Soil Moisture Content Estimation From Hyperspectral Remote Sensing Data

Ketaki Vinay Jambhali, *Graduate Student Member, IEEE*, Bikram Koirala, *Member, IEEE*, Zakaria Bnoulkacem, *Graduate Student Member, IEEE*, and Paul Scheunders, *Senior Member, IEEE*

***Abstract*—Because of its significant absorption power, especially in the shortwave infrared optical region, water dominates the optical reflectance properties of water-bearing materials. This allows us to study a material's water-related features, such as its moisture content, from optical reflectance. In this study, we proposed a framework to estimate soil moisture content from PRISMA hyperspectral remote sensing data. The proposed framework requires a dry endmember spectrum and an endmember spectrum of high soil moisture along with ground truth moisture content, obtained from ground measurements. The method takes into account the complex interaction of light with soils, the large variation of environmental conditions, leading to spectral variability and the soil-specific behavior of water. The framework is extensively validated using ground-measured soil moisture data from the International Soil Moisture Network database. A total of 1418 PRISMA images corresponding to 151 ground stations were analyzed. From 518 retained images, a total root-mean-squared error of 8.682 % and R^2 of 0.385 was obtained.**

***Index Terms*—Hyperspectral, machine learning regression, PRISMA, remote sensing, soil moisture content (SMC).**

I. Introduction

SOIL moisture is an essential variable for the study of the regional water cycle, agricultural irrigation management, climate change, and environmental monitoring [1]. As the main source of water for agriculture and natural vegetation, soil water plays a key role in crop production by serving as a solvent for nutrients, such as sodium, potassium, carbon, and nitrogen [2]. Soil moisture has a substantial impact on plants and soil biota, which also affect the global nutrient cycles and ecosystem services, such as crop productivity [3]. It also plays a very important role in the exchange of mass and energy between the Earth's surface and the atmosphere [4], [5], [6], [7], [8], [9]. In all these domains, timely and accurate assessment of soil moisture content (SMC) is critical [10], [11], [12].

Received 15 July 2024; revised 4 March 2025; accepted 22 August 2025. Date of publication 29 August 2025; date of current version 15 September 2025. This work was supported by the Research Foundation-Flanders - Project under Grant G031921N. The work of Ketaki Vinay Jambhali was supported by the Research Fund of the University of Antwerp. The work of Bikram Koirala is a Postdoctoral Fellow of the Research Foundation Flanders, Belgium under Grant FWO:1250824N-7028. The work of Zakaria Bnoulkacem is a Doctoral Fellow of the Research Foundation Flanders, Belgium under Grant FWO:1SH1M24N. *(Corresponding author: Ketaki Vinay Jambhali.)*

The authors are with the Imec-Visionlab, Department of Physics, University of Antwerp (CDE), 2000 Antwerpen, Belgium (e-mail: ketakivinay.jambhali@uantwerpen.be; bikram.koirala@uantwerpen.be; zakaria.bnoulkacem@uantwerpen.be; paul.scheunders@uantwerpen.be).

Digital Object Identifier 10.1109/JSTARS.2025.3603841

The term "soil moisture content" refers to the amount of water present on land surfaces in the pores of the soil. Multiple factors, including soil type and nearby vegetation, as well as meteorological conditions, influence the amount of soil moisture [13]. In turn, a variety of soil and plant dynamics are influenced by SMC [14].

Traditional methods, such as the thermo-gravimetric technique and the calcium-carbide approach, require both field sampling and laboratory analysis [15]. Although these approaches have high accuracy, they also have disadvantages. They are destructive, require complicated sampling procedures, and a large number of repeated experiments [16]. Commonly used ground-based techniques (time-domain reflectometry, frequency domain sensors, etc.) estimate volumetric SMC [17], [18], [19]. Although these methods can be adapted to monitor temporal variations of volumetric SMC locally (in a small area), they cannot be applied to map the spatial distribution of soil moisture on a large scale. Compared to these traditional methods, remote sensing technologies make it possible to monitor soil moisture near the surface on a large scale.

Microwave remote sensing allows continuous, large-scale estimation of soil moisture, because moisture strongly affects soil dielectric properties and longer wavelengths penetrate relatively deeply into the ground [7], [8], [12]. Moreover, surface vegetation has no influence on the longer wavelength microwave radiation (see [5]). Large-scale soil moisture data have been collected by the Soil Moisture Active Passive and Soil Moisture and Ocean Salinity satellite missions [19], from which soil moisture can be estimated on a global scale [20]. Numerous algorithms have been developed to estimate SMC from microwave radiation. However, the resulting data products often exhibit limited spatial resolution (10–20 km) [21], even when collected from airborne systems (see [17]), making them less suitable for monitoring, e.g., small catchment areas (see [19]). Moreover, information is needed about the surface roughness and the dielectric constant of the soil.

Thermal infrared remote sensing uses wavelengths ranging from approximately 3500 to 14 000 nm to estimate SMC. The use of thermal remote sensing for estimating SMC has been limited in the past due to high acquisition costs. Currently, thermal images with high spatial and temporal resolution are affordable due to advances in low-cost remote sensing platforms, such as autonomous aerial vehicles, which have contributed to understanding the variability of soil conditions [22]. Often, the methods using thermal infrared remote sensing images are

empirical in nature and heavily depend on local meteorological factors, such as wind speed, air temperature, and humidity [23].

