, IA 50011 USA (e-mail: john.prueger@ars.usda.gov).

Dara Entekhabi is with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: darae@mit.edu).

Digital Object Identifier 10.1109/JSTARS.2020.3033591

I. Introduction

CONCERNS over recent and projected changes in our planet's climate have cast a spotlight on the urgent need for atmospheric models accurate enough to provide reliable climate projections under a variety of policy options. The development of such models relies on a solid understanding of the global hydrological cycle, of which one key observable over land is soil moisture. Additionally, an accurate knowledge of the spatial distribution of soil moisture is integral to more immediate applications such as weather forecasting, agricultural productivity enhancement, water resources management, flood area mapping, and ecosystem health monitoring [1], [2]. Access to accurate and timely data on surface soil moisture is vital to all of these purposes.

The Soil Moisture Active Passive (SMAP) satellite observatory [3], launched in January 2015, was developed to provide frequent global observations of soil moisture using L-band ($\sim$1.4 GHz) microwave measurements collected by on-board radiometer and synthetic aperture radar instruments. It provides global estimates of soil moisture in the top 5 cm of the earth's surface every 2–3 days. Details on the SMAP mission are provided in [1] and [4]. As described there, the mission requirements for SMAP included a retrieval accuracy placing an upper limit of 0.04 m^3/m^3 on the overall unbiased root-mean-square error (ubRMSE). This is to be achieved at a spatial scale no greater than 10 km for vegetation water content of less than 5 kg/m^2. Failure of the on-board radar in July 2015 made this requirement more challenging, but led to the development of an enhanced radiometer-derived soil moisture product (L2SMP_E) [2], [5], [4]. The L2SMP_E product is posted in the

EASEGrid-2.0 projection[1] [6], [7] at 9 km, with each 9-km soil moisture pixel reflecting data aggregated over a 33-km contributing domain (i.e., primary spatial area contributing to the radiometer brightness temperature response) centered on the pixel, as shown in [4, Fig. 3].

The primary method of evaluating the SMAP enhanced soil moisture product has been to compare SMAP soil moisture retrievals with measurements from networks of well-calibrated *in situ* soil moisture sensors at a number of test sites, referred to as core validation sites (CVSs), deployed in various ecosystems around the globe [8]. These CVSs provide reference pixels where the average soil moisture can be estimated at the SMAP spatial scales [9]. In order to be considered a CVS, each network must have been verified as reliably providing a spatial average of soil moisture at the SMAP footprint scale. Multiple studies have characterized the reliability and spatial variability of the *in situ* measurements at these sites, and established the representativeness of the measurements at large scales [10]–[12].

At each CVS, 10–30 (typically) permanent soil moisture measurement stations have been deployed, with most such stations located within a single 33-km contributing domain grid of interest, here called the SMAP grid. At certain CVSs, an additional 40–50 temporary soil moisture measurement stations have been deployed, but only for a limited period of time.

To assess the performance of the SMAP enhanced soil moisture product, the *in situ* soil moisture measurements within each CVS must be upscaled to a single value at the 33 km pixel size to be directly compared to the SMAP soil moisture retrieval. Undersampling of the range of variability of soil moisture within the field-of-view of the radiometer results in representativeness errors of the ground-truth. The problem is nontrivial due to the major difference in scales between the 33-km enhanced SMAP contributing domain grid and localized point-scale *in situ* sensor measurements. Data collected at each *in situ* soil moisture measurement station typically accurately reflect soil moisture only in the immediate vicinity of the station [13], such that sensor networks must ordinarily be spatially dense in order to fully capture the spatial heterogeneity of geophysical variables within the site [14].

Upscaling at the SMAP CVSs is currently tailored to each site, comprising a weighted average of the network's permanent *in situ* sensor measurements [9]. For most CVS sites, including three of the four sites considered here, weights are chosen based on geometric considerations such as Voronoi (aka Theissen) polygons [15]. Alternatively, weights can be selected by considering local soil texture and vegetation characteristics based on the local hydrological conditions. The accuracy of the upscaled values produced by these permanent-network-based CVS scaling functions (SFs) is typically verified by a combination of intensive field campaigns and the installation of dense but temporary sensor networks [16].

Accuracy of the SMAP retrievals is then assessed by calculating error statistics relative to each CVS site's upscaled *in situ* soil moisture measurements. Error statistics have included [4], [17] time-series correlation, bias, rms error (RMSE), ubRMSE, and (more recently, [8]) overall mean absolute bias (MAB). On this basis, it has been shown, for example, that when SMAP L2SMP_E retrievals are evaluated relative to CVS SF upscaled *in situ* measurements, the resulting ubRMSE statistics, taken over all sites, are approximately as accurate as the original L2SMP retrievals, thereby meeting SMAP mission requirements despite the finer grid size of the enhanced product [18]. In these assessments, however, it is important to understand any error sources that may corrupt the CVS SF upscaled soil moisture estimates themselves, since these errors can distort the SMAP performance statistics. It has been found, in particular, that error statistics collected from some sites are significantly worse than those collected from other sites [9].

The spatial distribution of soil moisture is a complex and nonlinear function of (typically large-scale) meteorological variables and (typically smaller-scale) static land characteristics that include topography (elevation, slope, aspect, and flow accumulation), soil texture, and land cover [14], [19]. Because of the latter, upscaled estimates derived from *in situ* sensor networks most closely represent the true mean soil moisture within a given SMAP grid if the sensors are dense and evenly spaced [14]. It is possible to produce accurate upscaled estimates based on sparser networks, but only if sensor locations have been optimized based on known persistent soil moisture patterns within the site (e.g., [20], [21]). In many locations, however, it is difficult or impossible to maintain either type of arrangement, especially for the duration of a long-term validation effort. Given the large geographic extent of the 33-km SMAP grid, it is generally the case that many otherwise desirable sensor locations are inaccessible due to such hindrances as property boundaries, paving, topography, surface water, and dense vegetation. These unavoidable limitations on permanent *in situ* sensor station locations often make it impossible to fully capture the heterogeneity of the site's surface characteristics and the natural variability of soil moisture, resulting in an incomplete spatial representation that can lead to biases in the CVS SFs upscaled soil moisture estimates. Since these biases act as a source of uncertainty in the error statistics used to quantify the accuracy of SMAP soil moisture retrievals, it is important to establish a ceiling as to how large they might be. The objective of this study, therefore, was to develop a method to quantify the error performance of the CVS SFs, and to illustrate the use of the method at four of the CVS sites.

Study site characteristics are described in Section II. Our bias assessment methodology is presented in Section III, results are discussed in Section IV, discussion is provided in Section V, and conclusions are provided in Section VI.

II. Study Sites

We have investigated *in situ* CVS SF biases for four different SMAP CVSs. These are the sites for which dense temporary network reference data were available. Study site characteristics are briefly summarized in Table I and their locations are shown in Fig. 1. Detailed descriptions of the study sites are provided in the Appendix.

At all four CVSs, the analysis included all stations within the 33-km SMAP grid and some stations outside of, but close to, the

[1] [Online]. Available: https://nsidc.org/data/ease/ease_grid2.html

TABLE I
CHARACTERISTICS OF STUDY SITES

Site Name	Coordinates	Land Cover	Climate[1]	Permanent Nodes/Sensors	Temporary Nodes
Carman, MB	(49.60, -97.98)	Row Crops	Cold (Dfb)	9/20	50
South Fork, IA	(42.42, -93.41)	Row Crops	Cold (Dfa)	15/15	39
Walnut Gulch, AZ	(31.75, -110.03)	Desert Scrub	Arid (BSk)	28/28	49
Little River, GA	(31.67, -83.60)	Row Crops/Evergreens	Temperate (Cfa)	33/33	48

[1] Koeppen–Geiger climate classification.

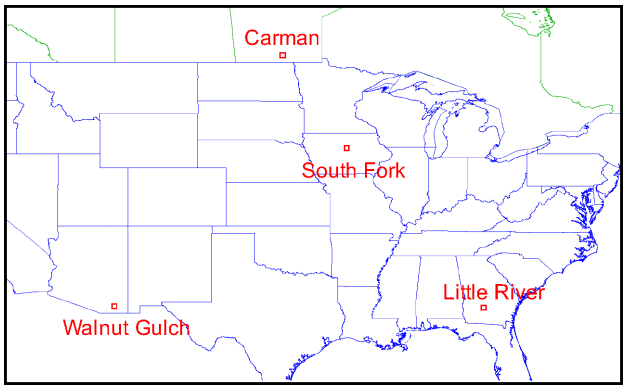

Fig. 1. Locations of study sites.

SMAP grid. This gave us the benefit of the soil moisture data offered by those outside sensors in cases where there appeared to be no significant difference in landcover, topography, or soil characteristics relative to the characteristics seen within the grid boundary, and also took advantage of the spatial autocorrelation of soil moisture (e.g., [22]).

III. METHODOLOGY

In previous work [13], we developed and validated a regression-based (RB) method of upscaling capable of providing accurate estimates of upscaled soil moisture even from sparse, irregularly spaced networks of sensors. This research applies this well-tested method to assess biases in the soil moisture estimates derived from permanent *in situ* sensor stations and the CVS SF weighted-average. We do this by comparing the permanent station CVS SF biases to biases in the RB method's soil moisture estimates when the latter are based on measurements taken from dense networks of temporary soil moisture sensors installed during various field campaigns. In doing so, we assume that the dense temporary network constitutes the best available source of reference data based on which the closest estimate of the true field mean soil moisture can be calculated. In other words, the data from the dense temporary networks provide our "ground truth."

A number of different methods have been developed for upscaling soil moisture measurements from *in situ* soil moisture networks to satellite footprint scale, as has been described in [9], [13], and [14]. Here, we used a random forests RB upscaling method based on the dense network of temporary stations at each CVS to derive upscaled estimates of soil moisture. The RB estimates were then compared to CVS SF soil moisture estimates based on permanent *in situ* sensor measurements at the same CVS. This comparison made it possible to quantify the bias uncertainty of the CVS SF soil moisture estimates.

A. RB Upscaling Method

As described in [13], the RB upscaling method relies on repeated runs of a well-known algorithm called random forests [23], [24]. Random forests is a highly versatile and well-tested machine-learning regression algorithm that has been applied to a diverse array of classification and nonlinear regression problems.

For a given trial run with a defined set of run parameters including start time and stop time, iteration time span, and geographic boundaries of the area included in the analysis, the RB upscaling method runs random forests once for each iteration time interval within the overall start-to-stop time period.

