Estimating High-Resolution Soil Moisture Over Mountainous Regions Using Remotely-Sensed Multispectral and Topographic Data

Lei Fan, Amen Al-Yaari, Frédéric Frappart, Jian Peng, Jianguang Wen, Qing Xiao, Rui Jin, Xiaojun Li, Xiangzhuo Liu, Mengjia Wang, et al.

Estimating High-Resolution Soil Moisture Over Mountainous Regions Using Remotely-Sensed Multispectral and Topographic Data

Lei Fan, Amen Al-Yaari, Frédéric Frappart, Jian Peng, Jianguang Wen, Qing Xiao, Rui Jin, Xiaojun Li, Xiangzhuo Liu, Mengjia Wang, Xiuzhi Chen, Lin Zhao, Mingguo Ma, and Jean-Pierre Wigneron, *Fellow, IEEE*

***Abstract*—A surface soil moisture (SM) condition at high spatiotemportal resolutions is required by regional Earth system applications. Here, we mapped daily 1-km SM in the Babao River Basin in the northwest of China during the summers from 2013 to 2015 using a random forest (RF) method by merging SM information retrieved from *in situ* measurements, optical/thermal remote sensing, and topographical indices. Relative importance analysis was used to determine the optimal predictors for estimating high-resolution SM. A specific RF model (RF_{VI}+sup) was constructed using the optimal predictors including remote sensing albedo, apparent thermal inertia (ATI), normalized difference vegetation index, normalized difference infrared index 5, soil adjusted vegetation index, and topographical indices (aspect and elevation). The RF_{VI+sup} also accounted for missing observations of the thermal index (e.g., ATI) over the mountainous regions. In the comparison between the SM estimates using the new RF_{VI+sup} model and other RF models, the spatial coverage of available estimates increased from 14% to 64% over the study region, the correlation coefficient values were improved to 0.75, the unbiased root-mean-squared difference values decreased to 0.032 m^3/m^3. Thus, the proposed RF method provided accurate SM estimates with high spatiotemporal resolution over the mountainous regions, by merging multiresource datasets from *in situ* measurements, remotely-sensed, and topographical indices.**

***Index Terms*—High resolution, mountainous regions, optical index, random forest (RF) method, soil moisture (SM), thermal index.**

Manuscript received December 9, 2021; revised February 23, 2022; accepted April 8, 2022. Date of publication April 26, 2022; date of current version May 18, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFA0608703, in part by the National Natural Science Foundation of China under Grant 42171339, Grant 41830648, and Grant 41801247, and in part by the Postdoctoral Start-Up Project of Southwest University under Grant SWU020016. *(Corresponding author: Lei Fan.)*

Lei Fan and Mingguo Ma are with the Chongqing Jinfo Mountain Karst Ecosystem National Observation and Research Station, School of Geographical Sciences, Southwest University, Chongqing 400715, China (e-mail: fanlei20088@163.com; mmg@swu.edu.cn).

Amen Al-Yaari is with University of Grenoble Alpes, IRD, CNRS, Grenoble INP, IGE, Grenoble, France (e-mail: amen.alyaari@gmail.com).

Frédéric Frappart is with UMR 1391 ISPA, INRAE, F-33140 Villenave d'Ornon, France, and also with the Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS), 31400 Toulouse, France (e-mail: frederic.frappart@gmail.com).

Jian Peng is with the Department of Remote Sensing, Helmholtz Centre for Environmental Research−UFZ, 04318 Leipzig, Germany, and also with the Remote Sensing Centre for Earth System Research, Leipzig University, 04103 Leipzig, Germany (e-mail: jian.peng@ufz.de).

Jianguang Wen and Qing Xiao are with the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100049, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: wenjg@radi.ac.cn; xiaoqing@radi.ac.cn).

Rui Jin is with the Key Laboratory of Remote Sensing of Gansu Province, Heihe Remote Sensing Experimental Research Station, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China (e-mail: jinrui@lzb.ac.cn).

Xiaojun Li, Xiangzhuo Liu, Mengjia Wang, and Jean-Pierre Wigneron are with UMR 1391 ISPA, INRAE, F-33140 Villenave d'Ornon, France (e-mail: xiaojun.li@inra.fr; xiangzhuo.liu@inrae.fr; mengjia.wang@inrae.fr; jean-pierre.wigneron@inra.fr).

Xiuzhi Chen is with the Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou 510275, China (e-mail: chenxzh73@mail.sysu.edu.cn).

Lin Zhao is with the School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: lzhao@nuist.edu.cn).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2022.3166974

I. Introduction

SURFACE soil moisture (SM) is the key to adjusting land surface energy partition, controlling vegetation transpiration, and surface runoff [1]–[5]. As a consequence of these needs, many microwave remote sensing observations were used to estimate SM time series at a global scale [6], [7]. The application of passive microwaves at *L*-band has been considered one of the most promising methods [8]–[11], although higher frequencies of passive and active microwave observations also demonstrated good potential [12]. However, current passive microwaves missions (e.g., SMOS [9] and SMAP [8]) are limited in spatial resolution (lower than 25 km), greatly limiting potential applications at regional scales (1–10 km). In an effort to produce higher resolution SM datasets, the SMAP and active microwave Sentinel-1 observations were merged to produce the most recent 3-km SM product (L2_SM_SP) [13], [14]. However, these SM products using active microwave observations are limited over the mountain regions, which can be attributed to two aspects: 1) Sentinel-1 has a low temporal resolution (6 or 12 days), which results in the low temporal resolution (equal or larger than 12 days over China) [15]. Especially over the Northwest regions of China, very few high-quality SM information from L2_SM_AP is available; 2) the effect of mountains on the backscatter of SAR or Sentinel-1 is too complex to be modified, which leads

to the high uncertainties in the backscatter, further decreases the accuracy of high-resolution SM [16].

Currently, downscaling microwave-derived coarse-resolution SM products by the synergistic coupling of optical/thermal-infrared datasets is the widely used approach to estimate high-resolution ($\sim$1 km) SM information [17]. These downscaling approaches can be classified as empirical regression methods [18]–[20] and semiphysical methods [21], [22], both relying on the indirect relationships between SM and optical/thermal-infrared indices. In the downscaling method, a range of indices are used to provide high-resolution SM proxies, such as surface albedo, apparent thermal inertia (ATI), optical vegetation indices, land surface temperature (LST), temperature vegetation dryness index [23]–[25], vegetation temperature condition index [26], [27], and the soil evaporative efficiency [21], [22], [28].

The performance of downscaling approaches is limited by the following two issues.

1) the mathematical functions could not describe the complex relationship between SM and optical/thermal-infrared indices over the mountainous regions. Previous analysis indicated high variations in the relationship between SM and optical/thermal-infrared indices [20], [27], [29]. Moreover, the relationship between SM and optical/thermal-infrared indices could be more complex over mountainous regions, because of the complex interactions among meteorological, topographical, and vegetation factors [30]. Few studies were made to evaluate these optical/thermal-infrared indices on the estimation of SM over the mountainous regions.
2) uncertainties associated with the coarse-resolution microwave SM products may degrade the accuracy of the downscaled high-resolution SM [17]. In particular, over mountainous regions, topographic effects could further increase the uncertainties in the optimization of some parameters (e.g., vegetation properties and surface roughness) used in the retrieval models for estimating microwave SM products [9].

A promising alternative strategy to estimate high-resolution SM is to extrapolate the *in situ* measurements to larger scales by using geostatistical techniques [30]. Using the wireless sensor networks (WSNs) [31], [32], several upscaling methods were proposed to retrieve high-resolution SM, including block kriging [33], [34], the Bayesian maximum entropy method [35], [36], and the Bayesian linear regression (BLR) method [37]–[39]. The latter was successfully used to estimate high-resolution SM over the mountainous regions [38], [39], due to its ability to overcome the issue of temporal discontinuity in high-resolution SM products. Note that all upscaled methods require the representative *in situ* measurements, for instance, to be located in the shady and sunny slopes for different elevations.

For BLR upscaling approach, a reference map of "true" SM was first calculated as the ATI-retrieved SM, then the calibrated relationship between the field-measured SM and the reference SM was used to obtain high-resolution (1-km) SM [37], [39]. A simple mathematical function with fixed empirical coefficients was used to retrieve SM using ATI. This function could not be an optimum expression reflecting the actual relationship between SM and ATI as it is likely that the SM patterns are not spatially uniform due to heterogeneity in the meteorological conditions, the local land use types, and the vegetation cover, especially in mountainous regions. This could increase the uncertainties in the reference ATI-retrieved SM, which is a key to the BLR estimation [38].

An alternative method to compute high-resolution SM maps is machine learning (ML) techniques [e.g., artificial neural network and random forest (RF)] [40], [41]. ML techniques started to be used for downscaling satellite SM products [41]–[44], due to its ability to merge the SM information derived from several optical-derived variables [45]. Previous investigations have found, in particular, that the RF algorithm can be used to merge multisource observations and address the issue of a nonlinear relationship between predictors and predictions [46], [49]. The RF method is able to overcome the overfitting issues by generating independent regression trees through randomly selecting training samples, due to the low correlation between each independent tree [50].

This study aims to estimate the daily 1-km resolution SM over Babao River Basin, by proposing a framework to combine SM information from multispectral images and topographic information. The proposed method is to decrease the potential uncertainties in the SM estimations over the mountainous regions. Our approach relies on remote sensing and field-measurement information (e.g., WSN measurements, aspect, and elevation from ASTER digital elevation model and MODIS reflectance and LST data). The WSN measurements, MODIS reflectance and LST, and topography data are presented in Section II; the proposed RF method and the metrics for evaluating the RF-derived SM are given in Section III; and results, discussions, and conclusions are given in Section IV–VI, respectively.

II. Study Area and Datasets

A. Study Area and Field Campaigns

The Babao River Basin, located upstream of the Heihe River Basin [see Fig. 1(a)], is the main water source for agriculture in the mid- and downstream areas of the Heihe River Basin [51], [52]. The developments of the Heihe River Basin have been limited by the scarcity of water resources [51]. Thus, high-resolution SM information of the Babao River Basin will be critical for the water resource management in the whole Heihe River Basin.