Water has a high absorption capacity in the shortwave infrared (SWIR) (900–2500 nm) wavelength region. To exploit this property of water, the reflectance in the 350–2500 nm range has been used in optical remote sensing to estimate SMC [9], [24]. Despite the limited penetration depth (a few micrometers to a few millimeters) of the incident light in the SWIR wavelength region, the reflected light acquired by a hyperspectral camera can be utilized to estimate the SMC of the topmost layer of the soil at a high spatial resolution [23]. Unlike active microwave sensors, the main advantage of using the optical part of electromagnetic spectrum is that solar radiation acts as a natural illumination source. The optical remote sensing techniques also have certain disadvantages: weather conditions affect the images they produce and vegetation on soil surfaces affects the accuracy at which SMC is estimated [19].

In earlier work, the impact of soil moisture on reflectance was mainly studied qualitatively. A decrease in reflectance when artificially moistening soil, due to internal reflections of the reflectance in the water film covering the soil particles, was experimentally observed [25], [26], [27]. These observations were made on unsaturated soils [8], [28]; specular reflection can cause the reflectance of an oversaturated soil to increase.

Since then, different methods were developed to estimate soil SMC from spectral reflectance [12], which can be grouped into empirical methods and methods based on physical modeling.

Empirical methods make use of spectral indices [12], [29], [30], [31], statistical methods [32], [33], exponential functions [8], [9], [34], wavelet analysis [35], and multivariate analysis [36]. Most of these methods use a limited number of wavelength bands to estimate SMC. It was concluded that SWIR bands are better suited for estimating SMC, due to the strong absorption bands of molecular water around 1400 and 1900 nm [7], [9].

Physical models describe the interaction of light with moist soils. Most of the physical modeling techniques use the entire VNIR and SWIR wavelength region (400–2500 nm) [37]. The Bach model [38] and the MARMIT model [10] describe the reflectance of a wet soil by using the soil reflectance of a dry soil sample and a parameter L that denotes the active thickness of the water layer. Inversion leads to an estimate of L. The relationship between L and the SMC is then learned with a supervised regression model using a training dataset with known SMC. In [23], the Sadeghi model (SM) was derived from the Kubelka–Munk theory. This model estimates SMC of moist soil by using the spectral reflectance of a dry and saturated soil sample.

These methods were mainly validated with data obtained in a controlled laboratory environment. In field conditions, the use of the entire wavelength range is required because the strong absorption bands of molecular water around 1400 and 1900 nm are masked by atmospheric absorption. Moreover, environmental conditions may vary a lot, and a large variety of soil types is encountered. The existing methods are not invariant to these variations. To address these challenges, in recent work [39], we proposed a methodology (NRAL) that is invariant to the changes in the acquisition and illumination conditions and soil type. NRAL has been validated extensively on laboratory data. Similar to SM, NRAL relies on the spectral reflectance of a dry and maximally wet soil sample.

In this work, we will explore the potential of hyperspectral remote sensing imaging for SMC estimation. Because the developed NRAL approach entirely relies on the light-absorption properties of water in the SWIR range, it is critical to have access to a sufficient number of distinct wavelength values in the range 1100–1900 nm. This limits the possible choices of remote sensing sensors. In preliminary work [40], we have used Hyperion EO 1 hyperspectral images to estimate SMC. Hyperion data have a number of disadvantages: they are provided as level 1R data, requiring atmospheric correction to transform to surface-level reflectance, and they contain a lot of cloud cover. For this reason, in this work, PRISMA hyperspectral data are applied. PRISMA images are of better quality than Hyperion EO 1, with almost no cloud cover.

Validation is performed through a large number of SMC measurements by ground stations from the International Soil Moisture Network (ISMN) [41] and spectral reflectance data were obtained from a large amount of spatially and temporally aligned PRISMA hyperspectral images. Preprocessing involves the masking of vegetation pixels. The SMC estimation is done based on the NRAL method. The rest of this article is organized as follows. In Section II, we elaborate on the preparation of the PRISMA hyperspectral data and the ground truth SMC data. In Section III, the methodology for the estimation of the SMC is described. In Section IV, the experiments and results are elaborated. Finally, Section VI concludes this article.

II. Data Acquisition and Preparation

A. Soil Moisture Ground Stations

Soil moisture data were downloaded from the ISMN [41]. The ISMN is a collaborative effort to create and maintain a global in-situ soil moisture database. This database is critical for evaluating and improving global satellite products, as well as land surface, climate, and hydrological models. Multiple national organizations globally upload their in-situ soil moisture measurements on the ISMN portal for further research purposes. It is therefore the most global source of ground-observed soil moisture data that can be applied for our study. To date, ISMN groups 77 networks, each with a different number of ground stations in different locations. The number of stations per network varies from 1 to 460. Datasets from participating networks are harmonized in both time [measurements translated to coordinated universal time (UTC)] and measurement units [stored in the database as fractional volumetric soil moisture (m^3/m^3)] [41].

Most stations use different types of HydraProbe sensors to measure SMC. The HydraProbe sensor measures the soil's electrical properties to determine its water content. Because water has a higher dielectric constant than dry soil, the amount of water in the soil can be deduced from this. The ISMN collects in-situ soil moisture data using various data acquisition procedures, such as sensor installation depths and installation positions (vertical and horizontal), to monitor soil moisture at a given

depth or over a depth interval. Depending on the monitoring station, depths range from the upper soil layer (0–0.1 m) up to 2 m. Because hyperspectral remote sensing images are applied, only measurements from the upper soil layer will be relevant, and we only retain stations that provide SMC data from the top 5 cm soil layer.