For each iteration within the run, the algorithm [25] extracts *in situ* measurement data from each selected soil moisture sensor station within the area of interest and temporally averages them (and, in the case of stations with redundant sensors, sensor averages them) over the time interval of the iteration (throughout this analysis, the time interval was set to one day). It then assembles a stack of spatial variable data layers for the site and iteration time interval and extracts the values of the data layers for each station location. This yields a pool of N samples, where N is the number of stations used in the iteration. Each sample consists of a station's averaged soil moisture measurement accompanied by a vector of data layer values for the station location. Random forests uses the samples to construct a regression model that estimates soil moisture for each station location as a function of the data layer values. The resulting regression model is then applied to every pixel within the site's 33-km SMAP grid in order to obtain the soil moisture estimate for that pixel. This process yields a raster of soil moisture estimates that are then spatially averaged over all fine-resolution pixels within the 33-km SMAP grid to yield the grid's upscaled estimate of soil moisture for the iteration's time span.

This RB method can be applied systematically to any CVS for which station locations have been chosen to achieve a diverse sampling of values for the different data layers (e.g., various land cover types, elevation values, etc.). Site-specific modifications are chiefly confined to accommodating differences in CVS geographic boundaries, local spatial data layers, and *in situ* measurement data characteristics.

1) Random Forests: Random forests is a supervised machine learning algorithm that constructs a large number, a "forest" of decision trees based on various subsets of the available dataset,

with each tree calculating an estimate of the observable in question, in this case soil moisture, for all station locations. Random forests then averages across the trees' individual estimates for each station location to produce the location's overall regression model result. This approach has been found to improve predictive accuracy and reduce overfitting relative to what is achievable using a single decision tree. More information on random forests is provided in the Appendix.

Random forests works best when the forest contains a large number of trees. Here, as in [13], the regression model has implemented a forest of 300 decision trees because previous trial runs had shown accuracy to stop improving after about 200 trees and we wanted to allow some margin for additional run variability. Also as in [13], this analysis used three randomly selected data layers to split each decision tree node (i.e., in the notation of the Appendix, $n_{\text{split}} = 3$), a number seen as a compromise between creating strong individual trees and reducing correlation between trees.

The random forests regression used in our RB upscaling method was implemented using the scikit-learn Python library [26].

2) Random Forests Predicted Soil Moisture: Once constructed, each iteration's regression model is applied for every fine-resolution pixel within the SMAP grid. This is done by running the vector of data layer values for each pixel in the raster through all decision trees in the forest and averaging the resulting soil moisture predictions to obtain the soil moisture estimate for that pixel.

The spatial variable data layers used in the analysis do not contribute equally to the accuracy of the final result. Random forests evaluates the relative importance of the various input data layers in its regression computation. This is calculated based on the expected fraction of samples to which the spatial variable contributes and the decrease in residual sum of squared errors from splitting them to create a normalized estimate of the predictive power of that spatial variable [27].

A flow chart, taken from [13, Fig. 5] (with a few modifications), which summarizes the steps involved in the RB upscaling method of generating upscaled soil moisture estimates is shown in Fig. 2.

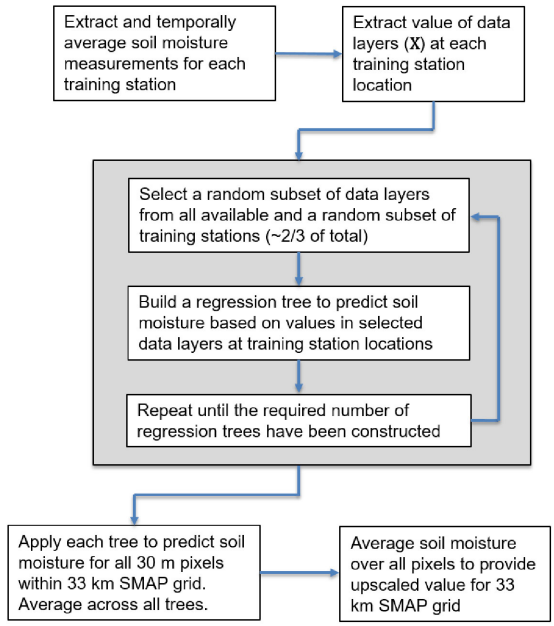

Fig. 2. Random forests upscaling method for a single time period. Steps inside the grey box are implemented within the core random forests algorithm. (This figure is a slightly modified version of [13, Fig. 5].)

B. Spatial Variable Data Layers

As mentioned, the spatial variables upon which the random forests nonlinear regression model is based are extracted from a stack of input data layers. The stack is constructed in the EASEGrid-2.0 map projection with a 30 m pixel spacing, the latter having been chosen to roughly accord with the resolution of most of the input data layers.

Specific data layer types were chosen to characterize the following geophysical parameters:

1) landcover;

2) elevation, along with elevation-derived slope, aspect, and flow accumulation;

3) soil texture, including clay fraction, sand fraction, and bulk density; and

4) weather, including precipitation and surface temperature (the latter a driver of evapotranspiration).

The ability of these spatial variables to affect the spatial distribution of soil moisture has been confirmed in studies using land surface models [28]. The spatial distribution of land cover types, in particular, is directly dependent on local soil texture, topography, and climate conditions that also closely correlate with soil moisture [29]. Local variations in elevation can influence soil moisture by affecting such subsurface parameters as depth to water table and through their correlation with slope, aspect, and flow accumulation, which largely determine runoff characteristics. Soil texture variables have a strong impact on soil moisture through their ability to directly affect the soil's capacity to absorb and retain water.

Landcover data for the Carman CVS were taken from the Agriculture and Agri-Food Canada crop inventory [30]. This crop inventory, which is provided at 30 m resolution, was derived by applying a decision-tree classification to optical (Landsat-8, Sentinel-2, Gaofen-1) and radar (RADARSAT-2) satellite images. Landcover data for the South Fork and Little River CVSs were taken from the cropland data layer (CDL) [31], a crop data layer produced based on satellite imagery collected during the growing season from Landsat 8, the Disaster Monitoring Constellation, ISRO's ResourceSat-2, and ESA's Sentinel-2. Landcover for the Walnut Gulch CVS was taken from the National Land Cover Database (NLCD) [32], [33], which is derived from Landsat data at a 30 m resolution. Since the NLCD is highly monochromatic at the Walnut Gulch CVS, particularly at the station locations, the use of an alternative, somewhat more detailed source of landcover, the GAP Land Cover Dataset,[2] was tried, but ultimately abandoned since it did not improve results.

Elevation data for the Carman CVS were taken from the Shuttle Radar Topography Mission (SRTM, SRTMGL1 V003

[2][Online]. Available: https://gapanalysis.usgs.gov/gaplandcover

[34]) at 1 arcsecond resolution. Elevation data for the South Fork, Walnut Gulch, and Little River CVSs were taken from the National Elevation Dataset (NED, [35]) at 1 arcsecond resolution. For all four CVSs, slope at each location was calculated using least squares to fit a plane to a window of elevation data around the location, then calculating the slope of the plane relative to the local tangent to the ellipsoid [36]. Following reprojection into the EASEGrid-2.0 projection, aspect was calculated using the RSGISLib function rsgislib.elevation.aspect (see [37]). Flow accumulation, that is, the accumulated flow of water into each pixel in the output raster as determined by summing the number of pixels from which water flows into the pixel (following a precursory calculation of flow direction out of each pixel in the raster), was calculated using the Hydrology toolset in ESRI's ArcGIS Desktop Spatial Analyst Toolbox (see [38]). Pixels with a high flow accumulation are areas of concentrated flow and can be used to identify stream channels while pixels with a flow accumulation of zero are local topographic highs and can be used to identify ridges.

Soil texture (i.e., clay fraction, sand fraction, and bulk density) for the top soil horizon of the Carman CVS was taken from the Canadian Soil Information Service Detailed Soil Survey (CanSIS DSS, [39]). Similar soil texture variables for the South Fork, Walnut Gulch, and Little River CVSs were taken from the USDA's Soil Survey Geographic Database (SSURGO, [40]). In both cases, GDAL's gdal_rasterize utility (see [41]) was used to rasterize the detailed soil data shape files to the 30-m resolution used in our analysis.

In the case of Carman, total precipitation, and skin temperature (i.e., temperature at the earth/air interface) were taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis [42]. ERA-interim data are provided at 3-h intervals and a pixel spacing of 0.075° (the nominal resolution for the dataset is 80 km). Daily averages of these data were calculated prior to use. ERA5, a next-generation weather dataset that offers higher accuracy and resolution than the ERA-interim, currently has a pixel spacing of 0.25° (the nominal resolution for the dataset is 31 km) that is too large for our application, but smaller spacings are expected soon. Precipitation and mean temperature data for the South Fork, Walnut Gulch, and Little River CVSs were taken from the PRISM Climate Group (see [43]); these data are available daily at a resolution of about 4 km. PRISM mean temperature values constitute daily averages and are interpreted as air temperatures at 2 m above the Earth's surface.

C. Assessment of Uncertainty in CVS SF Bias

As previously mentioned, the primary goal of this analysis was to assess biases in the CVS SF for soil moisture estimates. Toward this end, the RB upscaling method was run 30 times for each CVS site with a specified number of training stations, N. That is, the total number of temporary stations available at each site was divided into two mutually exclusive pools of stations, with one pool of N stations used for training the random forests regression and the remaining pool of stations used for validation. Given this split, it was expected that small values of N would result in accurate validation soil moisture estimates, but some loss of accuracy in RB soil moisture estimates, whereas large values of N would result in somewhat less accurate validation estimates but more accurate RB estimates.

For each trial run t soil moisture measurements from a random (with replacement) selection of N temporary *in situ* soil moisture stations are used to train a random forests RB model from which a RB upscaled estimate of soil moisture, $\mathrm{sm}_{\mathrm{RB}}(N,t,i)$, is calculated for each time interval i within the overall duration of the run. Soil moisture measurements from the remaining, unselected, temporary stations are averaged together to produce an independent validation "truth reference" estimate of soil moisture, $\mathrm{sm}_V(N,t,i)$. Additionally, soil moisture measurements from the permanent *in situ* soil moisture stations at the site are combined to compute a CVS SF estimate of soil moisture $\mathrm{sm}_{\mathrm{SF}}(i)$ in accordance with the method outlined in [9].

For any given time interval of any given trial run, therefore,

1) a randomly chosen subset of the temporary stations, A, is used to derive a RB soil moisture estimate $\mathrm{sm}_{\mathrm{RB}}(N,t,i)$;

2) the remaining, mutually exclusive subset of temporary stations, that is, the complement of A, is used to derive a validation "truth reference" estimate of soil moisture, $\mathrm{sm}_V(N,t,i)$; and

3) the full set of permanent stations at the site are used to derive a CVS SF estimate of soil moisture, $\mathrm{sm}_{\mathrm{SF}}(i)$.

Note that both $\mathrm{sm}_{\mathrm{RB}}(N,t,i)$ and $\mathrm{sm}_V(N,t,i)$ are derived using the relatively dense samples of soil moisture available from temporary station networks. This spatial density makes it possible to obtain a well-distributed set of training samples to support the RB upscaling while still maintaining enough validation samples to support a reasonably accurate "truth reference" of averaged validation station measurements.