The Babao river basin, mainly covered with grasslands [see Fig. 1(b)], has an area of about 2500 km^2 and an elevation ranging from $\sim$2600 to 5100 m [see Fig. 1(a)] [53]. The runoff in summer-wet months is mostly from precipitation. The spatial–temporal variability of SM is mainly affected by the heterogeneous spatial distribution of precipitation and topographic conditions, such as shady and sunny slopes [32].

As one of the core study areas in the Heihe Watershed Allied Telemetry Experimental Research (HiWATER) project, the whole Babao River Basin was instrumented with the Ecological and Hydrological WSN [EHWSN; Fig. 1(a) and Table S1] for monitoring SM and soil temperature [31]. EHWSN, consisting of 37 *in situ* stations, was deployed according to an optimal

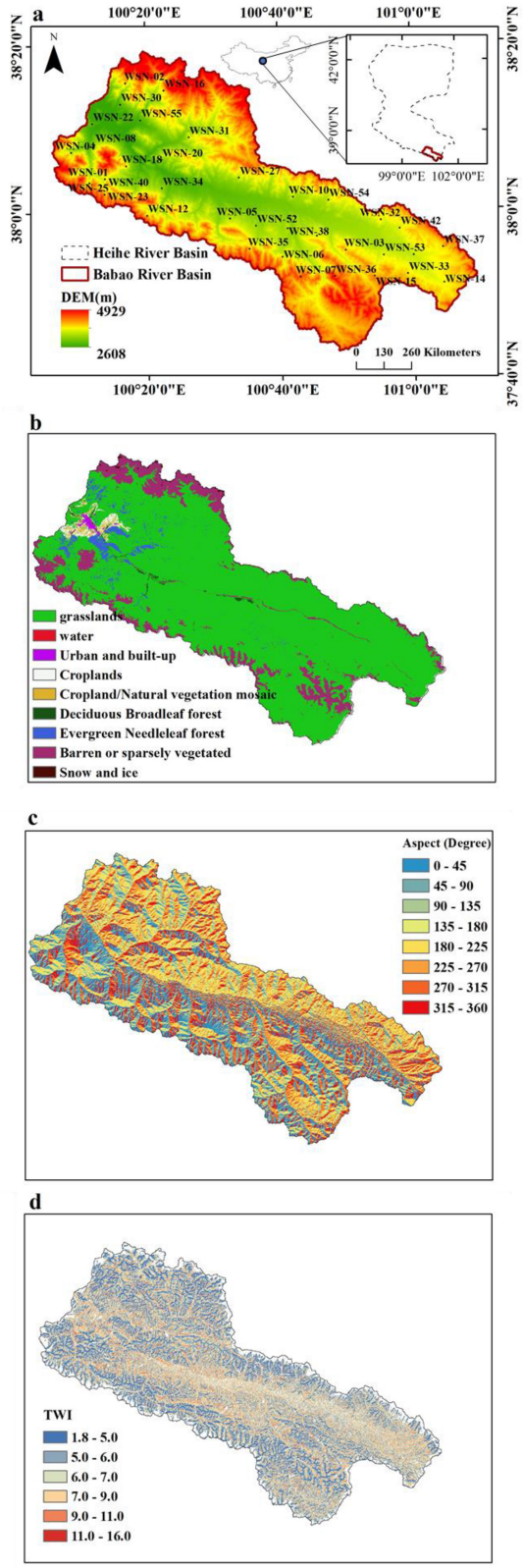


Fig. 1. Study region. The spatial distributions of *in situ* measurements stations of the study region were shown in (a), the corresponding land cover, terrain aspect, and TWI were indicated in (b), (c), and (d), respectively. In (a), elevation information is also shown and the WSN sites are marked as black points. In (b), land cover types data at a spatial resolution of 30 m provided by the HiWATER project [52].

design to better capture the spatial variations of SM [54]. The optimal sampling design of the network is obtained using a spatially simulated annealing algorithm by considering the multivariate (cross-) correlation and spatial trend. The results demonstrate that this sampling method can capture the SM dynamics over the mountainous regions [55].

SM and soil temperature at depths of 4, 10, and 20 cm at 5-min intervals were measured by EHWSN from 2013 to 2015. The study period was 15th July to 15th October of 2013, 2014, and 2015 to avoid snow periods covering from November to May [33]. *In situ* measurements at a depth of 4 cm within a time window of 1 h were matched with the instantaneous overpass times of the MODIS LST product. *In situ* measurements from 2013 to 2014 were used for training the SM model, and *in situ* measurements in 2015 were used to validate the SM estimates.

B. Optical/Thermal-Infrared Indices

Optical vegetation indices can be used for monitoring the vegetation water status, which indicates an increase or decrease in the spatial variability of SM, due to that vegetation can affect the upward/downward flow of water vapor [56]. MODIS 500-m daily nadir BRDF-Adjusted surface reflectance product (MCD43A4) [57] was used here to calculate several spectral indices (see Table S2) that are impacted directly or indirectly by SM: NDVI, albedo, soil adjusted vegetation index (SAVI), visible atmospheric resistant index (VARI), normalized difference water index (NDWI), normalized difference infrared index 6 (NDII6), normalized difference infrared index 7 (NDII7), and global vegetation moisture index (GVMI). The lower NDWI/NDII6 values indicate the lower leaf vegetation water content. The 500-m resolution spectral indices were aggregated into 1 km to match the spatial resolution of the LST product by a simple average method.

The thermal-infrared indices, such as the difference between LST at daytime and nighttime (ΔLST) and ATI, have been proved to be related to SM over regions with low vegetation density [33], [38], [58]. ΔLST was calculated using the day- and night-time daily LST provided by the MODIS 1-km Terra LST product (MOD11A1) [59]. Higher SM content has high thermal inertia, determining its higher resistance to temperature variations. Thus, higher SM has lower ΔLST values. ATI was retrieved using LST from MOD11A1 and albedo derived from MCD43A4.

C. Topographic Indices

Over the mountainous regions, the spatial patterns of SM are also influenced by the topographic conditions, such as terrain aspect and elevation [30], thus topographic indices could reflect SM information. The topographic wetness index (TWI) [60] has been proved to be a good indicator of SM over the mountainous regions where surface flow dominates the processes of water transportation [61]–[63]. It is defined as $\ln(a/\tan\beta)$ where a is the local upslope area draining through a certain point per unit contour length and $\tan\beta$ is the local slope [60]. TWI combines local upslope contributing area and

slope and has been commonly used to quantify topographic control on hydrological processes. This potentially suggests that TWI can better represent the distribution of SM than the slope. Furthermore, the calculation of TWI was based on slope, namely that TWI has included the slope information. TWI is not directly related to the variability of SM but can identify hydrological flow paths which generally are related to the spatial distribution of SM. Thus, TWI was selected in this study instead of a slope.

Aspect, elevation, and TWI (see Table S2) were calculated using 30-m Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER) Global Digital Elevation Model (GDEM). These topographic indices were aggregated into the 1-km resolution data to match the spatial resolution of LST product using a simple average method.

III. Method

A. RF Algorithm

The ML RF method has been applied to retrieve various climate variables (e.g., SM, evapotranspiration, precipitation and fire) [46], [50], [64]–[66]. The complex and highly nonlinear relationship between the predictors and the response variable can be described by the RF method due to its advantage in involving the adaptive and decorrelated decision rules [50]. In the RF algorithm, the relationship between the predictors and the response was described using lots of decision trees. The RF algorithm produces these decision trees in the training phase, by dividing the input dataset into a number of regression trees, known as forest, where each tree is generated using a bootstrap sample method. Then, the mean prediction of the individual trees is generated as the output. A bootstrap sample includes two-thirds of the training input data, and the remaining samples (about one-third) are used for the validation of each tree. This is one of the key features of the RF algorithm to estimate the model generalization error. This approach solves the overfitting issue usually occurring in the traditional regression methods, with the help of the bagging method that merges the information from lots of decision trees.

Also, the RF method provides an analysis of the relative importance of each predictor for the response variable. Each predictor is permuted using the out-of-bag observations to estimate the importance of each predictor. It is performed for every tree in the RF framework, and changes in the prediction error are estimated. The performance of the new model is improved significantly, suggesting that the predictor is important for the performance of the original model used to estimate the response variable. The overall average of this measure is usually divided by the standard deviation value of the overall ensemble, in order to obtain a normalized measure of the predictor importance [50].

In this study, the optimal predictors (remote sensing indices, Table S2) as input of the RF model were determined based on their relative importance to the SM prediction. Then the relationship between WSN-measured SM and SM predictors was trained in the framework of the RF model. Last, the trained RF model was used to estimate the high-resolution (1-km) SM.

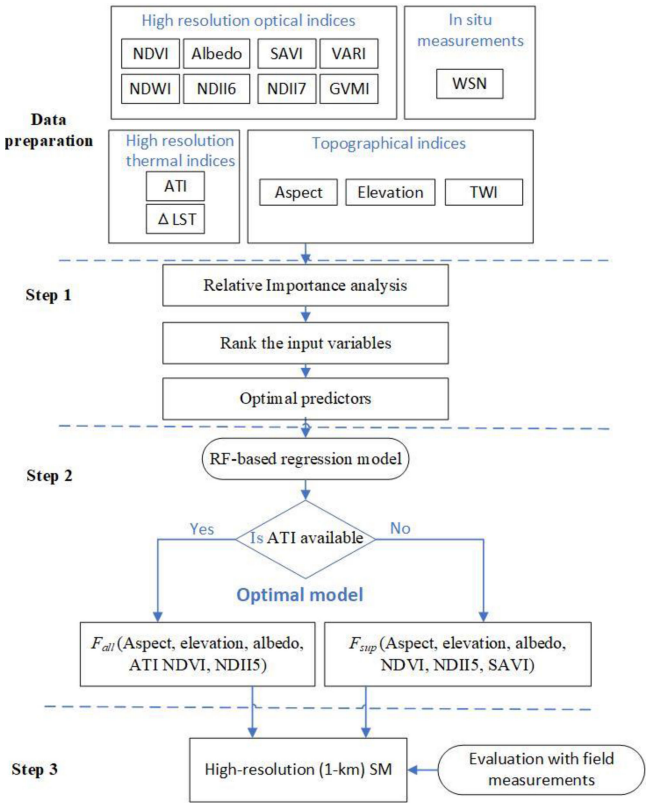


Fig. 2. Flowchart of the framework in this study.