The frequency of soil moisture measurements varies from network to network. Based on the frequency of data uploads, the networks can be classified as near real-time (almost daily), regular (once a month), or irregular. Because the PRISMA mission started in March 2019, we kept soil moisture measurements from June 2019 to December 2023. To simplify the process of finding corresponding hyperspectral images, we kept only the stations that provide hourly SMC information, regardless of whether they are real-time (daily for the entire period), regular (including stations providing hourly data for one day per month) or irregular updates. A total of 1077 stations from 37 networks were retained.

B. PRISMA Hyperspectral Images

The ISMN measurements serve as reference SMC data to validate the proposed approach for estimating SMC from hyperspectral remote sensing data. Real remote sensing spectral reflectance data are obtained from the PRISMA hyperspectral sensor. Launched on 22 March 2019, PRISMA is a medium-resolution hyperspectral imaging satellite developed, owned, and operated by Agenzia Spaziale Italiana. The sensor has a spatial resolution of 30 m. It is a pushbroom scanner with a swath width of 30 km and a field of view of 1000 km on either side. PRISMA records 239 distinct spectral channels with a spectral interval from 400 to 2500 nm. It operates in two blocks: 66 bands in the VNIR and 173 bands in the SWIR.

In this study, L2D products were utilized. For this purpose, geometric correction with ground control points and a digital terrain model was added to the processing stages, resulting in surface reflectance projected on a cartographic UTM system.

To correlate PRISMA images with the SMC measurements from the reference ground station, the required PRISMA images need to be filtered based on the location and date of SMC measurements by the ground stations of the selected 37 networks. Moreover, images should be free of cloud cover. A QGIS database was developed by matching the location of ground stations with the polygon vector layer of all PRISMA images available from the mission's start until December 2023. This layer also provides additional information about the images, such as the date of acquisition, cloud cover, and coordinates of the image's four corners.

The images are first filtered by location and date. Ground stations were linked to PRISMA images based on proximity. All images that overlap with a ground station were selected. It was then verified for each image whether a valid ground-station soil moisture measurement was available on the same day and hour as the image acquisition date. If found, the ground station SMC was selected at the same hour the satellite passed overhead.

Then, cloud cover was checked and only images containing less than 20% cloud cover were retained (more than 90% of the PRISMA images have a cloud cover less than 20%).

A final filter is based on the number of images available for each station. As will become clear from the next section, in the proposed approach for estimating SMC from hyperspectral data, moist soil is described as a binary mixture of an air-dried and a maximally wet soil sample, with known spectral reflectance (endmembers) and SMC. Because the endmember SMC is obtained from the ground stations, only stations with three or more matching PRISMA images lead to usable datapoints.

After filtering, we identified 151 ground stations that met these requirements. At each station, the number of matched images ranged from 3 to 39. The spatial distribution of the identified ground stations is shown in Fig. 1. Fig. 2 shows the networks used and the number of ground stations they contain. The three largest networks, SCAN, SNOTEL, and USCRN from which we select our stations, all update in real time and provide hourly updates of SMC.

The final reference database is created, which contains the image name, date, and hour of acquisition, cloud cover, station, network, and ground station soil moisture for that specific hour. Then, the PRISMA images were downloaded. A total of 1418 hyperspectral images with overlapping ground station and SMC measurements were downloaded. Fig. 3 illustrates the proposed acquisition methodology.

The VNIR and SWIR PRISMA L2D data cubes are stored as (height, 66, width) and (height, 173, width), respectively, where the image height and width varied between images.

C. Data Pre-Processing

The SWIR data cube has 173 bands, but the first two are bad bands and have been removed, leaving a total of 171 bands. The SWIR data cube was also accompanied by a SWIR error matrix, showing errors in pixel radiance processing. Bad pixels, showing an error value different from zero, were masked.

Not all acquired images consist entirely of bare ground. In particular, some images contain a significant amount of vegetation, which can largely influence the estimated water content in an image. To eliminate this effect, vegetation pixels must be masked. To detect vegetation pixels, the Normalized Difference Vegetation Index (NDVI) was applied. NDVI determines the difference between a near-infrared and a red band, which is high for vegetation. Typically, bands around 660 and 800 nm are selected, and a threshold of 0.9 for the NDVI value is selected to determine vegetation pixels. The NDVI was applied to all PRISMA VNIR data cubes and selected vegetation pixels were masked.