Error statistics for both the random forest RB estimates and the CVS SF estimates were calculated relative to the validation "truth reference" estimates of soil moisture. These included RMSE, bias (mean error), and unbiased RMSE. The resulting error statistics were averaged across trial runs in order to reduce the effects of significant run-to-run variability.

Error calculations based on the random forests RB results were as follows:

$$\mathrm{RMSE}_{\mathrm{RB}}(N) = \frac{1}{N_t}\sum_t \sqrt{\frac{1}{N_i}\sum_i \left(\mathrm{sm}_{\mathrm{RB}}(N,t,i) - \mathrm{sm}_V(N,t,i)\right)^2} \quad (1)$$

$$\mathrm{bias}_{\mathrm{RB}}(N) = \frac{1}{N_t}\sum_t \frac{1}{N_i}\sum_i \left(\mathrm{sm}_{\mathrm{RB}}(N,t,i) - \mathrm{sm}_V(N,t,i)\right) \quad (2)$$

$$\mathrm{ubRMSE}_{\mathrm{RB}}(N) = \sqrt{\left(\mathrm{RMSE}_{\mathrm{RB}}(N)\right)^2 - \left(\mathrm{bias}_{\mathrm{RB}}(N)\right)^2} \quad (3)$$

and those based on the CVS SF results were

$$\mathrm{RMSE}_{\mathrm{SF}}(N) = \frac{1}{N_t}\sum_t \sqrt{\frac{1}{N_i}\sum_i \left(\mathrm{sm}_{\mathrm{SF}}(i) - \mathrm{sm}_V(N,t,i)\right)^2} \quad (4)$$

$$\text{bias}_{\text{SF}}(N) = \frac{1}{N_t}\sum_{t}\frac{1}{N_i}\sum_{i}\left(\text{sm}_{\text{SF}}(i) - \text{sm}_V(N,t,i)\right) \tag{5}$$

$$\text{ubRMSE}_{\text{SF}}(N) = \sqrt{\left(\text{RMSE}_{\text{SF}}(N)\right)^2 - \left(\text{bias}_{\text{SF}}(N)\right)^2}. \tag{6}$$

In these equations, N_i is the number of time intervals in each trial run and N_t is the number of trial runs for each value of N. The calculations resulted in a single value for each error statistic as a function of N, and were repeated for a sequence of N_{trains} values of N.

Two different bias metrics useful for assessing the overall uncertainty in CVS SF bias for each site were derived from the $\text{bias}_{\text{RB}}(N)$ and $\text{bias}_{\text{SF}}(N)$ error statistics. The underlying assumption in defining these metrics is that the bias derived by the proposed model based on dense networks of temporary stations constitutes a lower bound (or best-achievable) value of bias since the dense networks provide the best available estimates of the true mean of the soil moisture fields. The bias metrics are as follows.

1) Mean difference of biases: This metric is derived by averaging $\text{bias}_{\text{SF}}(N)$ and $\text{bias}_{\text{RB}}(N)$ over all values of N and subtracting the results

$$\text{bias}_{\text{SF}} = \frac{1}{N_{\text{trains}}}\sum_{N}\text{bias}_{\text{SF}}(N)$$

$$\text{bias}_{\text{RB}} = \frac{1}{N_{\text{trains}}}\sum_{N}\text{bias}_{\text{RB}}(N) \tag{7}$$

$$D_{\text{mean}} = \text{bias}_{\text{SF}} - \text{bias}_{\text{RB}} \tag{8}$$

where N_{trains} denotes the number of different values of N used in the analysis. This metric helps to quantify how closely the overall bias in CVS SF estimates approaches the best achievable value of bias.

2) RMS difference of biases: This metric is formed by calculating the squared difference between

$\text{bias}_{\text{SF}}(N)$ and $\text{bias}_{\text{RB}}(N)$ for all values of N and then computing the root of the mean of the resulting squares

$$\text{SD}(N) = \left(\text{bias}_{\text{SF}}(N) - \text{bias}_{\text{RB}}(N)\right)^2 \tag{9}$$

$$D_{\text{RMS}} = \sqrt{\frac{1}{N_{\text{trains}}}\sum_{N}\text{SD}(N)}. \tag{10}$$

This metric helps to quantify the absolute value uncertainty in the bias of CVS SF estimates.

IV. Results

The following results center around the mid-range cases of $N = N_{\text{mid}} = 26$, 20, 24, and 24 temporary nodes for the Carman, South Fork, Walnut Gulch, and Little River CVSs, respectively, with multi-N statistics spanning the range $N_{\text{mid}} - 6$ to $N_{\text{mid}} + 6$. Run results were generated for many values of N outside the range of values included in these multi-N statistics (see, for example, Fig. 7), but it was decided that the statistics collected using this range of Ns were likely most representative of the best achievable bias in CVS SF soil moisture estimates.

A. Performance of RB Upscaling Method

1) Typical Results From a Single Run: A typical raster of fine-resolution (30 m) soil moisture values as predicted by the RB upscaling method is shown in Fig. 3. The figure covers the 33-km SMAP grid of the Carman CVS. This particular example shows the effects of a prominent gradient in soil texture and, to a lesser extent, topography, oriented in the northeast-to-southwest direction. Specifically, the northeastern portion of the grid has a high clay fraction and low elevation, while the southwestern portion of the grid has a high sand fraction and a high elevation. This spatial distribution results in far higher values of estimated soil moisture in the northeastern portion of the grid than in the southwestern portion.

Typical performance plots from a single-trial run of the RB upscaling method are shown in Fig. 4 for the case $N = N_{\text{mid}}$. The plots show soil moisture as estimated using 1) the RB upscaling method ($\text{sm}_{\text{RB}}(N,t,i)$) and 2) the CVS SF ($\text{sm}_{\text{SF}}(i)$) plotted versus the average of validation station measurements ($\text{sm}_V(N,t,i)$) for each of the four CVSs. It can be seen that results for both the RB upscaling and the CVS SF upscaling exhibit reasonable agreement with the corresponding mean validation station measurements. Individual run results were quite variable, however; performance statistics were averaged over many runs.

2) Layer Importance: Overall plots of the relative importance of the various input data layers to the regression, as computed by random forests, are shown in Fig. 5. The importance values shown in these plots are simply averages across all runs spanning the range $N = N_{\text{mid}} - 6$ to $N_{\text{mid}} + 6$. In all cases, it should be noted that the relative importance of each spatial data layer was highly dependent on the heterogeneity of its spatial distribution, not just on how directly that spatial variable impacts soil moisture. Highly uniform spatial data layers were, thus, found to be relatively unimportant.

Fig. 5(a) presents importance values for the Carman CVS. As was the case in [13], the most important spatial variables at the site are elevation and clay fraction. Here, the clay and elevation data layers both exhibit a prominent change in average values between the northeastern (low elevation and clayey) and southwestern (high elevation and sandy) portions of the site. Other sites do not exhibit similarly strong soil texture gradients. The importance of the elevation is likely the result of its covariance with soil texture because the general topography of the area is flat except for a gradual 20–30 m drop in average elevation that occurs (in tandem with the sand-to-clay transition in soil texture) in the vicinity of a line dividing the southwestern portion of the site from the northeastern portion of the site.

Fig. 5(b), which presents importance values for the South Fork CVS, indicates that slope and aspect are the most important spatial variables (although not by a wide margin). This may to some extent be a consequence of the region's somewhat undulating topography. The site includes prominent riverbeds that seem to have relatively steep slopes, but since stations were not actually installed in the riverbeds, it is not clear how the steepness of

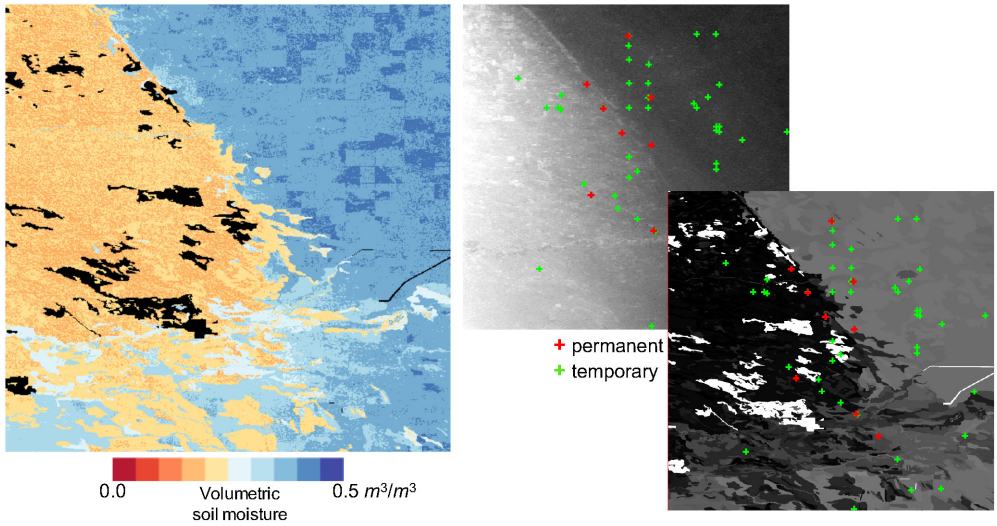

Fig. 3. Typical raster of soil moisture estimates (in m^3/m^3) from a single run (Carman, vertical orientation) of the RB upscaling method based on using randomly selected temporary stations for training and with the remaining stations averaged together for validation. Grayscale images in the upper right show the clay fraction and DEM input data layers with permanent (red) and temporary (green) station locations overlaid.