B. RF-Derived SM

The RF-derived SM algorithm is presented in Fig. 2. First, to build the optimal model for the high-resolution SM estimates, a selection was first performed to identify an optimal set of predictors representing the spatial and temporal variability of SM. Thirteen remote sensing indices, tested as predictors, were extracted from the pixels where the WSN nodes are located. These predictors were ranked based on their relative importance to SM mapping using the RF algorithm.

Based on the relative importance of the SM predictors, all predictors were reassembled as 13 sets to estimate SM: The first ranked variable was used to predict SM using the RF model, defined as $\mathrm{RF_I}$; the second, third, …, thirteenth ranked variable was added successively in the RF model to predict SM, defined as $\mathrm{RF_{II}}, \mathrm{RF_{III}}, \ldots, \mathrm{RF_{XL}}$, respectively. The relationship between the predictors and the *in situ* measured SM was calibrated using the RF regression model. The calibrated relationship was used in the RF model to predict SM. To determine the optimal RF model, the performances of the 13 RF predictions were evaluated by comparing them against the *in situ* measured SM, using statistical metrics including the correlation coefficient (R), unbiased root-mean-squared difference (ubRMSD), bias, and slope (see Section III-D).

Based on the statistical metrics, a model (referred to as $\mathrm{RF_{VI}}$) using six predictors [aspect, elevation, albedo, ATI, normalized difference vegetation index (NDVI), and normalized difference infrared index 5 (NDII5)] was determined to retrieve SM (as presented in Section IV-A). In the $\mathrm{RF_{VI}}$ model, the linking

model (f) between SM and the predictors is expressed as follows:

$$\mathrm{RF_{VI}} = f\,(\text{aspect, elevation, albedo, ATI, NDVI, NDII5}) \tag{1}$$

where $\mathrm{RF_{VI}}$ is the retrieved SM value.

Note that the calculation of ATI, as a predictor, requires the high-quality albedo and LST observations from both daytime and nighttime, resulting in lower availability of ATI than other optical indices. Namely, the optical indices could be obtained when ATI is missing. Accounting for the missing ATI data due to the missing LST observations affected by clouds [38], an RF model ($\mathrm{RF_{VI}^{sup}}$), as a supplementary model of $\mathrm{RF_{VI}}$, was trained using the predictors without ATI to provide SM estimates at the pixels where there are missing estimates from $\mathrm{RF_{VI}}$. The supplementary RF model $\mathrm{RF_{VI}^{sup}}$ was written as

$$\mathrm{RF_{VI}^{sup}} = f\,(\text{aspect, elevation, albedo, NDVI, NDII5, SAVI}) \tag{2}$$

where SAVI, ranked as the seventh importance prediction index of SM, was added to replace ATI. $\mathrm{RF_{VI}}$ was used in all pixels, except the pixels without ATI observations, where $\mathrm{RF_{VI}^{sup}}$ was used instead. The combined model merging $\mathrm{RF_{VI}}$ and $\mathrm{RF_{VI}^{sup}}$ was defined as $\mathrm{RF_{VI+sup}}$ (3) and was used to estimate 1-km resolution SM during the study period

$$\mathrm{RF_{VI+sup}} = \text{merging}\,(\mathrm{RF_{VI}}\,, \mathrm{RF_{VI}^{sup}})\,. \tag{3}$$

C. Validation

An independent validation method was used to evaluate SM estimates against *in situ* measurements: The *in situ* measurements from 37 WSN sites during 2013–2014 were used to calibrate the $\mathrm{RF_{VI+sup}}$ model, and the *in situ* measurements in 2015 were used to validate the SM estimates using the $\mathrm{RF_{VI+sup}}$ model.

D. Evaluation Metrics

The performance of the proposed RF models was evaluated using the following.

1) The ubRMSD. A nonbiased root-mean-squared error (ubRMSE) was proposed to remove the mean bias caused by the mean of the estimates or the amplitude of fluctuations in the estimates. Accounting for that the spatial SM distribution within each 1-km grid is heterogeneous, and the *in situ* measurements also include measurement uncertainties, the term "error" of ubRMSE was replaced by "difference," that is, ubRMSD [67], which was obtained using (4) shown at the bottom of this page, in a unit of $\mathrm{m^3/m^3}$.
2) The correlation coefficient (R), which was used to evaluate the consistency between the field measurements and SM estimates, can be calculated using (5) shown at the bottom of this page.
3) Slope, which was used to evaluate the agreement between the field measurements and SM estimates at each grid, can be calculated using simple linear regression (6)

$$\theta^{\text{In situ}} = \text{slope} \cdot \theta^{\text{upscale}} + \text{intercept} \tag{6}$$

where N is the number of the WSN-measured SM in the time series. $\theta_t^{\text{In situ}}$ and $\theta_t^{\text{upscale}}$ are the WSN-measured and the RF estimated SM on date t, respectively. $\overline{\theta^{\text{In situ}}}$ and $\overline{\theta^{\text{upscale}}}$ are the average SM values of the corresponding time series.

IV. Results

A. Predictor Selection

Aspect and elevation, as topography indices, were identified as the first and second most important predictors of the RF-based SM estimation, with mean relative importance of 33% and 18% out of 100%; thus much more important than other predictors (see Fig. 3). The results about the important role of topographic indices in the spatial pattern of SM were also reported by previous studies related to the SM disaggregation using topographic data [68]–[70]. By contrast, another topography index, TWI, was only the eighth most important predictor of SM in this area. The importance of optical and thermal indices (e.g., albedo, ATI, NDVI, NDII5, and SAVI) for predicting SM is higher than that of TWI, followed by VARI, NDII7, NDII6, ΔLST, and GVMI.

Thirteen sensitivity tests were made to evaluate the performances of RF models on the SM estimates, by using different groups of predictors based on the order of importance of the predictors (see Fig. 4). The aspect was as the input used in $\mathrm{RF_I}$, the elevation, albedo, ATI, NDVI, NDII5, SAVI, TWI, VARI, NDII7, NDII6, ΔLST, and GVMI were added successively from $\mathrm{RF_{II}}$ to $\mathrm{RF_{XL}}$. The model with the best predictive performance was determined by statistical metrics (e.g., lowest ubRMSD and bias, highest R values, and slope values closer to one), which were calculated by comparing the RF-derived estimates against the *in situ* measured SM. With increasing the number of predictors in the RF model, the R values between the SM estimations and the field measurements were improved from 0.87 ($\mathrm{RF_I}$) to 0.93 ($\mathrm{RF_{VI}}$) with ubRMSD values decreasing from 0.06 ($\mathrm{RF_I}$) to 0.043 $\mathrm{m^3/m^3}$ ($\mathrm{RF_{VI}}$), albeit slope and bias

$$\text{ubRMSD} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\theta_t^{\text{upscale}} - \overline{\theta^{\text{upscale}}}\right)^2 - \left(\frac{1}{N}\sum_{t=1}^{N}\left(\theta_t^{\text{In situ}} - \overline{\theta^{\text{In situ}}}\right)\right)^2} \tag{4}$$

$$r = \frac{\sum_{t=1}^{N}\left(\theta_t^{\text{upscale}} - \overline{\theta^{\text{upscale}}}\right)\left(\theta_t^{\text{in situ}} - \overline{\theta^{\text{in situ}}}\right)}{\sqrt{\sum_{t=1}^{N}\left(\theta_t^{\text{upscale}} - \overline{\theta^{\text{upscale}}}\right)^2 \cdot \sum_{t=1}^{N}\left(\theta_t^{\text{in situ}} - \overline{\theta^{\text{in situ}}}\right)^2}} \tag{5}$$

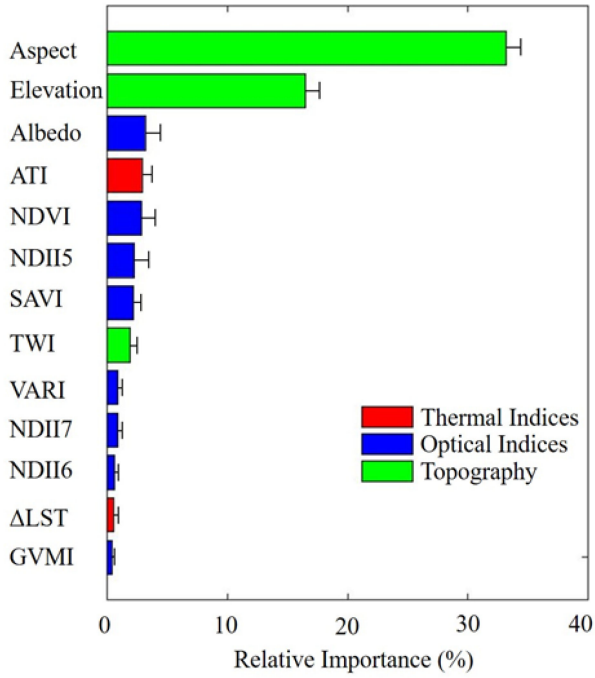


Fig. 3. Relative importance in the model. Error bars represent the standard deviation (SD; variability) of the relative importance (median+SD).