Since the PRISMA satellite has a swath of 30 km, and the images typically have a height and width of 1000 pixels each or larger, the images cover a large area, and only the region near the ground station can be considered. Furthermore, to make the estimation more robust, we choose to average the pixel spectra over an area near the ground station. On all preprocessed images, we use PRISMA's geolocation data (latitude and longitude) to determine the distance between each pixel and the actual ground station. Then, all spectra within a certain radius from the station are averaged. In the experimental section, we describe the procedure to determine the optimal radius over which the spectra should be averaged. After averaging, one

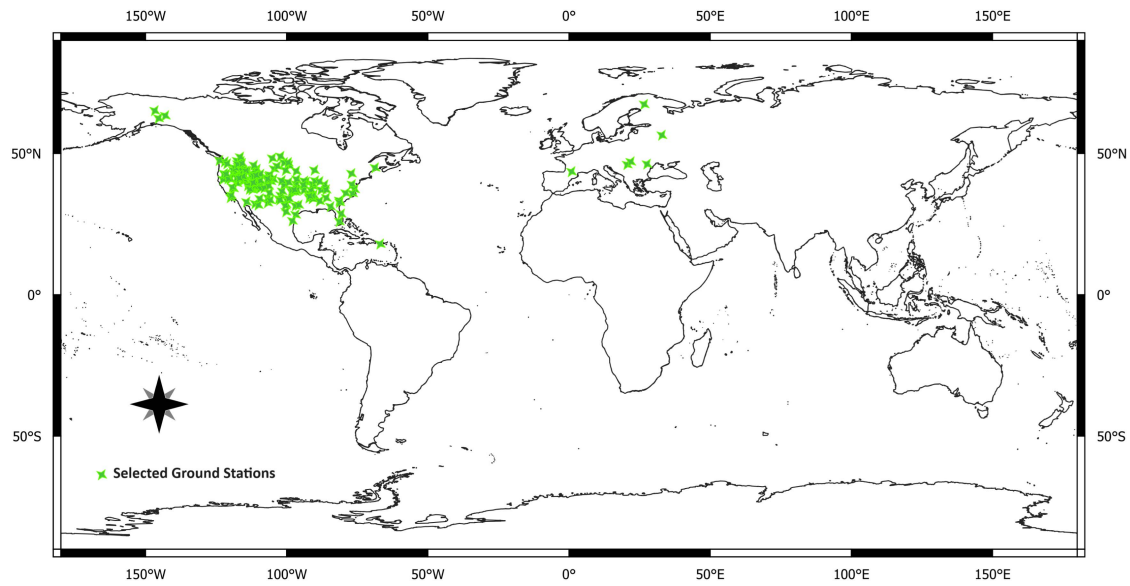

Fig. 1. Location of ground stations selected from the ISMN (WGS 84 EPSG : 4326). Shapefile obtained from [42].

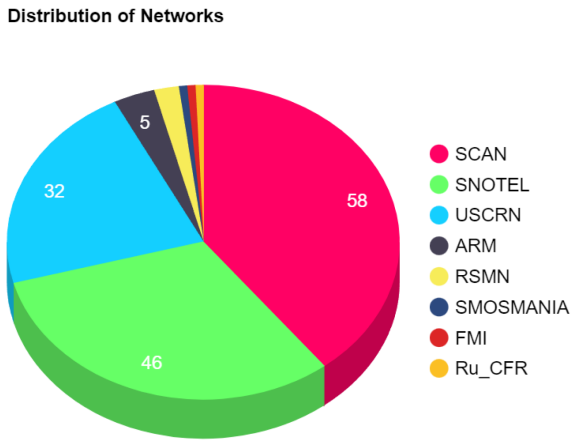

Fig. 2. Number of selected ground stations per network.

spectrum of 171 bands is obtained for each image. A data file is created for each ground station, in which the associated spectra are stored as columns. These files are used for further processing.

III. SMC Estimation

The spectral reflectance of the dry soil is highly dependent on the grain size and its distribution. Addition of small amount of water on the dry soil can significantly impact the overall reflectance of the mixture. The distribution of water in the moist soil defines the reflectance of the mixture. The distribution of water in the moist soil is partially embedded in the spectral reflectance of the soil, when fully saturated, but can vary significantly during drying because of the capillary and adsorptive forces. It thus makes sense to relate the SMC of a wet sample relative to its dry and saturated versions.

A. Sadeghi Model

One such method is provided by the SM. The SM is a physically-based model for soil moisture retrieval in the solar domain (350–2500 nm), based on the Kubelka–Munk two-flux radiative transfer theory. In the SWIR bands, the theory indicates a linear relationship between converted reflectance and soil water content, providing an easy-to-use algorithm in these bands. The accuracy of this model was evaluated and preliminary verified using laboratory-measured spectral reflectance data from various soils. Equation (1) relates R, the soil reflectance, to r, the ratio between the light absorption coefficient (m^{-1}) and the light scattering coefficient (m^{-1})

$$R = 1 + r - \sqrt{r^2 + 2 * r}. \tag{1}$$

Inverting the above equation yields

$$r = ((1 - R)^2/2R). \tag{2}$$

In [23], the following relation was derived:

$$\frac{\theta}{\theta_s} = \frac{r - r_d}{r_s - r_d} \tag{3}$$

where θ and θ_s are the SMC of the moist and saturated soils, respectively, and r, r_d, and r_s are the absorption ratios of the wet, dry, and saturated soils, respectively.

The fundamental disadvantage of the SM is that it is limited to a single (in theory any) wavelength. A wavelength of 2210 nm was suggested in [23]. This limitation prevents an accurate estimation of the SMC. Furthermore, it does not allow to correct for scaling effects due to environmental and acquisition conditions, which makes this model unsuitable for application in outdoor settings, let alone remote sensing.