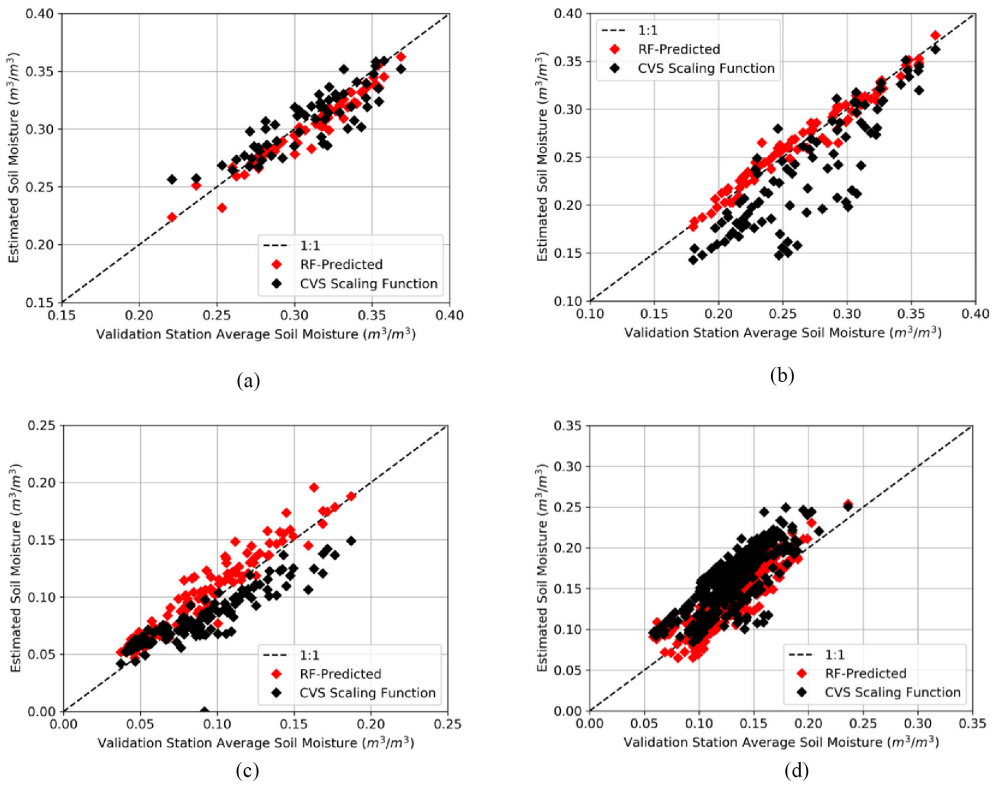

Fig. 4. Typical soil moisture estimation results from individual runs of 1) the RB (RF-predicted) upscaling method trained based on randomly selected temporary stations, and 2) the CVS SF based on permanent stations. All temporary stations not used for training are averaged together for validation. (a) Carman, vertical orientation. (b) South Fork. (c) Walnut Gulch. (d) Little River.

their banks could have contributed to the high importance of slope and aspect. The relative importance of temperature could reflect spatial gradients in local temperature occurring in the vicinity of the river channels.

Fig. 5(c), which presents importance values for the Walnut Gulch CVS, shows that the NLCD layer had virtually no importance. This is due to the highly monochromatic nature of the landcover throughout the Walnut Gulch region, particularly at the temporary station locations. The most important spatial variables were temperature, elevation and, to a lesser extent, slope, and aspect. Precipitation is only the sixth most important spatial variable for this region when considered over all dates. This is partly because precipitation was nonzero for only about half of the dates in the test duration. When layer importance is

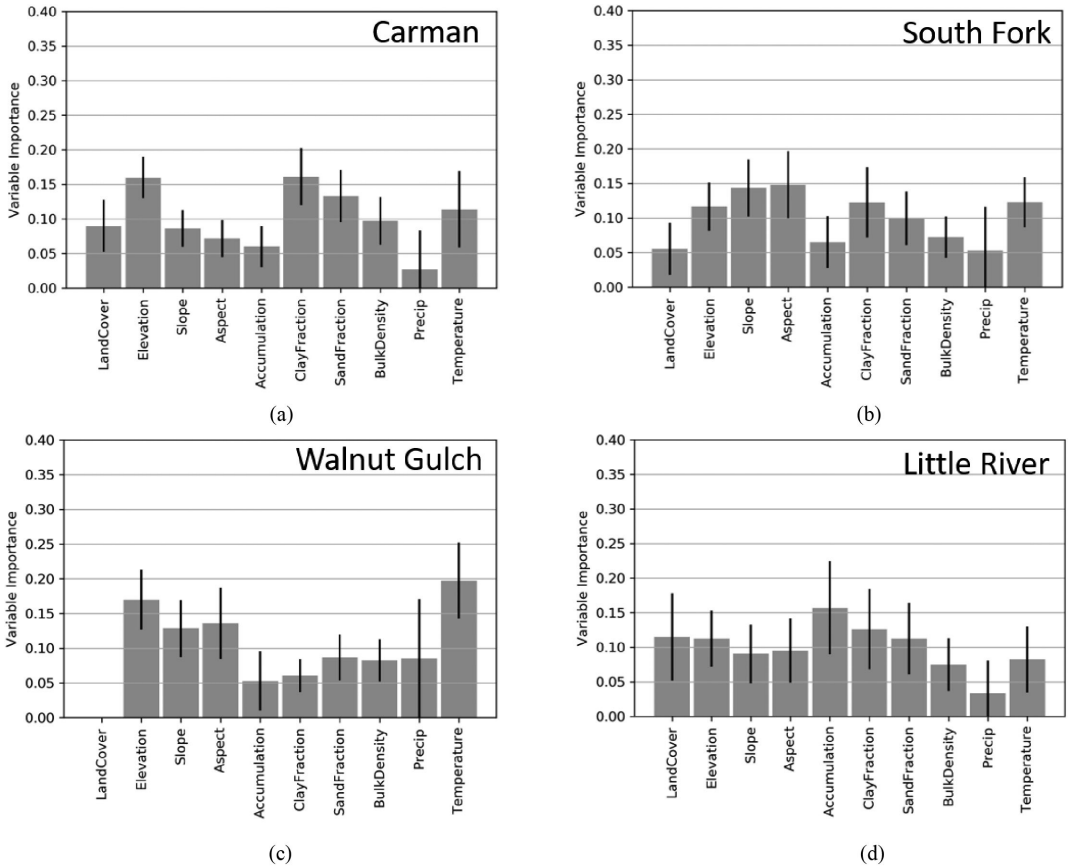

Fig. 5. Overall relative importance of the input data layers to the RB upscaling method. (a) Carman, vertical orientation. (b) South Fork. (c) Walnut Gulch. (d) Little River.

calculated with zero-precipitation dates excluded, precipitation becomes the third most important variable, after temperature and elevation.

Given the lack of importance of the NLCD data layer for the Walnut Gulch CVS, several additional runs were tried with the NLCD data layer replaced by one of several different data layers based on metrics of surface and topsoil rock content available in the SSURGO database. The metrics included percent area of rocky outcrops, percent area coverage by surface rock fragments, percent weight of large and small rock fragments within the top soil horizon, soil weight fraction unable to pass a number 4 sieve, and soil weight fraction unable to pass a number 10 sieve. While the layer importance for each of these trial "rock fraction" layers was greater than that of the NLCD data layer (see Fig. 6), overall error performance results were no better than had been obtained using the NLCD. Since there was significant uncertainty as to whether any of these SSURGO-derived variables constitute an appropriate indicator of rock fraction, however, further investigation into how best to account for the effects of rock fraction on soil moisture might yield performance improvements.

Fig. 5(d), which presents importance values for the Little River CVS, shows that flow accumulation was the most important spatial variable for this CVS, as a consequence of the presence of a dense network of large and small streambeds tracing throughout the site. The streams also affect the soil composition: soils of the wetland forests lining the streambeds

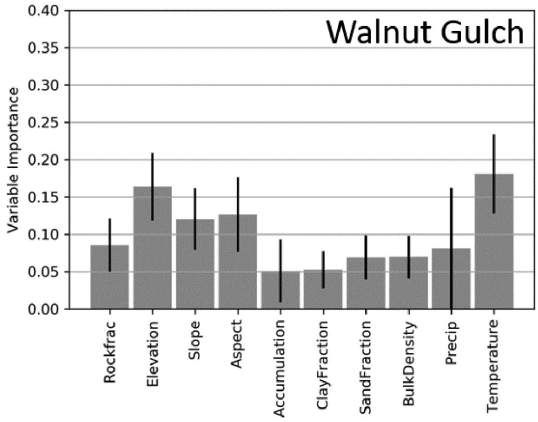

Fig. 6. Overall relative importance of the input data layers to the RB upscaling method for Walnut Gulch with a "rock fraction" layer in place of the NLCD. In the absence of information to the contrary, we assumed the "rock fraction" layer to include approximately equal fractions contributed from 1) area fraction of rocky outcrop, 2) area fraction covered by surface rock fragments, 3) weight fraction of large rock fragments within the top soil horizon, 4) weight fraction of small rock fragments within the top soil horizon, and 5) soil weight fraction unable to pass a number 4 sieve within the top soil horizon. These categories are available from the SSURGO database.

contain relatively high percentages of clay, while soils of the floodplains adjacent to these wetlands contain very high percentages of sand. These spatial relationships likely explain the

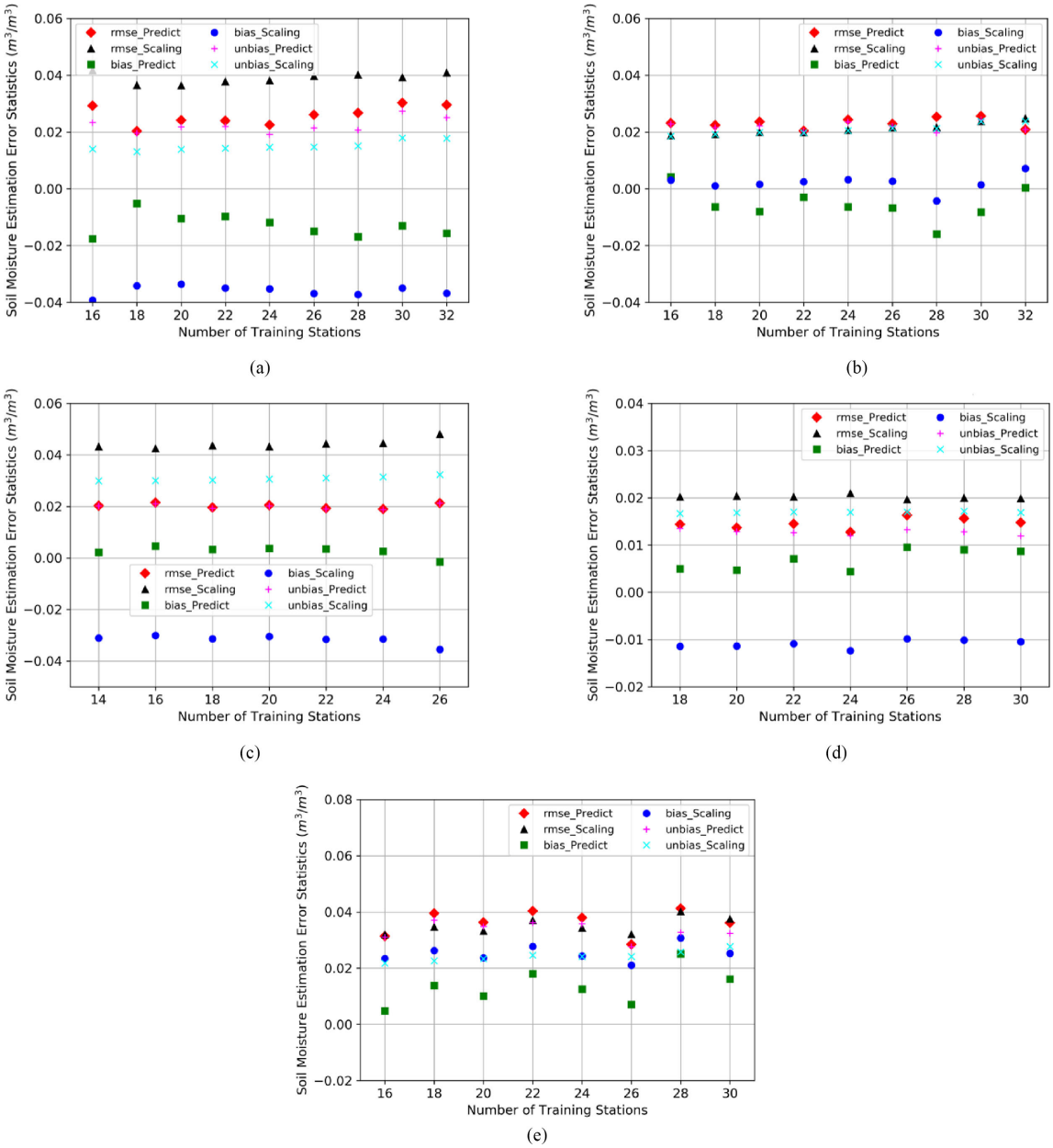

Fig. 7. Performance results: RF-predicted and CVS SF error statistics (in m^3/m^3) versus the number of temporary stations used to train the random forests regression. (a) Carman, horizontal orientation. (b) Carman, vertical orientation. (c) South Fork. (d) Walnut Gulch. (e) Little River.

relatively high importance of the clay fraction and sand fraction data layers.