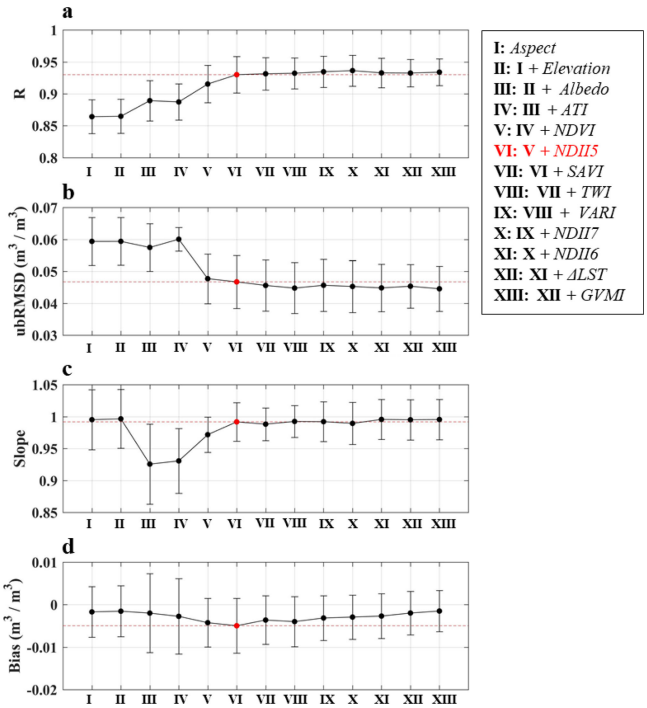


Fig. 4. Sensitivity tests of SM predictions by using different predictors based on their relative importance for SM prediction.

values were almost similar by comparing the RF_I and RF_{VI} estimates. Also, it can be seen (see Fig. 4) that all statistical metrics, except bias, did not improve much from RF_{VI} to RF_{XL}, indicated by the approximately constant R, ubRMSD, and slope values. This implies that RF_{VI} is very close to the optimal model for estimating high-resolution SM over the study region.

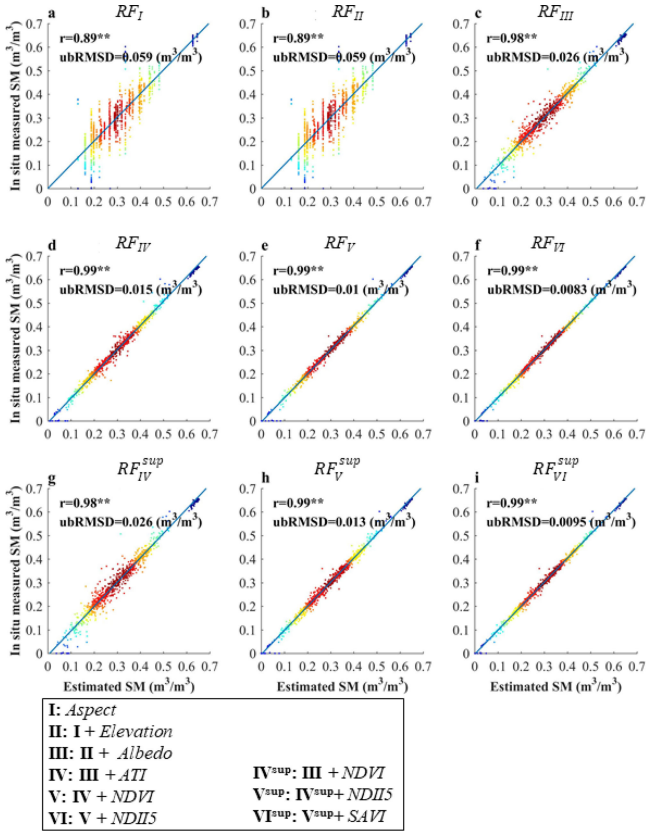


Fig. 5. Performances of RF models and the supplementary RF models using different predictors.

To evaluate the behaviors of RF_I, RF_{II}, RF_{III}, RF_{IV}, RF_V, and RF_{VI} for SM prediction, the scatter plots between the RF-derived SM estimates and the *in situ* measurements are shown in Fig. 5(a)–(f), together with the corresponding R and ubRMSD values. Relative to RF_I [see Fig. 5(a)] and RF_{II} [see Fig. 5(b)] estimations, RF_{III} [see Fig. 5(c)] agreed well with *in situ* measurements with SM predictions close to the 1:1 line, with improved metrics ($R = 0.98$ and ubRMSD $= 0.56$ m^3/m^3). The performance of the RF_{VI} model [see Fig. 5(f)] to estimate SM was found to be the best with the highest R of 0.99 and lowest ubRMSD of 0.0083 m^3/m^3. Thus, SM patterns could be accurately captured by RF_{VI} using six predictors: 1) aspect, 2) elevation, 3) albedo, 4) ATI, 5) NDVI, and 6) NDII5.

Accounting for the missing ATI observations, three supplementary RF models that did not include the ATI predictor were evaluated. These three supplementary RF models were defined as RF_{IV}^{sup} [see Fig. 5(g)], RF_V^{sup} [see Fig. 5(h)], and RF_{VI}^{sup} [see Fig. 5(i)] using the inputs of aspect, elevation, albedo, NDVI, NDII5, and SAVI. Relative to RF_{IV} [see Fig. 5(g)] and RF_V [see Fig. 5(h)], the corresponding supplementary RF models [see Fig. 5(g) and (h)], which did not include the ATI predictor have slightly lower statistical metrics. However, as the supplementary model of RF_{VI}, RF_{VI}^{sup} [see Fig. 5(i)] has an identical performance ($R = 0.99$ and ubRMSD $= 0.0095$ m^3/m^3) relative to RF_{VI}. Thus, the model combining RF_{VI} and RF_{VI}^{sup} (hereafter, RF_{VI+sup}) could be used to estimate accurately the SM patterns over the study region.

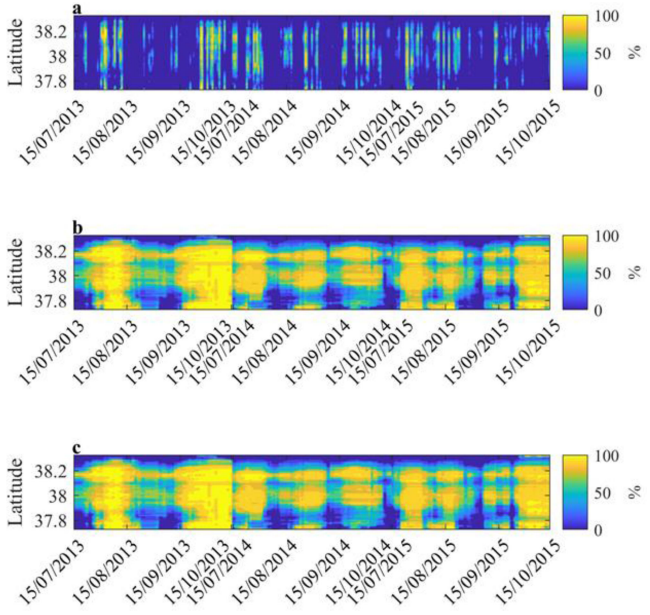


Fig. 6. Hovmöller diagrams showing available observations from the (a) RF_{VI}, (b) $RF_{VI,sup}$, and (c) RF_{VI+sup} estimates in the Babao River Basin. Available observations are in the unit of percent, calculated using the ratio between available estimates and the whole area in each latitude.

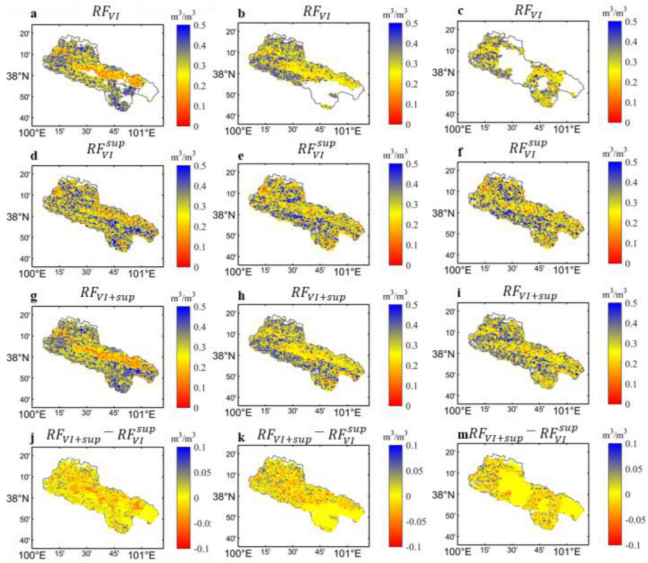


Fig. 7. Spatial distribution of the SM estimates from the RF_{VI}, $RF_{VI,sup}$, and RF_{VI+sup} models on (a), (d), and (g) 4th October, (b), (e), and (h) 8th August, and (c), (f), and (i) 12th August in 2013 over the Babao River Basin. The corresponding SM differences between RF_{VI+sup} and $RF_{VI,sup}$ were shown in (j), (k), and (m), respectively. The spatial coverages of the RF_{VI} estimates are 79% [see Fig. 4(a)], 74% [see Fig. 4(d)], and 58% [see Fig. 4(g)] of the study region, respectively.

B. RF-Based SM Estimation

The 1-km SM estimates over the Babao River Basin were derived using the RF_{VI} model (using predictors including aspect, elevation, albedo, ATI, NDVI, and NDII5), the RF_{VI}^{sup} model (using predictors including aspect, elevation, albedo, NDVI, NDII5, and SAVI) and the RF_{VI+sup} model combining the RF_{VI} and RF_{VI}^{sup} models, respectively. The RF_{VI} estimations [see Fig. 6(a)] were found to be often missing over the study period (with low spatial coverage of 14% over the study regions), especially in September. Besides, the SM patterns are difficult to observe over the southernmost and northernmost parts of the study area using the RF_{VI} model due to many missing data. The missing RF_{VI} estimates can be attributed to the missing ATI observations. This is due to the limited availability of LST, used in the calculation of ATI, because of the presence of clouds and rain over the study region most of the year [38]. As a supplementary estimate of the missing RF_{VI} estimations, the RF_{VI}^{sup} model [see Fig. 6(b)] provided a higher spatial coverage (62%) of SM estimates, albeit with missing observations over the northernmost of the study region. To replace the missing RF_{VI} estimates with the corresponding RF_{VI}^{sup} estimates at pixel-scale, the RF_{VI+sup} estimates [see Fig. 6(c)] have a high spatial coverage, in good agreement with that of RF_{VI}^{sup}.