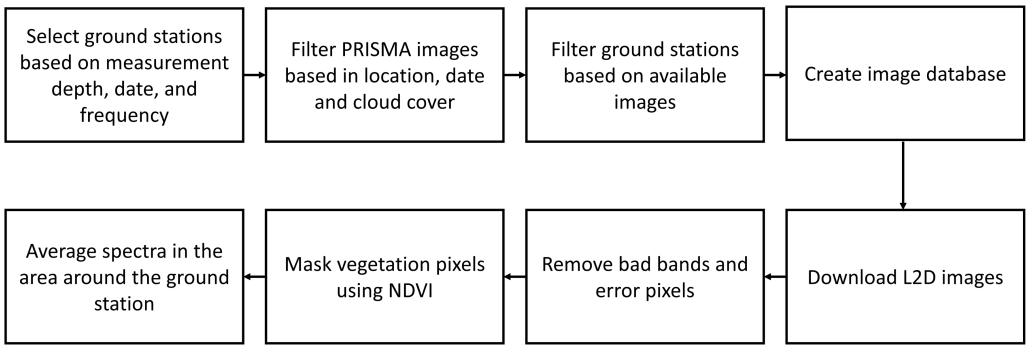

Fig. 3. Flowchart of the data acquisition and preparation.

B. Normalized Relative Arc Lengths (NRAL)

In [39], a data-driven approach for SMC estimation was proposed, using the spectral reflectance of an air-dried and a saturated soil sample. Moist soil is considered as a binary mixture of the air-dried and saturated soil samples, with known spectral reflectance (endmembers) and ground truth SMC. Because the data manifold generated by a number of soil samples with varying SMC forms a (nonlinear) curve in spectral space between the two endmembers, the position of a moist soil sample on this curve, relative to the two endmembers can be regarded as a proxy for its SMC. To estimate the SMC by NRAL, the following three steps are taken.

1) Step 1—Normalizing the Spectra: To obtain the relative position of a moist soil sample on the curve between the two endmembers, its distance from the endmembers along the curve (geodesic distance) is required. Because a curve between two points can be approximated by a piecewise linear curve, the entire length of the curve (i.e., the geodesic distance between the two points) can be approximated by summing the lengths of all straight lines (i.e., the Euclidean distances) between intermediate points. The more intermediate points are available, the better the approximation. However, to estimate the SMC of a spectrum, typically only that spectrum and the two endmembers are available, and thus the approximation will be bad. In that case, the relative position can be obtained simply by projecting all spectra onto the unit hypersphere, i.e., by dividing each spectrum by its length ($R \rightarrow (R/\|R\|)$. On the unit sphere, the arc length between any two spectra is simply given by the angle between them and may be determined by computing the arc cosine of their dot product. So, no intermediate points are required. Another advantage of the projection is that all spectra become invariant to scaling effects, the most prominent effects caused by environmental variations (e.g., illumination variations).

2) Step 2—Projecting the Spectra Onto the ARC Between the Two Endmembers: It is not guaranteed that after projection, a moist sample will lie on the arc linking the two endmembers (see Fig. 4). In [39], the spherical law of cosines was applied to obtain the following:

$$\cos b_1 = \frac{\sin(b_1 + b_2)}{\sqrt{\left(\frac{\cos c'}{\cos c} - \cos(b_1 + b_2)\right)^2 + \sin^2(b_1 + b_2)}} \tag{4}$$

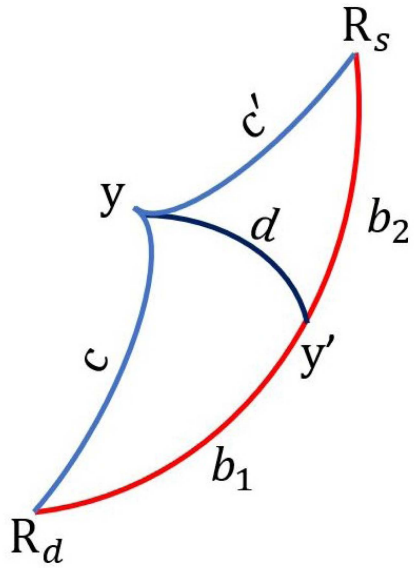

Fig. 4. Red curve: Arc connecting the dry endmember $\mathbf{R}_d$ and the saturated endmember $\mathbf{R}_s$; blue curves: the arcs connecting the moist soil ($\mathbf{y}$) with the endmembers. c and c' denote the arc lengths between $\mathbf{y}$ and the endmembers, respectively. $\mathbf{y}'$ denotes the projection of $\mathbf{y}$ on the arc, and b_1 and b_2 denote the true arc lengths between $\mathbf{y}'$ and the endmembers.

where $b_1 + b_2 = \arccos(R_d^T R_s)$. Then, the vector containing the relative arc lengths of the sample between the saturated and the dry endmember is given by

$$\hat{\mathbf{a}} = \begin{bmatrix} \frac{b_2}{b_1+b_2} \\ \frac{b_1}{b_1+b_2} \end{bmatrix}. \tag{5}$$

3) Step 3—Calibrating the Relative ARC Length: A direct relationship exists between the relative arc length and the SMC. This relationship is expected to be the same for all soils that can hold the same amount of water in their saturated state form. This amount, however, can vary significantly between soils. The relative arc length can be calibrated against the SMC of the saturated soil, by multiplying the relative arc length of the moist soil $\hat{\mathbf{a}}$ with the SMC of the saturated soil (θ_s) ($\hat{\theta} \rightarrow \hat{\mathbf{a}} \times \theta_s$). This way, a more general relationship between the arc length and the SMC is obtained, which is similar for different soils.