3) Overall Error Performance of the RB and CVS SF Upscaling Methods: Overall error performance statistics for the random forests RB upscaling method using the temporary network and the CVS SF using the permanent network are shown in Fig. 7. The error performance is not a strong function of N for any of the sites. This is consistent with previous results [13] demonstrating that random forests generally works well even with sparse training data as long as those data are spatially representative.

Error performance statistics were also collected for Little River with the addition of either NDVI or EVI dynamic layers. The results were slightly better for most higher values of N but slightly worse for lower values of N; overall, the changes were minor.

Summary statistics for the RB upscaling method using the temporary networks averaged across all runs spanning the range $N = N_{\text{mid}} - 6$ to $N_{\text{mid}} + 6$ and the CVS SF using the permanent networks are shown in Table II.

B. Uncertainty in CVS SF Bias

In Table III, the uncertainty in the CVS SF bias has been assessed according to the two different metrics defined in Section III-C. Results are averaged over all runs spanning the range $N = N_{\text{mid}} - 6$ to $N_{\text{mid}} + 6$.

TABLE II
SUMMARY ERROR STATISTICS (RMSE, BIAS, UNBIASED RMSE) FOR THE RANDOM FORESTS RB UPSCALING METHOD AND CVS SF AVERAGED OVER ALL RUNS FOR N IN THE RANGE $N_{\text{MID}} - 6$ TO $N_{\text{MID}} + 6$ (ALL UNITS M^3/M^3) (h: HORIZONTAL SENSORS, v: VERTICAL SENSORS)

	Temporary networks with RF scaling			Permanant networks with CVS scaling		
Site Name	**RMSE**	**bias**	**ub**	**RMSE**	**bias**	**ub**
Carman (h)	0.026	-0.013	0.022	0.039	-0.036	0.016
Carman (v)	0.023	-0.007	0.022	0.022	0.002	0.021
South Fork	0.020	0.003	0.020	0.044	-0.032	0.031
Walnut Gulch	0.015	0.007	0.013	0.020	-0.011	0.017
Little River	0.037	0.015	0.034	0.036	0.026	0.025

The dense temporary network RF scaling results are taken as the gold standard for assessing the CVS biases.

TABLE III
METRICS FOR CVS SF BIAS (h: HORIZONTAL SENSORS, v: VERTICAL SENSORS)

Site Name	**Mean difference of biases** (m^3/m^3)	**RMS difference of biases** (m^3/m^3)	**Percent of allowed ubRMSE**
Carman (h)	-0.022	0.022	56.2
Carman (v)	0.009	0.009	22.6
South Fork	-0.034	0.034	85.7
Walnut Gulch	-0.018	0.018	44.7
Little River	0.011	0.011	28.0

Only the mean difference of biases metric provides information about the sign of the bias uncertainty; here, it shows that the horizontal Carman, South Fork, and Walnut Gulch have significant negative biases. The mean difference of biases metric and the rms difference of biases metric both indicate that the uncertainty in the CVS SF bias can be large, particularly in the case of the South Fork and horizontal Carman results. The magnitude of the value for South Fork, 0.034 m^3/m^3, for example, is more than 80% as large as the entire 0.04 m^3/m^3 upper limit on overall ubRMSE allowed for SMAP soil moisture retrievals.

V. DISCUSSION

A. Validation Dataset

The bias in the CVS SF was calculated relative to an average of soil moisture measurements collected by the validation dataset, that is, the subset of temporary stations not selected for training the random forests RB model and, thus, available for use as validation stations. Permanent stations were not included in the validation dataset since those are used to derive the CVS SF soil moisture estimates; as such, they do not constitute an independent source of soil moisture data from that produced using the CVS SF. Temporary stations used to train the random forests regression were also not included in the validation dataset because, having been used to calculate the RB soil moisture estimates, they do not constitute an independent source of soil moisture data from that derived using the random forests RB upscaling method. Had the stations used to train the random forests regression also been used for validation, the RB model's estimation bias, as calculated in (2), would have been unrealistically small. This would prevent the RB upscaling method from providing a realistic measure of best achievable bias against which the CVS SFs estimation bias could be assessed. With the validation dataset limited to unselected temporary stations, biases in the CVS SF estimates and those in the random forests upscaling estimates can both be quantified relative to the same set of validation measurements, facilitating a fair comparison.

B. Layer Importances

Soil texture is one of the primary factors in defining soil moisture as it affects the water retention capacity and infiltration rate of the soil; for a diverse landscape (sands to clays), it can account for up to 50% of the variability of the data signal. Therefore, it is not surprising that random forests found soil texture to be important at the Carman CVS. The relatively marginal change in elevation at the site has likely been molded by geological processes together with the soil texture.

The North American Monsoon drives much of the soil moisture variability at the Walnut Gulch CVS [44]. The importance of temperature is likely the result of the cooling of the atmosphere during the monsoon events in the summer and fast dry down rates while the weather warms up again. Temperature is correlated with rock fraction, vegetation cover and topography. Its dependence on rock fraction results from the fact that, since rock is dry, it heats up quickly, leading to higher daytime temperatures. The dependence of temperature on topography follows since flat areas get some wind, but hills are ocluded. Furthermore, within the Walnut Gulch region, elevation is believed to correlate with rock fraction. These relationships, in concert with the relatively strong topographic relief seen in the region, likely explain the relative importance of the topographic variables at the site.

The relative importance of flow accumulation at the Little River CVS is not surprising since floodplains bordering the numerous large and small streambeds throughout the site likely dominate the soil-moisture signature. Specifically, there is a dense network of large and small streambeds tracing throughout the site, which have distinct characteristics in terms of soil wetness and vegetation. Woodlands and wetlands characterize these areas; some stations were targeted in these wetland forests to capture inundation. Therefore, as the flow accumulation captures these features better than the NLCD it becomes the most important layer.

C. Error Statistics

Consistent with Fig. 7, the RMSE and bias error statistics for the RB estimates based on temporary stations are similar to or better than those of the CVS SFs based on permanent networks, although this is not always the case for the unbiased RMSE statistics. Specifically, for the vertically oriented sensors at the Carman CVS and the (horizontally oriented) sensors at South Fork and Walnut Gulch, the unbiased RMSE statistics for the regression-based upscaling are similar to or better than those of the original CVS SF using the permanent network. For the horizontally oriented sensors at Carman and the (horizontally oriented) sensors at Little River, however, the unbiased RMSE statistics for the RB upscaling are worse than those of original CVS SF. In both of the latter cases, the RMSE values for the two methods are at least roughly comparable but the bias errors of the CVS SF are much larger than those obtained using the RB upscaling. As can be seen in (3) and (6), this results in smaller values of unbiased RMSE for the CVS SF estimates than for the random forests RB estimates.

In the case of Little River, the CVS SF error statistics are comparable to those obtained using the RB method. This good performance may partly have resulted from a successful effort to apply measurements collected from temporary stations in woodland areas (which, due to practical limitations, contain no permanent stations) to correct for biases in the CVS SF estimates that might otherwise have resulted from a failure to capture the effects of woodland areas within the 33-km SMAP grid. Random forests offers an alternative and potentially more systematic and uniform way of accounting for the effects of such landscape heterogeneity across the sites, in that the method itself does not require adaptation to the characteristics of each individual site and only needs site-specific ancillary data layers.

VI. Conclusion

Recent analyses have highlighted the importance of maintaining a clear accounting for biases in characterizing the performance of soil moisture retrievals [45]. Since CVS SF upscaled soil moisture estimates are used in the calibration and validation of the SMAP soil moisture products, any bias errors corrupting the estimates have the potential to impact the accuracy assessment and, therefore, mitigation of the SMAP products. It is consequently very important to determine the range of values CVS SF bias errors can assume at a given site.

We have derived the following two new metrics, 1) the mean difference of biases and 2) the rms difference of biases, to assess bias errors in CVS SF soil moisture estimates. Both measures offer a straightforward and consistent means of quantifying biases in the soil moisture estimates generated using the original CVS SF and can be used to bound a significant source of SMAP soil moisture retrieval bias. Computation of the metrics is systematic and can be applied to any site that, for at least a limited period of time, has installed a relatively dense network of soil moisture sensors, or even a relatively sparse network of sensors as long as the latter still realizes sufficient spatial representativeness to achieve a diverse sampling of values for the different data layers. Metric (1) provides a measure of how closely biases in the CVS SF soil moisture estimates approach the best achievable value of bias. Metric (2) helps to determine the absolute value uncertainty in the biases. The metrics have been applied to evaluate biases in the upscaled soil moisture estimates of four different CVSs where dense temporary reference networks were deployed.

The results of this analysis show that, for at least some CVSs, the magnitude of the uncertainty in the CVS SF bias can be more than 80% as large as the upper limit on SMAP's entire allowable overall ubRMSE. If left uncorrected, such large CVS bias uncertainties could prove to be a significant confounding factor in the overall bias assessment for soil moisture estimates from SMAP.

The results of this research demonstrate the importance of quantifying the bias uncertainties in the CVS SF soil moisture estimates. As has been shown, this can be accomplished by combining soil moisture measurements collected from dense (but possibly temporary) networks of sensors according to a systematic approach such as the random forests RB upscaling method. The resulting bias uncertainty estimates can be applied to correct for and thereby reduce biases in the CVS scaling function soil moisture estimates used to validate SMAP's soil moisture retrievals.

Appendix

A. SMAPVEX16-MB Site in Carman, Manitoba

The Carman CVS is operated by Agriculture and Agri-Food Canada (AAFC). It is located in the Canadian Red River Watershed, an agricultural region with land cover including such crops as soybeans, wheat, and canola. Southwestern portions of the site are very sandy, while northeastern portions are typically about 20 m lower in elevation and very clayey. As described in [46], the region often experiences extreme variations in soil moisture, sometimes drought, and sometimes excessive moisture. The Carman CVS, including the location of permanent and temporary stations within it, is shown in Fig. 8.