To illustrate the performance of the RF models, the SM estimates using RF_{VI+sup} [see Fig. 7(c), (f), and (i)] over 3 days (8th August, 12th August, and 4th October in 2013) are shown in Fig. 7, together with the RF_{VI} [see Fig. 7(a), (d), and (g)] and RF_{VI}^{sup} [see Fig. 7(b), (e), and (h)] estimates. It can be seen that RF_{VI} is able to provide detailed spatial SM patterns, but only covering 79% [see Fig. 7(a)], 74% [see Fig. 7(d)], and 58% [see Fig. 7(g)] of the study region. In contrast to RF_{VI}, the RF_{VI+sup} model provided a higher spatial coverage, covering 98%, 98%, and 96% of the study region. Additionally, the spatial distribution of RF_{VI+sup} [see Fig. 7(b), (e), and (h)] agreed well with that of RF_{VI} [see Fig. 7(a), (d), and (g)], suggesting that RF_{VI}^{sup} can be a good substitute of RF_{VI} to capture the spatial distribution of SM. Moreover, compared with RF_{VI} and RF_{VI}^{sup}, we can observe that RF_{VI+sup} is able to provide more SM information related to its spatiotemporal distribution. The SM differences between RF_{VI+sup}^{sup} and RF_{VI}^{sup} showed that the spatial pattern of SM estimates using RF_{VI}^{sup} are generally close to that using RF_{VI+sup}^{sup} [see Fig. 7(j), (k), and (m)], except RF_{VI}^{sup} have higher SM estimates than RF_{VI+sup}^{sup} in the middle of the study region.

C. Evaluation of the RF Performance

The accuracy of the SM values estimated using the RF_{VI+sup} model was evaluated quantitatively by comparing them with *in situ* measured SM in 2015 using an independent validation method (see Fig. 8). The SM estimates had a good overall correlation with the *in situ* measurements with a median *R*-value of 0.75 and a range of 0.67 to 0.80 for all stations, relative to the RF_I estimates that present no correlation with the *in situ* measured SM. Also, the overall ubRMSD values were improved from 0.071 m^3/m^3 (RF_I) to 0.030 m^3/m^3 (RF_{VI+sup}), except over WSN-25 (ubRMSD = 0.049 m^3/m^3), WSN-37 (ubRMSD = 0.046 m^3/m^3), and WSN-38 (ubRMSD = 0.045 m^3/m^3). The SM estimates had slope values closer to one for all stations, with a median value of 0.83, ranging from 0.72 to 1.5. These statistical metrics indicated that the RF_{VI+sup} model provides accurate estimates of SM.

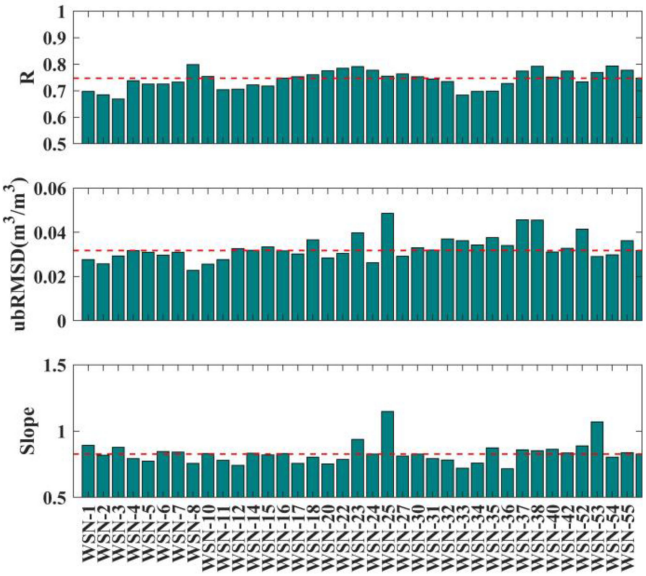


Fig. 8. Statistical metrics of the SM estimations based on the combined model of RF_{VI} and $RF_{VI,\ sup}$ (hereafter, RF_{VI+sup}) against the *in situ* SM measurements. Red dotted line represents the median value of the statistical metrics.

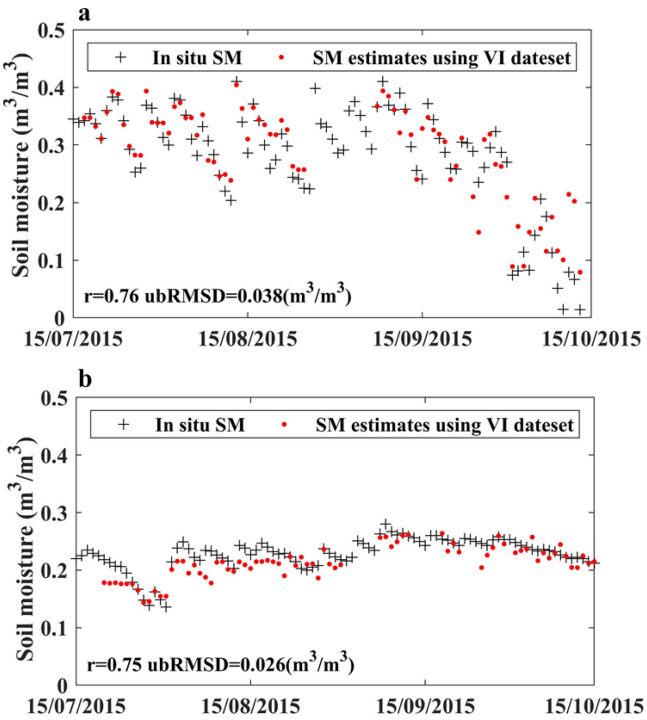


Fig. 9. SM estimates using several RF models against the *in situ* data from WSN-35 in shady slopes and WSN-10 in sunny slopes.

As an example, the time series of RF predictions (e.g., $RF_{I,}$ $RF_{II,}$ $RF_{III,}$ RF_{IV}, RF_{VI}, and RF_{VI+sup}) and *in situ* measurements at the WSN-35 site in shady slopes and WSN-10 site in sunny slopes were presented in Fig. 9. In comparison with the estimates based on RF_{VI+sup} model and other RF models at WSN-35 site [see Fig. 9(a)], the ubRMSD value decreased from 0.086 to 0.038 m^3/m^3 while the *R*-value increased from no-correlation to 0.76 (*P*-value$<$0.01). The improved performance of RF_{VI+sup} was also found at the WSN-10 site [see Fig. 9(b)], providing higher *R* values and lower ubRMSD values than the other SM estimates. Moreover, better agreements were observed between the time series of the *in situ* measured SM and the RF_{VI+sup} estimates, compared with other estimates at both WSN sites (see Fig. 9).

However, SM underestimation by RF_{VI+sup} can be observed at most sites (except WSN-25 and WSN-53 sites), as indicated by slope values < 1 (ranging from 0.72 to 0.94). Note that WSN-25 also has a high ubRMSD value, indicating inaccurate estimations by RF_{VI+sup}. The poor performance of RF_{VI+sup} over these specific sites could be attributed to the high uncertainties in optical indices and the low repressiveness of the *in situ* measurements (see Section V).

V. Discussion

Here, the RF method was proposed to improve SM estimates over mountainous areas, by merging remotely-sensed multispectral and topographic data. The proposed RF model produced accurate SM estimates over mountainous regions in the Babao River Basin. Aspect and elevation were identified as the two most important predictors to estimate SM over the mountainous region, implying that topographic factors are the key to the spatial distribution of SM over the study region associated with precipitation, which affects the SM patterns, as noted by the previous study in the same study region [38]. According to Fig. 5 in Fan *et al.* (2019), at low elevation, the average values of SM from the shady slope were 0.33 m^3/m^3, higher than the sunny slope (0.23 m^3/m^3). Similar results also were observed at high elevations that the SM from shady (0.31 m^3/m^3) was higher than the sunny slope (0.21 m^3/m^3). Based on the *in situ* measurements above, the SM at high elevation (0.31 m^3/m^3 and 0.21 m^3/m^3 from shady and sunny slopes) generally are lower than low elevation (0.33 m^3/m^3 and 0.23 m^3/m^3 from shady and sunny slopes) regions. These results indicated that the elevation and aspect are the key factors for the distribution of SM at the basin scale.

However, it is worth to be noted that the SM estimates using the aspect and elevation have the worst accuracy. This could be partly attributed to that aspect and elevation, as static predictors, cannot capture the SM dynamics, although the spatial pattern of SM could be determined by aspect and elevation over the mountainous regions [71]. This further suggests the important analysis of the RF method could be more related to the spatial pattern of SM than temporal patterns. Furthermore, spectral indices at high-resolution (1-km) were used in the RF approach to further improve RF estimates. This could be partly explained by that the RF algorithm has a strong ability to build up nonlinear relationships between remote sensing indices and SM [48].

The relationship between spectral indices and SM could not be well described using simple linear or nonlinear functions over the mountainous regions [49]. The linear/nonlinear functions used in previous works [38], [39] could result in uncertainties in the calibrated relationship between SM and optical/thermal

indices, further decreasing the accuracy of SM estimates. This could be the main reason for low slope values with a median value of 0.61 ranging from 0.17 to 1.31 by Fan *et al.* [38] over the same study region, relative to the RF estimates of the present study, with a median slope value of 0.83. The RF model presented here partly succeeded to overcome the above-mentioned limitations by using a decision tree rule to accurately describe the complex and highly nonlinear relationships between the predictors and the SM predictions [50].

However, poor performance of RF_{VI+sup} can be found at WSN-23, WSN-25, WSN-37, WSN-38, and WSN-53 sites. This poor performance of RF_{VI+sup} could be partly attributed to the uncertainties in the remote-sensed indices [36]. An accurate estimation of remote-sensed predictors is the foundation for RF prediction. To calculate these predictors, it is assumed that the study area is free of clouds during the period ranging from sunrise to the satellite overpass time. However, such cloud-free conditions are not well satisfied due to temporary cloud cover and cloud movements. The presence of clouds during the study period is one of the main factors that will increase the uncertainties in the estimation of remote-sensed predictors.

Moreover, the available observations of ATI are strongly limited by the availability of both daytime and nighttime LST observations (see the calculation of ATI in Table S2). Although spectral indices were used in the RF models (RF_{VI+sup}) to provide more SM information over the pixels with missing ATI data, the cloudy conditions still decrease the available number of spectral indices, leading to missing RF_{VI+sup} estimates. Especially, over the mountainous regions with high elevation, the missing optical indices due to the cloudy conditions resulted in a low number of RF_{VI+sup} SM estimates. Because the limited predictors (e.g., VIs or LST) could decrease the number of trained datasets, further decrease the accuracy of RF_{VI+sup}. For example, the available VIs observations ($\sim$60 days) at WSN-37 and WSN-38 during 2013–2014 were lower by $\sim$30% than other sites with good performances having more VIs observations ($\sim$90 days). In addition, the MODIS product (MCD43A4) is used to calculate the daily VIs, but note that MCD43A4 is a 16-day weighted MODIS product. The rapid variations of vegetation water content and SM at a daily scale could be not successfully captured by this product.