Experiments on a large number of laboratory datasets have shown that this normalized relative arclength is an accurate

approximation of the SMC and that the relationship between the normalized relative arc length and the SMC is the same for different soils (see [39], Fig. 7).

To be applicable for remote sensing data, the proposed approach requires some modifications. First of all, it is highly unlikely that situations will arise where the soil is fully saturated. Step 3 of the procedure can be modified by calibrating the relative arc lengths against the SMC of the wettest available soil spectrum. Each ground station has a number of associated PRISMA images. From each of these images, an average spectrum is obtained. For each ground station, the spectrum corresponding with the highest ground SMC measurement of that station is chosen as the wet endmember.

Also for the dry endmember, some modifications are required. It cannot be guaranteed that a spectrum associated with a ground station measurement of zero SMC corresponds to a dry spectrum, and the smallest amount of water can alter the spectrum dramatically. Therefore, the spectrum associated with the smallest SMC of a station cannot be used as dry endmember. The general relationship between the arc length and the SMC is similar for different soils. This suggests that, although the position of the wettest spectrum obtained from different stations can vary a lot on this curve due to variability in soil grain size (texture) and grain size distribution, dry spectra obtained from different stations (if they would be available) are expected to be concentrated on one extreme of this curve. Therefore, we fix the dry endmember for all stations, by selecting one spectrum from all spectra related to zero SMC measurements of all ground stations.

The dry endmember is selected by visually inspecting and comparing spectra to dry spectra of laboratory measurements. Several dry spectra show similar behavior and the particular choice was observed to be not critical for the final results. In this work, the dry endmember was chosen from station Jornada Exp Range, which is part of the SCAN network. The related PRISMA image is selected on the date 27 May 2020, with a 0% cloud cover. The ground station SMC for this date is 0%.

IV. Experiments and Results

In this section, we perform several experiments to validate the potential of the PRISMA hyperspectral remote sensing data and the use of NRAL for estimating SMC. The measure to indicate performance is the root-mean-squared error (RMSE), i.e., the error between the estimated SMC ($\hat{\theta}$) and the ground station SMC (θ)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i - \hat{\theta}_i)^2} \times 100 \tag{6}$$

where n is the number of spectra. Another applied measure is the R^2 measure, defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2}{\sum_{i=1}^{n}(\hat{\theta}_i - <\theta>)^2} \tag{7}$$

where $<\theta>$ is the average ground station SMC.

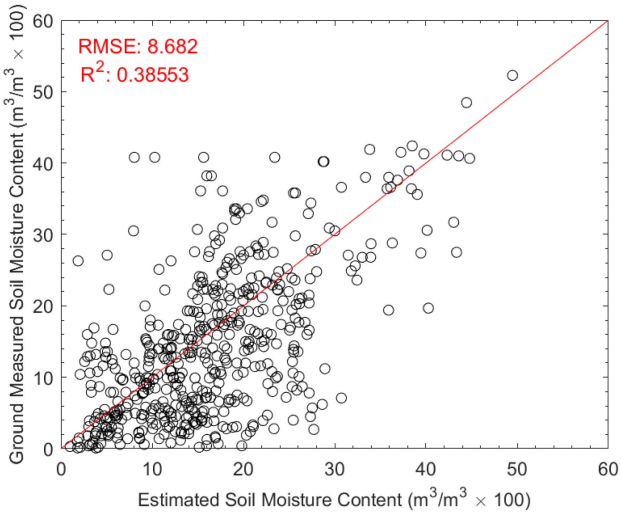

Fig. 5. Ground measured versus estimated SMC.

A. Step-by-Step Illustration of NRAL on a Single Station

To illustrate the use of NRAL for SMC estimation, we selected a number of spectra and their ground station SMC from one particular station: "Powder Mill." Fig. 6(a) shows the spectra. From Fig. 6, it is clear that water dominates the spectra in the entire SWIR wavelength region, offering great potential to estimate water content, even though the atmospheric water absorption bands have been removed.

Fig. 6 also shows that the ordering between the spectra does not correspond to the ordering between their corresponding ground station SMCs. This is due to scaling effects arising from differences in environmental conditions between the images. Step 1 of the NRAL procedure removes these effects by normalizing the spectra [see Fig. 6(b)]. The order of the normalized spectra corresponds to the order of the ground station SMC.

To compare the remote sensing situation to the laboratory environment, in Fig. 7, we show measured and normalized spectra and their corresponding ground-truth water content of a Goethite powder sample, measured in laboratory by a spectroradiometer. Although acquired in a controlled environment, the general behavior of the spectra is similar to the PRISMA spectra.

In addition, this first normalization step allows the determination of the relative position of the spectra between the endmembers. In the first column of Table I, the cosine angles between the spectra at the Powder Mill station and the dry spectrum are given. The second step of NRAL projects the spectra onto the arc between the dry and wet endmembers to obtain the relative position of each spectrum (as shown in the second column of the table). In the third step, these relative positions are calibrated against the SMC of the wet endmember to obtain the estimated SMC and results are shown in the third column. To compare, we also show the estimated SMC by the SM, as shown in the fourth column. The fifth column shows the ground station SMC.