The site included nine permanent stations, all contained within the 33-km SMAP grid. Each permanent station implemented three horizontally oriented Stevens HydraProbe soil moisture sensors at 5 cm depths plus three vertically oriented Stevens HydraProbes extending through the 0-6 cm depth range, with measurements collected at 15-min intervals (the permanent stations also had horizontal sensors installed at three other depths, but this analysis includes only those installed at the

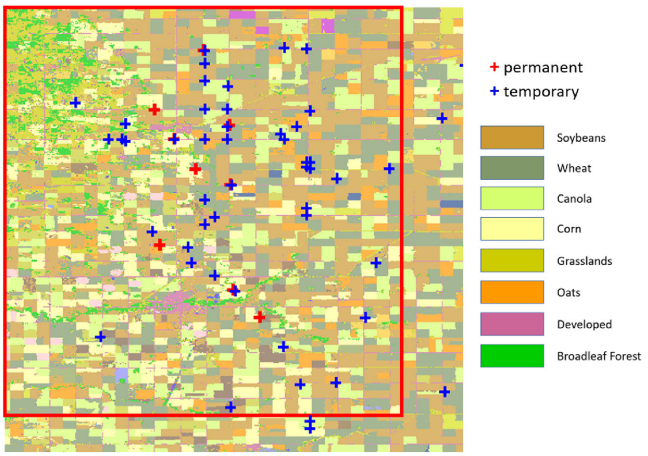

Fig. 8. Carman CVS. Red crosses denote permanent stations, blue crosses denote temporary stations, and the square-shaped red line denotes the CVS's 33-km SMAP grid. The background is taken from the AAFC crop inventory, and depicts an assortment of row crops.

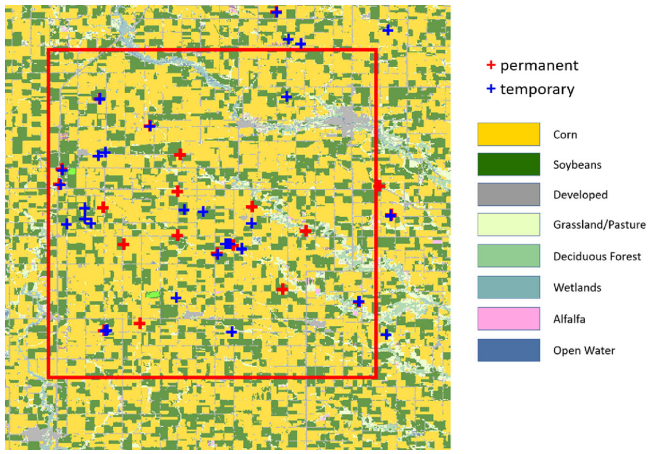

Fig. 9. South Fork CVS. Red crosses denote permanent stations, blue crosses denote temporary stations, and the square-shaped red line denotes the CVS's 33-km SMAP grid. The background is taken from the CDL, and is dominated by corn and soybeans.

TABLE IV
NONZERO STATION WEIGHTS FOR CARMAN CVS SF (EACH STATION IMPLEMENTS THREE HORIZONTAL SENSORS AND THREE VERTICAL SENSORS)

Perm Station	Weight
1	17.5
2	13
3	13
5	13
6	13
7	17.5
8	13

5 cm depth). Horizontally oriented sensors make it possible to sample soil moisture at a specific depth, while vertically installed sensors capture surface soil moisture behavior that corresponds better with what the SMAP radiometer observes [8].

Sensor weights applied to the permanent station soil moisture measurements in the CVS upSF were selected by the CVS partners based on soil type and land cover [9]; they are listed in Table IV.

Of the nine permanent stations (which included 20 horizontal and 20 vertical soil moisture sensors at 5 cm depth that were operational during the time interval analyzed here), only seven were assigned nonzero weights so as to be included in the CVS SFs weighted average; these included 17 horizontal and 17 vertical operational sensors. The stations were distributed on both sides of the gradient in soil texture and elevation, thereby capturing enough spatial variability to be able to achieve reasonable results.

The Carman CVS also included 50 temporary stations, of which 45 were within the 33-km SMAP grid, that sampled different agricultural fields for a two-month period in summer 2016. Each temporary station implemented one horizontally oriented Stevens HydraProbe at 5 cm depth and one vertically oriented Stevens HydraProbe at 0–6 cm depth, with measurements taken at hourly intervals. Temporary station data were collected throughout the SMAPVEX16-MB field campaign in June–July 2016 [47], [48].

B. SMAPVEX16-IA Site in South Fork, Iowa

The South Fork CVS (see Fig. 9) is located in an experimental watershed operated by the USDA-ARS Hydrology and Remote Sensing Laboratory, the National Laboratory for Agriculture and Environment, and the University of Iowa. The watershed is predominantly covered by corn and soybean croplands; small amounts of pasture, deciduous forest, and wetlands are also present. The region has an undulating topography (the "pothole region of Iowa"), but with low relief and poor surface drainage; large ditches have been dug to get water out of the fields as quickly as possible, however. It lies in a temperate, water-limited northern plain characterized by the typical Midwest seasonal precipitation pattern exhibiting maximum rainfall and runoff during the summer (growing season) months [49]. The soils vary at small spatial scales between being relatively sandy and relatively clayey loams and silty loams with low permeability.

The network includes 15 permanent stations of which 14 were within the 33-km SMAP grid. Each permanent station had one Stevens HydraProbe soil moisture sensor installed horizontally at a 5 cm depth, collecting soil moisture at hourly intervals. Sensor weights applied to the permanent station soil moisture measurements in the CVS upscaling function were based on a Voronoi diagram [9], in which the area is partitioned into regions based on the distance of the stations to each other within the SMAP grid (Thiessen Polygons in [15]).

The installation also included 39 temporary stations for the growing season of 2016 of which 31 were within the 33-km SMAP grid. Each temporary station had one Stevens HydraProbe installed horizontally at a depth of 5 cm, collecting soil moisture at half-hour intervals. Temporary station data were collected throughout the SMAPVEX16-IA field campaign in May–August 2016 [48].

C. SMAPVEX15 Site in Walnut Gulch, Arizona

The Walnut Gulch CVS (see Fig. 10) is operated by the Southwest Watershed Research Center of the USDA

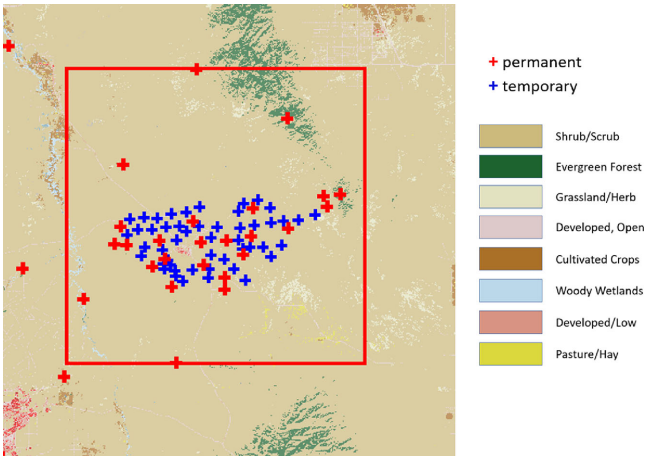

Fig. 10. Walnut Gulch CVS. Red crosses denote permanent stations, blue crosses denote temporary stations, and the square-shaped red line denotes the CVS's 33-km SMAP grid. The background is taken from the NLCD, and shows the region to be predominantly scrub/shrub, with smaller patches of evergreen forest and grasslands.

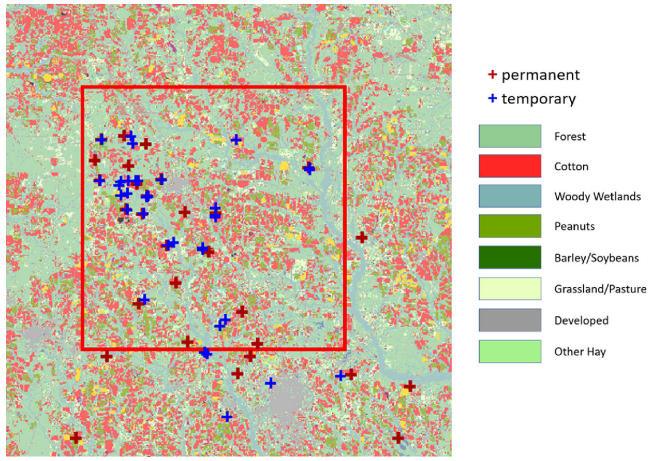

Fig. 11. Little River CVS. Red crosses denote permanent stations, blue crosses denote temporary stations, and the square-shaped red line denotes the CVS's 33-km SMAP grid. The background is from the CDL, and shows the region to include cotton and other row crops, evergreen forest, woody wetlands, and pasture.

Agricultural Research Service. It is in a semiarid rangeland dominated by desert open shrub, although it also includes small areas of evergreen forest, desert grasses, pasture, and wetland. The Walnut Gulch site exhibits large topographic variations, with elevations ranging from 1100 to 2000 m over the site. Soils are relatively sandy (sand fraction over 60%) throughout the region, mostly consisting of well-drained, sandy, and gravelly loams [50]. Rock fraction, a land surface parameter that tends to correlate with elevation within the region, is also indicative of soil moisture and hydrology. As described by Colliander *et al.* [51], most annual precipitation comes during the summer monsoons in the form of highly localized thunderstorms that result in a great deal of spatial variability in precipitation.

A total of 28 permanent stations were in the region included in the analysis, of which 25 were contained within the 33-km SMAP grid. Each permanent station had one Stevens HydraProbe soil moisture sensor installed horizontally at a depth of 5 cm and collected soil moisture at half-hour intervals. Sensor weights applied to the permanent station soil moisture measurements in the CVS upscaling function were based on a Voronoi diagram. Only 19 permanent stations (all within the SMAP grid) had nonzero weights so as to be included in the CVS SFs weighted average.

A total of 49 temporary stations were deployed in 2015 and included in the analysis, all contained within the 33-km SMAP grid. They had Stevens HydraProbe sensors installed horizontally at 5 cm depth and measured soil moisture at half-hour intervals. Temporary station data were collected over four months, July–October 2015, bracketing the SMAPVEX15 field campaign of August 2015 [51].

D. Little River Site Near Tifton, Georgia

The Little River CVS (see Fig. 11), operated by the USDA-ARS Southeastern Watershed Research Laboratory, is in a humid agricultural region largely covered by forest and row crops such as peanuts and cotton. The watershed also contains significant stands of evergreen forest and lesser amounts of woody wetlands, pasture, and deciduous forest. As described in Bosch *et al.* [52] and Sullivan *et al.* [53], the region contains numerous small low-gradient streams. Since slopes are shallow, water is slow to drain out of the area, and riparian areas are frequently inundated. Soils are mostly loamy sand (sand fraction over 60%) throughout the region.