The missing remote sensing indices will decrease the robustness of the trained RF model, which needs massive training data so that model training can be "optimal" [72]. The train data exert a substantial control on the learning process associated with the relationship between the SM predictions and multiple indices. To obtain an accurate description of this relationship, the training data should include numerous pixels encompassing SM values associated with different surface conditions, such as different vegetation cover types, terrain slopes, and elevations [73]. Thus, a limited number of training data could increase uncertainty in the SM estimation.

Another reason for the inaccurate estimations over these specific sites could be low spatial representativeness of the point-scale *in situ* measurements. In this study, the SM spatial distribution within each 1-km grid was assumed to be homogenous. However, strong spatial heterogeneities of SM at a subkilometer-scale could result in low representativeness of the *in situ* measurements for SM conditions at 1-km grid [31]. Especially, over complex mountainous terrain, SM has much strong heterogeneities at spatiotemperal dynamics. This mismatch between the 1-km grid and point-scale field-measurements could increase uncertainties in the construction of the relationship between SM and the predictors.

Note that our proposed approach was tested over a semiarid region with low vegetation cover. Indeed, when dense vegetation covers the soil, LST and remotely-sensed spectral indices mostly reflect the status of vegetation, and thus, the ability of optical/thermal indices to capture the SM dynamics may become weaker [21]. So, uncertainties in the approach presented here could be higher in high-density vegetation regions than those considered in the present study. Also, most spectral indices, such as drought indices, are sensitive to the water stress conditions, but not to the wet (or water saturation) conditions [28], [74], [75]. Thus, the applicability of our proposed approach should be tested in wet regions. Moreover, the global-scale high-resolution SM method should be further developed by including the coarse resolution SM products. How to account for the uncertainties caused by the coarse resolution SM products will be key to global high-resolution SM.

Moreover, the SM underestimation was found in this study, mainly caused by the limited sensitivity of optical-thermal indices to SM. As noted by Fan *et al.* [38], the limited ability of optical-thermal indices to the extreme dry/wet SM conditions could potentially decrease the performance of SM estimations. It is worth to be noted that passive microwave has a high sensitivity to estimate SM, albeit with low resolution. However, with the development of downscaling passive microwave datasets, the combination of high-resolution passive microwave observations and optical indices in the proposed framework could improve the SM estimation over the mountainous regions [17].

Note that it is inevitable that the overlapped information among vegetation indices. For example, NDVI and SAVI both were calculated using the reflectance from MODIS bands 1 and 2, potential meaning that the temporal variations in NDVI could be similar to SAVI, although the importance analysis (see Fig. 2) showed the different importance between NDVI and SAVI for predicting SM. In addition, the similar importance between NDII6 and NDII7 for predicting SM could be explained by the overlapped information represented by the two indices, partly due to the close information between MODIS bands 6 and 7, which are used for the calculation of NDII6 and NDII7, respectively.

Besides the RF method, other ML methods including artificial neural networks, cubist, deep belief networks, and support vector machines have various advantages of combining SM information [76], which has the potential to improve the SM estimation. For example, an artificial neural network, as a self-adaptive and self-learning method, can describe the complex nonlinear relationship between SM and predictors [77], and the cubist method can rapidly analyze big data [78]. Thus, different ML

methods will be tested for improving the SM estimation over the mountainous regions in the future work.

VI. Conclusion

High-resolution SM methods have generally limited performance over mountainous regions because of the high spatiotemporal variations of SM in a mountainous environment. To overcome these limitations, we proposed a regional learning strategy to improve the SM retrievals, by merging optical/thermal remote sensing data and topography data. Remote sensing albedo, ATI, NDVI, NDII5, SAVI, and topographical indices (aspect and elevation) were identified as the optimal predictors for capturing temporal SM changes at pixel scale. The proposed RF model developed in this study showed a strong capability to capture the spatiotemporal patterns of SM, relative to methods only using optical data or topographical indices. The RF-derived SM estimates we computed at 1 km exhibited a high estimation accuracy with overall ubRMSD, R, and slope values of 0.032 m^3/m^3, 0.75, and 0.83, respectively. Our SM estimates have the potential to be used for agriculture applications over the Babao River Basin.

However, our approach is limited by the availability and accuracy of the used predictors (e.g., optical and thermal indices), mainly influenced by the cloudy condition. To obtain continuous spatiotemporal SM values, one possible solution is to consider active microwave data (e.g., Sentinel-1), which are available in all-weather conditions [13], [15]. Terrestrial Water Resources Satellite, as a new Chinese mission, would provide high-resolution continuous *L*-band observation, which would provide new opportunities for high-resolution SM mapping in the mountain areas [79], [80]. This mission concept for producing multiscale SM products has been demonstrated in the Soil Moisture Experiment in the Luan River [81]. Finally, the RF model developed here, as an ML method, could be improved using longer time series of *in situ* SM measurements and remote sensing indices.

References

[1] A. Al-Yaari *et al.*, "Assessment and inter-comparison of recently developed/reprocessed microwave satellite soil moisture products using ISMN ground-based measurements," *Remote Sens. Environ.*, vol. 224, pp. 289–303, Apr. 2019.

[2] L. Fan *et al.*, "Satellite-observed pantropical carbon dynamics," *Nature Plants*, vol. 5, no. 9, pp. 944–951, Jul. 2019.

[3] J. Peng *et al.*, "A roadmap for high-resolution satellite soil moisture applications-confronting product characteristics with user requirements," *Remote Sens. Environ.*, vol. 252, Jan. 2021, Art. no. 112162.

[4] S. I. Seneviratne *et al.*, "Investigating soil moisture–climate interactions in a changing climate: A review," *Earth-Sci. Rev.*, vol. 99, no. 3, pp. 125–161, May 2010.

[5] J.-P. Wigneron *et al.*, "Tropical forests did not recover from the strong 2015–2016 El Niño event," *Sci. Adv.*, vol. 6, no. 6, 2020, Art. no. eaay4603.

[6] X. Li *et al.*, "Compared performances of SMOS-IC soil moisture and vegetation optical depth retrievals based on tau-omega and two-stream microwave emission models," *Remote Sens. Environ.*, vol. 236, Jan. 2020, Art. no. 111502.

[7] T. E. 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, Nov. 2013, doi: 10.2136/sssaj2013.03.0093.

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

[9] Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, J. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1729–1735, Aug. 2001.

[10] J.-P. Wigneron *et al.*, "L-band microwave emission of the biosphere (L-MEB) model: Description and calibration against experimental data sets over crop fields," *Remote Sens. Environ.*, vol. 107, no. 4, pp. 639–655, 2007.

[11] J. P. Wigneron *et al.*, "Modelling the passive microwave signature from land surfaces: A review of recent results and application to the L-band SMOS & SMAP soil moisture retrieval algorithms," *Remote Sens. Environ.*, vol. 192, pp. 238–262, Apr. 2017.

[12] W. A. Dorigo *et al.*, "Evaluation of the ESA CCI soil moisture product using ground-based observations," *Remote Sens. Environ.*, vol. 162, pp. 380–395, Jun. 2015.

[13] 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, pp. 111–380, Nov. 2019.

[14] 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.

[15] H. Mao, D. Kathuria, N. Duffield, and B. P. Mohanty, "Gap filling of high-resolution soil moisture for SMAP/Sentinel-1: A two-layer machine learning-based framework," *Water Resour. Res.*, vol. 55, no. 8, pp. 6986–7009, Aug. 2019, doi: 10.1029/2019WR024902.

[16] B. Bauer-Marschallinger *et al.*, "Toward global soil moisture monitoring with Sentinel-1: Harnessing assets and overcoming obstacles," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 520–539, Jan. 2019.

[17] J. Peng, A. Loew, O. Merlin, and N. E. Verhoest, "A review of spatial downscaling of satellite remotely sensed soil moisture," *Rev. Geophys.*, vol. 55, no. 2, pp. 341–366, 2017.

[18] N. Chauhan, S. Miller, and P. Ardanuy, "Spaceborne soil moisture estimation at high resolution: A microwave-optical/IR synergistic approach," *Int. J. Remote Sens.*, vol. 24, no. 22, pp. 4599–4622, 2003.

[19] B. Fang, V. Lakshmi, R. Bindlish, T. J. Jackson, M. Cosh, and J. Basara, "Passive microwave soil moisture downscaling using vegetation index and skin surface temperature," *Vadose Zone J.*, vol. 12, no. 3, pp. 1–19, 2013.

[20] W. Zhao and A. Li, "A comparison study on empirical microwave soil moisture downscaling methods based on the integration of microwave-optical/IR data on the Tibetan Plateau," *Int. J. Remote Sens.*, vol. 36, no. 19/20, pp. 4986–5002, 2015.

[21] O. Merlin, J. P. Walker, A. Chehbouni, and Y. Kerr, "Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3935–3946, 2008.

[22] O. Merlin, C. Rudiger, A. Al Bitar, P. Richaume, J. P. Walker, and Y. H. Kerr, "Disaggregation of SMOS soil moisture in Southeastern Australia," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1556–1571, May 2012.

[23] J. Wang, Z. Ling, Y. Wang, and H. Zeng, "Improving spatial representation of soil moisture by integration of microwave observations and the temperature–vegetation–drought index derived from MODIS products," *ISPRS J. Photogramm. Remote Sens.*, vol. 113, pp. 144–154, 2016.

[24] J. Kim and T. S. Hogue, "Improving spatial soil moisture representation through integration of AMSR-E and MODIS products," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 446–460, Feb. 2011.

[25] L. Fan *et al.*, "Evaluation of the airborne CASI/TASI Ts-VI space method for estimating near-surface soil moisture," *Remote Sens.*, vol. 7, no. 3, pp. 3114–3137, 2015.