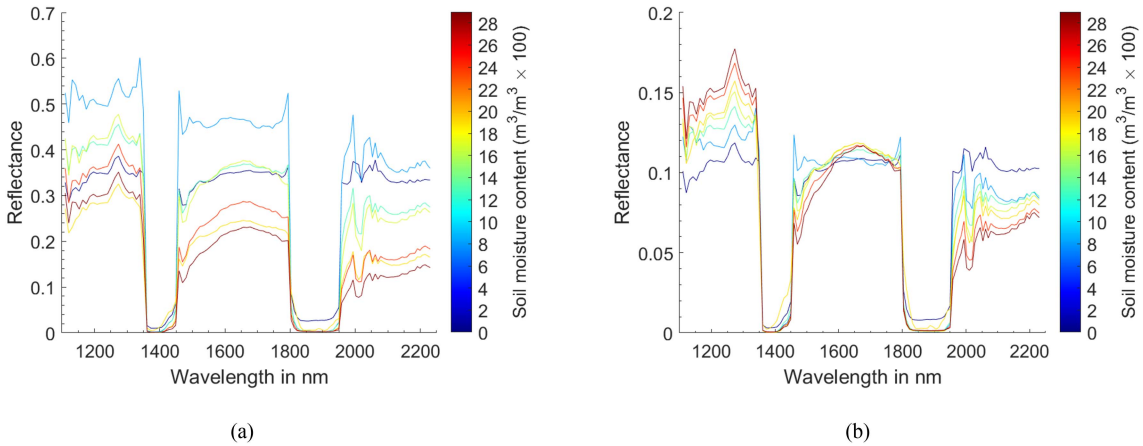

Fig. 6. PRISMA spectra for powder mill station. (a) Measured reflectance for station powder mill. (b) Normalized reflectance for station powder mill.

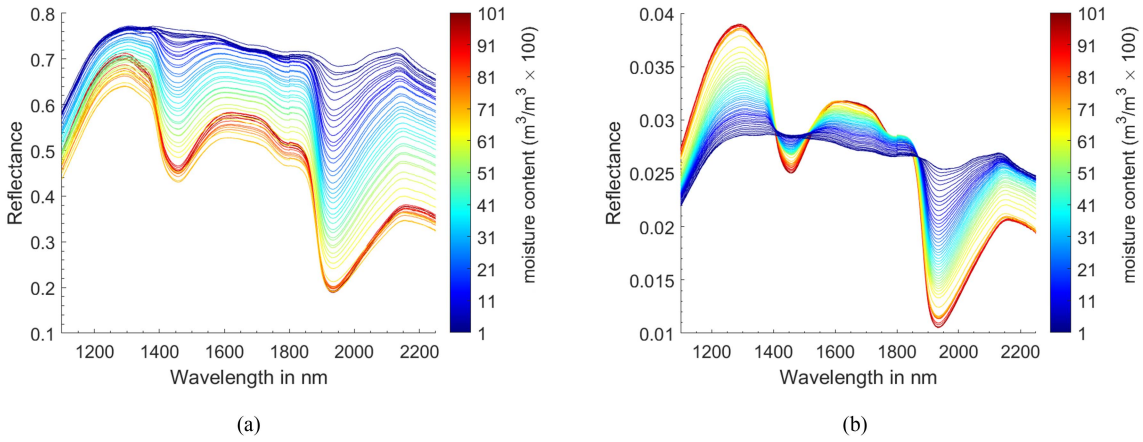

Fig. 7. Lab measured spectra of Goethite powder. (a) Measured reflectance for Goethite sample. (b) Normalized reflectance for Goethite sample.

TABLE I
APPLICATION OF NRAL TO THE SPECTRA OF POWDER MILL STATION

Cosine angle (degree)	Relative position	Estimated SMC NRAL	Estimated SMC Sadhegi	Ground station SMC
0.000	0.000	0.000	0.000	0.000
8.426	0.326	0.096	-0.049	0.083
10.308	0.485	0.142	0.086	0.128
12.730	0.607	0.178	0.104	0.171
13.941	0.647	0.190	0.258	0.190
18.021	0.875	0.257	0.230	0.236
20.536	1.000	0.293	0.293	0.293

B. Determination of the Optimal Spectral Averaging Radius

To determine the optimal radius from the station within which the spectra are averaged, we chose the eight stations with the largest number of corresponding images (i.e., 230 PRISMA images in total). We then chose different radii from the station (radii of 1, 3, and 5 km were chosen), and the spectra were prepared following the steps outlined in Section II. Vegetation pixels and bad pixels identified using the error matrix were masked to avoid interference. The pixels within the prescribed radii were then averaged to determine the final spectra for each image. When averaging, a small number of images was removed in case the area had insufficient soil pixels.

Then, SMC was estimated. For the radii of 3 and 5 km, similar results were obtained with RMSE values of 7.1234 and 8.0131, respectively. For the 1 km radius, an RMSE of 8.1078 was obtained but several images had to be removed because no soil pixels were available in 1 km radius. The radius of 3 km was chosen as the optimal spectral averaging radius for the entire dataset.