A total of 33 permanent stations were included in the analysis, of which 21 were contained within the 33-km SMAP grid. Each permanent station had one Stevens HydraProbe soil moisture sensor installed horizontally at a depth of 5 cm and collected soil moisture at half-hour intervals. The CVS upscaling function used a two-step approach to estimate the areal average soil moisture. First, sensor weights applied to the permanent station soil moisture measurements in the CVS upscaling function were based on a Voronoi diagram. Second, the average soil moisture was corrected with the temporary network measurements, which captured soil moisture conditions in woodland areas within the pixel that were not sampled by the permanent network.

A total of 48 temporary stations were included in the analysis, of which 43 were within the 33-km SMAP grid. Most stations implemented two or three sensors of two different types (one or two Stevens HydraProbes and one CS655); for these locations, the sensor measurements were averaged together. The sensors were installed horizontally at 5 cm depth and measured soil moisture at hourly intervals. Temporary station data were collected for over a year, starting approximately in May 2017 and continuing (in some instances) into May 2018.

E. More on Random Forests

As mentioned, random forests generates a "forest" of decision trees. Each decision tree is constructed in accordance with the classification and regression trees algorithm [54]. To do this, random forests first randomly selects N samples (with replacement, so as not to alter the characteristics of the pool as selections are made) from the available N-sample pool. It can be shown [55] that this results in the selection of approximately $2N/3$ unique

samples; these samples are used to train the regression. The remaining $\sim N/3$ unique samples are used to calculate accuracy statistics for the decision tree's regression model. In order to reduce correlation between trees, only a subset of the available data layers, n_{split} layers, is used to split each decision tree node. A decision tree is formed by starting with all of its samples placed in a single node, then successively splitting each (parent) node based on the values of n_{split} randomly selected data layers in such a manner as to minimize the mean squared error of the resulting (child) nodes [26]. That is, given a node containing N_s samples $\{y_s\}$, random forests chooses the split that minimizes

$$\text{MSE}(N_s) = \frac{1}{N_s} \sum_{s \in N_s} \left(y_s - \frac{1}{N_s} \sum_{s \in N_s} y_s \right)^2 .$$

Splitting continues until all nodes either contain a single sample or have a mean squared error less than a threshold value. Random forests average soil moisture estimates for each location across all decision trees to form an overall estimate of soil moisture at that location. Each such estimate can be compared to the location's *in situ* soil moisture value to assess accuracy.

REFERENCES

[1] D. Entekhabi *et al.*, "SMAP handbook, soil moisture active passive, mapping soil moisture and freeze/thaw from space, JPL CL#14-2285," Jet Propulsion Laboratory, Calif. Inst. Technol., Pasadena, CA, USA, Tech. Rep. JPL 400-1567, 2014.

[2] 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 National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado USA, 2020. [Online]. Available: https://doi.org/10.5067/Q8J8E3A89923

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

[4] S. Chan *et al.*, "Development and assessment of the SMAP enhanced passive soil moisture product," *Rem. Sens. Env.*, vol. 204, pp. 931–941, 2018.

[5] S. Chan, "Enhanced level 2 passive soil moisture product specification document, prime mission release," Jet Propulsion Laboratory, Calif. Inst. Technol., Pasadena, CA, USA, Tech. Rep. JPL D-56291, 2018.

[6] M. Brodzik, B. Billingsley, T. Haran, B. Raup, and M. Savoie, "Incremental but significant improvements for earth-gridded data sets," *ISPRS Int. J. Geo-Inf.*, vol. 1, no. 1, pp. 32–45, Jan. 2012.

[7] M. Brodzik, B. Billingsley, T. Haran, B. Raup, and M. Savoie, "Correction: Brodzik, M.J., *et al. EASE-Grid 2.0*: Incremental but Significant Improvements for Earth-Gridded Data Sets. *ISPRS International Journal of Geo-Information* 2012, *1*, 32–45," *ISPRS Int. J. Geo-Inf.*, vol. 3, no. 3, pp. 1154–1156, 2014.

[8] T. Jackson *et al.*, "Soil moisture active passive (SMAP) project: Calibration and validation for the L2/3_SM_P version 5 and L2/3_SM_P_E version 2 data products," Jet Propulsion Laboratory, Calif. Inst. Technol., Pasadena, CA, USA, Tech. Rep. JPL D-56297, 2018.

[9] A. Colliander *et al.*, "Validation of SMAP surface soil moisture products with core validation sites," *Rem. Sens. Env.*, vol. 191, pp. 215–231, 2017.

[10] J. Adams, H. McNairn, A. Berg, and C. Champagne, "Evaluation of near-surface soil moisture data from an AAFC monitoring network in manitoba, canada: Implications for L-band satellite validation," *J. Hydrol.*, vol. 521, pp. 582–592, 2015.

[11] M. Yee, J. Walker, A. Monerris, C. Rudiger, and T. Jackson, "On the identification of representative in situ soil moisture monitoring stations for the validation of SMAP soil moisture products in Australia," *J. Hydrol.*, vol. 537, pp. 367–381, 2016.

[12] F. Chen *et al.*, "Uncertainty of reference pixel soil moisture averages sampled at SMAP core validation sites," *J. Hydrometeorol.*, vol. 20, no. 8, pp. 1553–1569, 2019.

[13] D. Clewley *et al.*, "A method for upscaling in situ soil moisture measurements to satellite footprint scale using random forests," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2663–2673, Jun. 2017.

[14] 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, no. 2, pp. 1–20, Apr. 2012.

[15] S. Dingman, *Physical Hydrology*. Long Grove, IL, USA: Waveland Press, 2015.

[16] M. Cosh *et al.*, "Strategies for validating satellite soil moisture products using *in situ* networks: Lessons from the USDA-ARS watersheds," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 2015–2018.

[17] 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.

[18] A. Colliander *et al.*, "An assessment of the differences between spatial resolution and grid size for the SMAP enhanced soil moisture product over homogeneous sites," *Rem. Sens. Env.*, vol. 207, pp. 65–70, 2018.

[19] K. Vinnikov, A. Robock, S. Qiu, and J. Entin, "Optimal design of surface networks for observation of soil moisture," *J. Geophys. Res.*, vol. 104, pp. 19743–19749, 1999.

[20] M. Cosh, T. Jackson, P. Starks, and G. Heathman, "Temporal stability of surface soil moisture in the Little Washita River watershed and its applications in satellite soil moisture product validation," *J. Hydrol.*, vol. 323, pp. 168–177, 2006.

[21] M. Cosh, T. Jackson, S. Moran, and R. Bindlish, "Temporal persistence and stability of surface soil moisture in a semi-arid watershed," *Rem. Sens. Env.*, vol. 112, pp. 304–313, 2008.

[22] T. Burns, A. Berg, J. Cockburn, and E. Tetlock, "Regional scale spatial and temporal variability of soil moisture in a prairie region," *Hydrological Processes*, vol. 30, no. 20, pp. 3639–3649, 2016.

[23] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[24] L. Breiman and A. Cutler, "Random forests website at Berkeley," Accessed: 2020. [Online]. Available: https://www.stat.berkeley.edu/~breiman/RandomForests

[25] D. Clewley, J. Whitcomb, and M. Moghaddam, SoilSCAPE Upscaling Code v1.0.0 (Version v1.0.0). Zenodo, 2017. [Online]. Available: http://doi.org/10.5281/zenodo.556203

[26] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[27] Scikit-learn developers, "Scikit-learn user guide," Scikit-Learn GitHub Organization, San Francisco, CA, USA, Release 0.20.3, 2019.

[28] N. Chaney, J. Roundy, J. Herrera-Estrada, and E. Wood, "High-resolution modeling of the spatial heterogeneity of soil moisture: Applications in network design," *Water Resour. Res.*, vol. 51, no. 1, pp. 619–638, 2015.

[29] S. Bircher, N. Skou, K. Jensen, J. Walker, and L. Rasmussen, "A soil moisture and temperature network for SMOS validation in Western Denmark," *Hydrol. Earth Syst. Sci.*, vol. 16, pp. 1445–1463, 2012.

[30] Agriculture and A.-F. Canada, "Annual crop inventory," 2016. [Online]. Available: http://www.agr.gc.ca/atlas/aci/

[31] U. N. A. S. Service, "CropScape—Cropland Data Layer," 2018. [Online]. Available: https://nassgeodata.gmu.edu/CropScape/

[32] J. Fry *et al.*, "Completion of the 2006 National Land Cover Database for the Conterminous United States," *Photogrammetric Eng. Remote Sens.*, vol. 77, no. 9, pp. 858–864, 2012.

[33] S. Jin, L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian, "A comprehensive change detection method for updating the National Land Cover Database to circa 2011," *Rem. Sens. Env.*, vol. 132, pp. 159–175, 2013.

[34] NASA Land Processes Distributed Active Archive Center (LP DAAC), "SRTMGL1: NASA shuttle radar topography mission global 1 arc second V003." [Online]. Available: https://lpdaac.usgs.gov/products/srtmgl1v003/

[35] D. Maune, Ed., *The National Elevation Dataset* (Digital Elevation Model Technologies and Applications: The DEM Users Manual), 2nd ed. Rockville, MD, USA: American Society for Photogrammetry and Remote Sensing, 2007, pp. 99–118.