[26] J. Peng, A. Loew, X. Chen, Y. Ma, and Z. Su, "Comparison of satellite-based evapotranspiration estimates over the Tibetan Plateau," *Hydrol. Earth Syst. Sci.*, vol. 20, pp. 3167–3182, 2016.

[27] J. Peng, J. Niesel, and A. Loew, "Evaluation of soil moisture downscaling using a simple thermal-based proxy-the REMEDHUS network (Spain) example," *Hydrol. Earth Syst. Sci.*, vol. 19, pp. 4765–4782, 2015.

[28] Y. Malbéteau *et al.*, "Normalizing land surface temperature data for elevation and illumination effects in mountainous areas: A case study using ASTER data over a steep-sided valley in Morocco," *Remote Sens. Environ.*, vol. 189, pp. 25–39, 2017.

[29] W. Zhao, A. Li, H. Jin, Z. Zhang, J. Bian, and G. Yin, "Performance evaluation of the triangle-based empirical soil moisture relationship models based on Landsat-5 TM data and in situ measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2632–2645, May 2017.

[30] 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, 2012.

[31] R. Jin, X. Li, and S. Liu, "Understanding the heterogeneity of soil moisture and evapotranspiration using multiscale observations from satellites, airborne sensors, and a ground-based observation matrix," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2132–2136, Nov. 2017.

[32] Y. Ge, J. Wang, G. B. Heuvelink, R. Jin, X. Li, and J. Wang, "Sampling design optimization of a wireless sensor network for monitoring ecohydrological processes in the Babao River Basin, China," *Int. J. Geographical Inf. Sci.*, vol. 29, no. 1, pp. 92–110, 2015.

[33] J. Kang, R. Jin, and X. Li, "Regression kriging-based upscaling of soil moisture measurements from a wireless sensor network and multiresource remote sensing information over heterogeneous cropland," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 92–96, Jan. 2015.

[34] J. Wang, Y. Ge, Y. Song, and X. Li, "A geostatistical approach to upscale soil moisture with unequal precision observations," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2125–2129, Dec. 2014.

[35] S. Gao, Z. Zhu, S. Liu, R. Jin, G. Yang, and L. Tan, "Estimating the spatial distribution of soil moisture based on Bayesian maximum entropy method with auxiliary data from remote sensing," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 32, pp. 54–66, 2014.

[36] L. Fan *et al.*, "Mapping high-resolution soil moisture over heterogeneous cropland using multi-resource remote sensing and ground observations," *Remote Sens.*, vol. 7, no. 10, pp. 13273–13297, 2015.

[37] J. Qin, K. Yang, N. Lu, Y. Chen, L. Zhao, and M. Han, "Spatial upscaling of in-situ soil moisture measurements based on MODIS-derived apparent thermal inertia," *Remote Sens. Environ.*, vol. 138, pp. 1–9, 2013.

[38] L. Fan *et al.*, "Mapping soil moisture at a high resolution over mountainous regions by integrating in situ measurements, topography data, and MODIS land surface temperatures," *Remote Sens.*, vol. 11, no. 6, 2019, Art. no. 656.

[39] J. Kang, R. Jin, X. Li, C. Ma, J. Qin, and Y. Zhang, "High spatio-temporal resolution mapping of soil moisture by integrating wireless sensor network observations and MODIS apparent thermal inertia in the Babao River Basin, China," *Remote Sens. Environ.*, vol. 191, pp. 232–245, 2017.

[40] S. Ahmad, A. Kalra, and H. Stephen, "Estimating soil moisture using remote sensing data: A machine learning approach," *Adv. Water Resour.*, vol. 33, no. 1, pp. 69–80, Jan. 2010.

[41] J. Im, S. Park, J. Rhee, J. Baik, and M. Choi, "Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches," *Environ. Earth Sci.*, vol. 75, no. 15, 2016, Art. no. 1120.

[42] H. Jiang, H. Shen, H. Li, F. Lei, W. Gan, and L. Zhang, "Evaluation of multiple downscaled microwave soil moisture products over the central Tibetan Plateau," *Remote Sens.*, vol. 9, no. 5, 2017, Art. no. 402.

[43] P. Song, J. Huang, and L. R. Mansaray, "An improved surface soil moisture downscaling approach over cloudy areas based on geographically weighted regression," *Agricultural Forest Meteorol.*, vol. 275, pp. 146–158, Sep. 2019.

[44] Z. Wei, Y. Meng, W. Zhang, J. Peng, and L. Meng, "Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau," *Remote Sens. Environ.*, vol. 225, pp. 30–44, May 2019.

[45] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, "Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data," *Remote Sens.*, vol. 7, no. 12, pp. 16398–421, 2015.

[46] X. He, N. W. Chaney, M. Schleiss, and J. Sheffield, "Spatial downscaling of precipitation using adaptable random forests," *Water Resour. Res.*, vol. 52, no. 10, pp. 8217–8237, Oct. 2016, doi: 10.1002/2016WR019034.

[47] C. Pelletier, S. Valero, J. Inglada, N. Champion, and G. Dedieu, "Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas," *Remote Sens. Environ.*, vol. 187, pp. 156–168, 2016.

[48] T. Xu *et al.*, "Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale," *J. Geophys. Res., Atmos.*, vol. 123, no. 16, pp. 8674–8690, Aug. 2018, doi: 10.1029/2018JD028447.

[49] W. Zhao, N. Sánchez, H. Lu, and A. Li, "A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression," *J. Hydrol.*, vol. 563, pp. 1009–1024, 2018.

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

[51] X. Li *et al.*, "Watershed allied telemetry experimental research," *J. Geophys. Res., Atmos.*, vol. 114, no. D22, pp. 1–19, 2009.

[52] X. Li *et al.*, "Heihe watershed allied telemetry experimental research (HiWATER): Scientific objectives and experimental design," *Bull. Amer. Meteorol. Soc.*, vol. 94, no. 8, pp. 1145–1160, 2013.

[53] Z. Li *et al.*, "Environmental significance and hydrochemical processes at a cold alpine basin in the Qilian mountains," *Environ. Earth Sci.*, vol. 73, no. 8, pp. 4043–4052, 2015.

[54] G. Yong, W. Jianghao, W. Jinfeng, J. Rui, and H. Maogui, "Regression kriging model-based sampling optimization design for the eco-hydrology wireless sensor network," *Adv. Earth Sci.*, vol. 27, no. 9, pp. 1006–1013, 2012.

[55] Y. Ge, J. H. Wang, G. B. M. Heuvelink, R. Jin, X. Li, and J. F. Wang, "Sampling design optimization of a wireless sensor network for monitoring ecohydrological processes in the Babao River Basin, China," *Int. J. Geographical Inf. Sci.*, vol. 29, no. 1, pp. 92–110, Jan. 2015.

[56] T. Carlson, "An overview of the "triangle method" for estimating surface evapotranspiration and soil moisture from satellite imagery," *Sensors*, vol. 7, no. 8, pp. 1612–1629, 2007.

[57] C. Schaaf and Z. Wang, "MCD43A4 MODIS/Terra+ aqua BRDF/Albedo nadir BRDF adjusted ref daily L3 global-500m V006," NASA EOSDIS Land Process. DAAC, 2015.

[58] J. Qin *et al.*, "Inter-comparison of spatial upscaling methods for evaluation of satellite-based soil moisture," *J. Hydrol.*, vol. 523, pp. 170–178, 2015.

[59] Z. Wan, Y. Zhang, Q. Zhang, and Z.-L. Li, "Quality assessment and validation of the MODIS global land surface temperature," *Int. J. Remote Sens.*, vol. 25, no. 1, pp. 261–274, 2004.

[60] K. J. Beven and M. J. Kirkby, "A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant," *Hydrological Sci. J.*, vol. 24, no. 1, pp. 43–69, 1979.

[61] S. Lei, H. Chen, Z. Bian, and Z. Liu, "Evaluation of integrating topographic wetness index with backscattering coefficient of TerraSAR-X image for soil moisture estimation in a mountainous region," *Ecological Indicators*, vol. 61, pp. 624–633, 2016.

[62] F. Schmidt and A. Persson, "Comparison of DEM data capture and topographic wetness indices," *Precis. Agriculture*, vol. 4, no. 2, pp. 179–192, 2003.

[63] A. W. Western, R. B. Grayson, G. Blöschl, G. R. Willgoose, and T. A. McMahon, "Observed spatial organization of soil moisture and its relation to terrain indices," *Water Resour. Res.*, vol. 35, no. 3, pp. 797–810, 1999.

[64] S. Das, R. Chakraborty, and A. Maitra, "A random forest algorithm for nowcasting of intense precipitation events," *Adv. Space Res.*, vol. 60, pp. 1271–1282, 2017.

[65] S. Oliveira, F. Oehler, J. San-Miguel-Ayanz, A. Camia, and J. M. C. Pereira, "Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest," *Forest Ecol. Manage.*, vol. 275, pp. 117–129, Jul. 2012.

[66] K. Rao, W. R. L. Anderegg, A. Sala, J. Martínez-Vilalta, and A. G. Konings, "Satellite-based vegetation optical depth as an indicator of drought-driven tree mortality," *Remote Sens. Environ.*, vol. 227, pp. 125–136, Jun. 2019.

[67] A. Al-Yaari *et al.*, "Testing regression equations to derive long-term global soil moisture datasets from passive microwave observations," *Remote Sens. Environ.*, vol. 180, pp. 453–464, Jul. 2016.

[68] F. A. Busch, J. D. Niemann, and M. Coleman, "Evaluation of an empirical orthogonal function–based method to downscale soil moisture patterns based on topographical attributes," *Hydrological Process.*, vol. 26, no. 18, pp. 2696–2709, Aug. 2012, doi: 10.1002/hyp.8363.

[69] G. S. Cowley, J. D. Niemann, T. R. Green, M. S. Seyfried, A. S. Jones, and P. J. Grazaitis, "Impacts of precipitation and potential evapotranspiration patterns on downscaling soil moisture in regions with large topographic relief," *Water Resour. Res.*, vol. 53, no. 2, pp. 1553–1574, Feb. 2017, doi: 10.1002/2016WR019907.

[70] P. Nasta, D. Penna, L. Brocca, G. Zuecco, and N. Romano, "Downscaling near-surface soil moisture from field to plot scale: A comparative analysis under different environmental conditions," *J. Hydrol.*, vol. 557, pp. 97–108, Feb. 2018.