[36] D. Clewley, "CalcSlopeDegrees." 2015. [Online]. Available: https://github.com/MiXIL/calcSlopeDegrees

[37] P. Bunting, D. Clewley, R. M. Lucas, and S. Gillingham, "The remote sensing and GIS software library (RSGISLib)," *Comput. GeoSci.*, vol. 62, pp. 216–226. [Online]. Available: http://dx.doi.org/10.1016/j.cageo.2013.08.007

[38] Environmental Systems Research Institute (ESRI), "ArcGIS desktop website," Accessed: 2020. [Online]. Available: https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview

[39] Agriculture and A.-F. Canada, "The National Soil Database (NSDB)," 2013. [Online]. Available: http://sis.agr.gc.ca/cansis/nsdb/index.html

[40] Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Accessed: 2020. [Online]. Available: https://websoilsurvey.nrcs.usda.gov/

[41] GDAL/OGR contributors, *GDAL/OGR Geospatial Data Abstraction Softw. Libr.*, Open Source Geospatial Foundation, 2020. [Online]. Available: https://gdal.org
[42] European Centre for Medium-Range Weather Forecasts (ECMWF), "Reanalysis datasets, ERA-interim," Accessed: 2020. [Online]. Available: http://www.ecmwf.int/en/research/climate-reanalysis
[43] Oregon State University PRISM Climate Group, "PRISM climate data," Copyright, 2020. [Online]. Available: http://www.prism.oregonstate.edu
[44] D. Adams and A. Comrie, "The North American monsoon," *Bull. Amer. Meteorological Soc.*, vol. 78, no. 10, pp. 2197–2213, 1997.
[45] S. Zweiback *et al.*, "Estimating time-dependent vegetation biases in the SMAP soil moisture product," *Hydrol. Earth Syst. Sci.*, vol. 22, pp. 4473–4489, 2018.
[46] H. McNairn *et al.*, "SMAPVEX16-MB, SMAP validation experiment 2016 in Manitoba, Canada, experimental plan," *Agriculture Agri-Food Canada, Sci. Technol. Branch,* Winnipeg, MB, Canada, 2016.
[47] H. A. K. M. Bhuiyan *et al.*, "Assessing SMAP soil moisture scaling and retrieval in the Carman (Canada) study site," *Vadose Zone J.*, vol. 17, no. 1, 2018, Art. no. 180132.
[48] 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.
[49] E. Coopersmith, M. Cosh, W. Petersen, J. H. Prueger, and J. Niemeier, "Soil moisture model calibration and validation: An ARS watershed on the South Fork Iowa River," *J. Hydrometeorol.*, vol. 16, pp. 1087–1101, 2015.
[50] Southwest Watershed Resource Center and USDA-Agricultural Research Service, "Southwest watershed research center & walnut gulch experimental watershed," Southwest Watershed Research Center, Tucson, AZ, USA, 2007.
[51] A. Colliander *et al.*, "Validation and scaling of soil moisture in a semi-arid environment: SMAP validation experiment 2015 (smapvex15)," *Rem. Sens. Env.*, vol. 196, pp. 101–112, 2017.
[52] D. Bosch, O. Pisani, A. W. Coffin, and T. Strickland, "Water quality and land cover in the coastal plain little river watershed, Georgia, USA," *J. Soil Water Conservation*, vol. 75, no. 3, pp. 263–277, May 2020.
[53] D. G. Sullivan, H. L. Batten, D. Bosch, J. Sheridan, and T. Strickland, "Little river experimental watershed, Tifton, Georgia, United States: A geographic database," *Water Resour. Res.*, vol. 43, 2007, Art. no. W09471, doi: 10.1029/2006WR005836.
[54] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification Regression Trees*. New York, NY, USA: Chapman and Hall, 1993.
[55] J. Whitcomb, M. Moghaddam, K. McDonald, J. Kellndorfer, and E. Podest, "Mapping vegetated wetlands of Alaska using L-band radar satellite imagery," *Can. J. Remote Sens.*, vol. 35, no. 1, pp. 54–72, 2009. [Online]. Available: http://pubs.casi.ca/doi/abs/10.5589/m08-080

Jane Whitcomb (Member, IEEE) received the B.S. degree from Tufts University, Medford, MA, USA, in 1979 and the M.S. degree from the National Technological University, Fort Collins, CO, USA, in 2003, both in electrical engineering.

She participates in research with the University of Southern California, Los Angeles, CA, USA. This has included generating large-scale thematic mapping of boreal wetlands based on space-based *L*-Band synthetic aperture radar (SAR) and upscaling of *in situ* soil moisture measurements for calibration and validation of SMAP space-based radiometer soil moisture retrievals.

Daniel Clewley received the B.Sc. degree in physics, in 2006, M.Sc. degree in 2007, and Ph.D. degree in 2012, both in GIS and remote sensing from Aberystwyth University, Wales, U.K.

He works as Research Software Engineer with the Centre for Geospatial Applications, Plymouth Marine Laboratory, Plymouth, U.K. His work focuses on developing software and techniques for processing and generating products from Earth observation data spanning a range of application areas including soil moisture, vegetation mapping, and marine.

Andreas Colliander (Senior Member, IEEE) received the M.Sc. (Tech.), Lic.Sc. (Tech.), and D.Sc. (Tech.) degrees in electrical engineering 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. His research is focused on development of microwave remote sensing techniques. He is currently leading the calibration and validation of the geophysical retrievals of NASA's SMAP mission.

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. His current research interests include the monitoring of soil moisture from both *in situ* resources and satellite products. He conducts research on satellite calibration and validation for such missions as the soil moisture active passive and soil moisture ocean salinity missions.

Jarrett Powers received the B.Sc. degree in agriculture from the University of Manitoba, in 1989 and a Postgraduate Certificate in applied and theoretical geographic information systems from Simon Fraser University, in 2001.

From 1991 to 2012, he was employed with Agriculture and Agri-Food Canada (AAFC) – Prairie Farm Rehabilitation Administration, delivering environmental and water development programming in Manitoba and later managing a regional geomatics group that provided analysis for environmental and water studies. In 2012 he joined AAFC – Science and Technology Branch where he develops and provides scientific support number to soil science research including the management of AAFC.s soil moisture research network (RISMA).

Matthew Friesen received the bachelor's degree in environmental studies from the University of Manitoba, Winnipeg, MB, Canada.

He began his career with Agriculture and Agri-Food Canada, Ottawa, MB, Canada, in 2015, contributing to the understanding of satellite microwave retrievals by leveraging *in situ* soil moisture networks.

Heather McNairn received the Bachelor of Environmental Studies degree in geography from the University of Waterloo, Waterloo, ON, Canada, in 1987, the M.Sc. degree in soil science from the University of Guelph, Guelph, ON, Canada, in 1992, and the Ph.D. degree in geography from Université Laval, Quebec City, QC, Canada, in 1999.

She is a Senior Research Scientist with Agriculture and Agri-food Canada (AAFC), Ottawa, ON, Canada. She has 30 years of research experience in developing methods to monitor soils and crops using multispectral, hyperspectral, and Synthetic Aperture Radar (SAR) sensors. This experience has included the analysis of multifrequency (*X*-, *C*-, and *L*-Band), multipolarization, and fully polarimetric SAR data acquired from ground-based scatterometers, airborne SARs, and various satellite platforms. She has led numerous national and international research teams. She is an Adjunct Professor with Carleton University, Ottawa, ON, Canada, University of Manitoba, Winnipeg, MB, Canada, and Nipissing University, North Bay, ON, Canada.

Aaron A. Berg received the Ph.D. degree in earth system science from the University of California Irvine, in 2003.

He is currently a Canada Research Chair (Tier II) and a Professor with the Department of Geography, Environment and Geomatics, University of Guelph, Guelph, ON, Canada. He leads a research program that is broadly focused on the observation, modeling, and analysis of soil moisture anomalies using hydrological models and satellite observations from several remote-sensing platforms. More recently his research program has explored the use of UAV platforms using a variety of sensors for applications in agricultural remote sensing.

David D. Bosch received the Ph.D. degree in hydrology from The University of Arizona, Tucson, AZ, USA, in 1990.

In 1986, he joined the Agricultural Research Service. He leads a Watershed Research Program investigating the impacts of land use on water balance and quality. He has been active in the validation of remotely sensed soil moisture products, since 2000. He is currently a Research Hydrologist with the Southeast Watershed Research Laboratory, Agricultural Research Service, Tifton, GA, USA. His primary research interests include watershed and landscape scale hydrology, agricultural impacts on water quality, hydrologic and solute transport modeling of watershed processes, riparian buffer hydrology and solute transport, and developing new methods for assessing the impact of agricultural chemicals on ground and surface water supplies.

Alisa Coffin received the B.S. degree in environmental studies from Rutgers University, in 1985, the Master of Landscape Architecture (MLA) degree in landscape architecture from Harvard's Graduate School of Design, in 1993, and the Ph.D. degree in geography (with minor in road ecology) from the University of Florida, in 2009.

She is currently a Research Ecologist with the USDA-ARS Southeast Watershed Research Laboratory, Tifton, GA, USA. Her research works include agroecosystem processes and dynamics from field to watershed scales, incorporating remote sensing, and geospatial analysis. She works in multidisciplinary teams integrating results from diverse data sources using spatial frameworks to characterize outcomes related to differences in the management of agricultural systems.

Chandra Holifield Collins received the Ph.D. degree in soil water and environmental science with a minor in remote sensing from the University of Arizona, Tucson, AZ, USA, in 2006.

She is currently a Soil Scientist with the USDA-Agricultural Research Service's Southwest Watershed Research Center, Tucson, AZ, USA. Her research interests include image analysis and the use of remote sensing data for agricultural applications, with current focus on operational tools for rangeland management.

John H. Prueger received the Ph.D. degree in micrometeorology with the Utah State University, Logan, UT, USA.

He is currently a Research Soil Scientist with the USDA Agriculture Research Service, National Laboratory for Agriculture and the Environment, Soil, Water and Air Resources Research Laboratory, Wyndmoor, PA, USA. His research interests include soil science, agricultural plant science, and agronomy.

Dara Entekhabi (Fellow, IEEE) received the B.S. degree, in 1983 and the M.S. degrees in, 1985 and 1988 in geography from Clark University, Worcester, MA, USA, and the Ph.D. degree in civil and environmental engineering from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1990.

He is currently a Professor with the Department of Civil and Environmental Engineering and the Department of Earth, Atmospheric and Planetary Sciences, MIT. He is currently the Science Team Lead for the National Aeronautics and Space Administration's Soil Moisture Active and Passive (SMAP) mission that was launched in January 31, 2015. His research includes terrestrial remote sensing, data assimilation, and coupled land–atmosphere systems modeling.

Prof. Entekhabi is a Fellow of the American Meteorological Society and the American Geophysical Union. He is a member of the National Academy of Engineering.

Mahta Moghaddam (Fellow, IEEE) received the B.S. degree in electrical and computer engineering in 1986 from the University of Kansas, Lawrence, KS, USA, with highest distinction, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1989 and 1991, respectively.

She is the holder of the Ming Hsieh Chair of Electrical and Computer Engineering with the University of Southern California, Los Angeles, CA, USA. Prior to that, she was a Professor of Electrical and Computer Engineering with the University of Michigan, Ann Arbor, MI, USA, from 2003 to 2011 and with the NASA Jet Propulsion Laboratory (JPL) from 1991 to 2003. She has introduced new approaches for quantitative interpretation of multichannel radar imagery based on analytical inverse scattering techniques applied to complex and random media. She was a Systems Engineer for the Cassini Radar and was a Science Chair of the JPL Team X (Advanced Mission Studies Team). Her most recent research interests include the development of new radar instrument and measurement technologies for subsurface and subcanopy characterization, development of forward, and inverse scattering techniques for layered random media especially for soil moisture and permafrost applications, geophysical retrievals using signal-of-opportunity reflectometry, and transforming concepts of radar remote sensing to medical imaging and therapy systems.

Dr. Moghaddam is a member of the Soil Moisture Active and Passive (SMAP) Mission Science Team and a member of the Cyclone Global Navigation Satellite System (CYGNSS) Science Team. She was the Principal Investigator of the AirMOSS NASA Earth Ventures 1 mission and is a member of the National Academy of Engineering.