[71] C. Song, L. Jia, and M. Menenti, "Retrieving high-resolution surface soil moisture by downscaling AMSR-E brightness temperature using MODIS LST and NDVI data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 3, pp. 935–942, Mar. 2013.

[72] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[73] W. Zhao, H. Wu, G. Yin, and S.-B. Duan, "Normalization of the temporal effect on the MODIS land surface temperature product using random forest regression," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 109–118, Jun. 2019.

[74] N. Djamai, R. Magagi, K. Goïta, O. Merlin, Y. Kerr, and A. Roy, "A combination of DISPATCH downscaling algorithm with CLASS land surface scheme for soil moisture estimation at fine scale during cloudy days," *Remote Sens. Environ.*, vol. 184, pp. 1–14, 2016.
[75] L. Fan *et al.*, "Evaluation of microwave remote sensing for monitoring live fuel moisture content in the Mediterranean region," *Remote Sens. Environ.*, vol. 205, pp. 210–223, 2018.
[76] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015.
[77] M. Kumar, N. S. Raghuwanshi, R. Singh, W. W. Wallender, and W. O. Pruitt, "Estimating evapotranspiration using artificial neural network," *J. Irrigation Drainage Eng.*, vol. 128, no. 4, pp. 224–233, Aug. 2002.
[78] S. Chen, D. She, L. Zhang, M. Guo, and X. Liu, "Spatial downscaling methods of soil moisture based on multisource remote sensing data and its application," *Water*, vol. 11, no. 7, 2019, Art. no. 1401.
[79] T. Zhao *et al.*, "Soil moisture retrievals using L-band radiometry from variable angular ground-based and airborne observations," *Remote Sens. Environ.*, vol. 248, no. 20, 2020, Art. no. 111958.
[80] P. Guo, T. Zhao, J. Shi, H. Xu, X. Li, and S. Niu, "Assessing the active-passive approach at variant incidence angles for microwave brightness temperature downscaling," *Int. J. Digit. Earth*, vol. 14, no. 10, pp. 1273–1293, 2021.
[81] T. Zhao *et al.*, "Soil moisture experiment in the Luan river supporting new satellite mission opportunities," *Remote Sens. Environ.*, vol. 240, no. 20, 2020, Art. no. 111680.

Lei Fan received the Ph.D. degree in geographic information system from the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, in 2017.

He was a joint Ph.D. and Postdoc with Interactions Sol Plante Atmosphere, Institut National de Recherche Agronomiques, Bordeaux, France, from 2015 to 2016 and 2017 to 2019, respectively. He is currently a Professor with the School of Geographical Sciences, Southwest University, Chongqing, China. His research interests include the retrieval of soil moisture and vegetation optical depth, and how the carbon cycle responds to variations in water availability at large scales, surrounding the role of vegetation water content in predicting plant carbon fluxes and growth.

Amen Al-Yaari received the B.S. (Hons.) degree in geophysics (with excellent grade and first rank in the class) from Damascus University, Damascus, Syria, in 2008, the M.Sc. degree in water resource engineering from Lund University, Lund, Sweden, in 2011, and the Ph.D. (Hons.) degree in microwave remote sensing and hydrology from Sorbonne University, Paris, France, in 2015.

From 2015 to 2018, he was with the INRAe (Institut national de recherche pour l'agriculture, l'alimentation et l'environnement) as a Research Scientist, Bordeaux, France. In 2019, he joined Sorbonne University as an Assistant Professor. He spent one month at Middle East Technical University, Turkey, as a visiting scholar in 2019. His involvement in Soil Moisture and Ocean Salinity mission started in 2012, with the validation of soil moisture data and development of the scientific algorithms and statistical methods to construct long and homogeneous time series of soil moisture. He has authored or coauthored more than 50 papers in international peer-reviewed journals. His research interests include the theory and techniques for microwave remote sensing of the Earth, with emphasis on hydrology, water resources management, and vegetation monitoring.

Dr. Al-Yaari has been a member of the Editorial Board of *Geoscience* since 2019.

Frédéric Frappart received the Engineer degree in oceanography engineer from Ecole Nationale Supérieure des Techniques Avancées Bretagne (ENSTA Bretagne, formerly ENSIETA), Brest, France, in 2001, and the Ph.D. degree in geophysics and remote sensing from the Université de Toulouse, Toulouse, France, in 2006.

He has been a Researcher with the Observatoire Midi-Pyrénées (OMP), Toulouse, since 2010, in charge of the scientific applications of radar altimetry over land (hydrology and surface properties) for the Centre de Topographie des Ocans et de l'Hydrosphère, a French Observation Service dedicated to radar altimetry studies, and a member of the Scientific Definition Team of the NASA/CNES InSAR altimeter Surface Water and Ocean Topography (SWOT) mission for land hydrology from 2012 to 2015 and of the SWOT Science Team from 2016 to 2020. He is involved in GNSS-R activities in the Geodesy from space team at GET-OMP.

Jian Peng received the Ph.D. degree in earth science from the Max Planck Institute for Meteorology (MPI-M), Hamburg, Germany, in 2003.

Before joining the University of Oxford, he was a Research Scientist with the University of Munich (LMU) and a Postdoc Researcher with MPI-M. He is currently an Earth System Scientist and the Head of the Department of Remote Sensing, UFZ, Leipzig. He is also a Full Professor in hydrology and remote sensing with the University of Leipzig, Leipzig. His research interests include the quantitative retrieval of land surface parameters from remote sensing data, the assimilation of remote sensing data into climate and land surface process models, understanding land–atmosphere interactions using Earth system models and observational data, and quantification of climate change impact on water resources, in particular, focusing on estimation of high-resolution land surface water and energy fluxes from satellite observations, and the investigation of hydrological and climatic extremes as well as their impacts on ecosystems.

Dr. Peng is the Editor-in-Chief of *Geoscience Data Journal*. He was a recipient of numerous international awards, most recently in 2019, the Remote Sensing Young Investigator Award of the Swiss scientific publisher MDPI.

Jianguang Wen received the B.S. degree in geographic information system and the M.S. degree in geophysical prospecting and information technology from Jilin University, Changchun, China, in 2002 and 2005, respectively, and the Ph.D. degree in geographic information system from the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, in 2008.

He was a Research Assistant from 2008 to 2011 and a Research Associate from 2011 to 2018 and since 2018 has been a Research Fellow with Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. He is the author of one book, more than 70 articles, and seven inventions. His research interests include multiangle remote sensing mechanism, remote sensing modeling and parameter inversion, and remote sensing experiment.

Dr. Wen was the recipient of the First Class of Progress of Surveying and Mapping Technology Prize in 2018.

Qing Xiao received the B.S. degree in applied geophysics from Jilin University, Changchun, China, in 1993, and the M.S. and Ph.D. degrees in remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, China, in 1996 and 2002, respectively.

He is currently a Researcher with the Institute of Remote Sensing and Digital Earth, CAS. His main research interests include remote sensing radiative transfer mechanism and inversion, and remote sensing experiment in land surface parameters validation.

Rui Jin photograph and biography not available at the time of publication.

Xiaojun Li photograph and biography not available at the time of publication.

Xiangzhuo Liu photograph and biography not available at the time of publication.

Mengjia Wang photograph and biography not available at the time of publication.

Xiuzhi Chen photograph and biography not available at the time of publication.

Lin Zhao received the B.S. degree in hydrogeology and engineering geology from Hebei GEO University, Shijiazhuang, China, in 1988, the M.S. degree in physical geography from the Lanzhou Institute of Glaciology and Geocryology, Chinese Academy of Sciences, Lanzhou, China, in 1991, and the Ph.D. degree in physical geography from Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, in 2003.

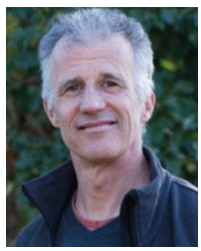
He was a Station Master of the Cryosphere Research Station on the Qinghai-Tibet Plateau, Chinese Academy of Sciences, Golmud, China, from 2003 to 2020. He is currently a Professor with the School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing, China. He is the author of two books and more than 100 articles. His research interests include land surface process simulation and remote sensing, geographic information system, meteorological application, and ecological environment remote sensing.

Dr. Zhao was the recipient of the Third Class of Progress of Science and Technology Prize of Gansu Province and was awarded as a national outstanding field scientific and technical worker.

Mingguo Ma received the B.S. degree in rural and urban planning from Resources and Environment College, Lanzhou University, Lanzhou, China, in 1998, and the Ph.D. degree in cartography from the University of Chinese Academy of Sciences, Beijing, China, in 2003.

He was a Research Assistant in 2003, a Research Associate from 2003 to 2009, and a Research Fellow from 2009 to 2014 with Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou. He is currently a Professor with the Chongqing Engineering Research Center for Remote Sensing Big Data Application, School of Geographical Sciences, Southwest University, Chongqing, China. He has authored or coauthored more than 200 papers. His research interests include remote sensing of ecological and land surface processes, and processing and application of remote sensing time-series data.

Dr. Ma has been a member of the Editorial Board of *Remote Sensing Technology and Application*. He is an Associate Editor for *Frontiers of Earth Science*.



Jean-Pierre Wigneron (Fellow, IEEE) received the M.Sc./Engineering degree in engineering from SupAéro, Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, France, in 1987, and the Ph.D. degree in remote sensing from the University of Toulouse, Toulouse, in 1993.

He is currently a Senior Research Scientist with the Institut National de Recherche Agronomiques (INRA), Bordeaux, France, a Co-Coordinator of remote sensing activities with INRA and within the Aquitaine Observatoire Aquitain des Sciences de l'Univers network, a Deputy Director of Ephyse, and the Head of the remote sensing team. He coordinated the development of the *L*-band microwave emission of the biosphere model for soil and vegetation in the Level-2 inversion algorithm of the European Space Agency-Soil Moisture and Ocean Salinity mission. He has authored or coauthored more than 100 papers in international peer-reviewed journals. His research interests include microwave remote sensing of soil and vegetation, radiative transfer, and data assimilation.

Dr. Wigneron has been an Associate Editor for *Remote Sensing of Environment* since 2014.