TABLE II
SELECTION OF SAMPLES

	Nr. of spectra
Initial nr. of spectra	1418
Spectra corresponding to zero ground station SMC	273
Spectra projected on dry endmember	79
Spectra projected on wet endmember	396
Endmembers (one dry and 151 wet)	152
Final nr. of spectra	518

C. Experiment on the Entire Data Set

Initially, from 151 corresponding ground stations, 1418 PRISMA images with overlapping ground stations were downloaded. From these images, spectra were extracted as the average of the pixels within a 3 km radius from the ground station. Not all of the obtained spectra provide a reliable SMC estimate. In particular, four groups of spectra are removed. First, spectra corresponding to a zero ground station SMC are removed, because their SMC estimation cannot be trusted. Given the fixed dry endmember, their estimate will likely be higher than zero. Second, when projected on the arc between the endmembers, some spectra fall outside the endmember range. NRAL projects these spectra onto either the dry or wet endmember, which, respectively, overestimates or underestimates their SMC. These spectra are removed. Finally, the one fixed dry endmember and the wet endmembers of each of the 151 stations are removed, because they produce by definition a zero error. Ultimately, 518 spectra were retained. Table II, shows in detail the selection of the spectra.

Fig. 5 shows the scatterplot between SMC from the ground station and SMC estimated by NRAL, using the 518 selected samples. An RMSE of 8.682 and R^2 of 0.385 were obtained.

D. SMC Maps

The goal of the previous experiment was to validate the SMC estimation approach using ground station SMC measurements. In this experiment, pixel-based SMC estimation is applied to generate spatial SMC maps. As an example, an image around the station harms way is chosen, and the region within a 3 km radius centered around the station is mapped. Fig. 8 shows the RGB image of the chosen area. The measured ground station SMC is 17.30, while the estimated SMC based on the averaged spectrum is 19.10. shows the obtained SMC map without removing vegetation pixels by NDVI and error pixels. Fig. 9 shows that SMC is higher in vegetated areas and lower in barren areas. Fig. 9 displays a saturated SMC value of 24.5 for a large region corresponding to vegetation, which is the SMC of the wet endmember for the harms way station. Fig. 10 shows the obtained SMC map after removing vegetation and bad pixels. It can be observed that a large amount of vegetation pixels has been masked. However, still a significant amount of spectra fall outside the endmember range. This shows that we were unable to remove all of the vegetation, probably due to a large amount of mixed vegetation-soil pixels.

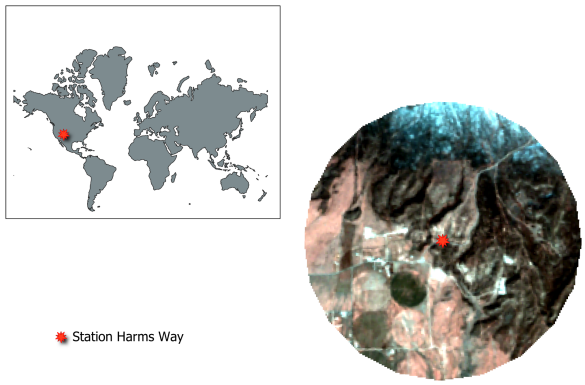

Fig. 8. RGB image of the area selected to generate an SMC map.

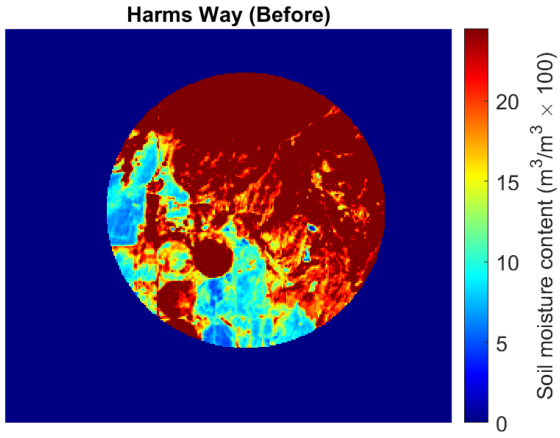

Fig. 9. SMC map generated without removing the error and vegetation pixels.

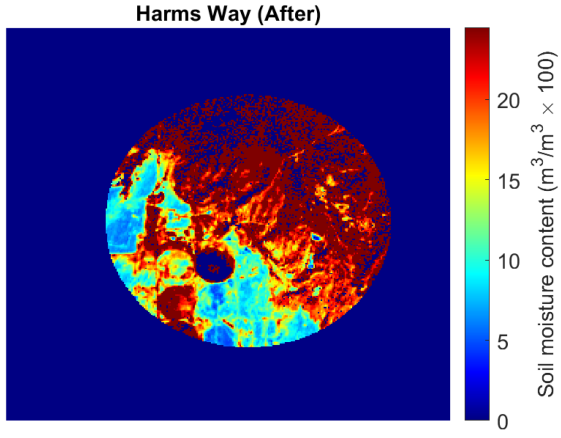

Fig. 10. SMC map after removing the error and vegetation pixels.

V. DISCUSSION

The presented results should be put in perspective, and the discrepancy between the remote sensing data estimates and the ground station measurements can be attributed to a number of factors, such as the quality of spectral reflectance data, when acquired via remote sensing, the limited spatial resolution, differences in environmental acquisition conditions, atmospheric correction effects, the influence of vegetation, etc.

1) First, the intrinsic limitation of spectral reflectance for soil moisture estimation is that it only contains information from a thin top layer of the soil, typically 100–300 μm. Although only stations with a measured SMC of the top 5 cm soil layer are retained, there may be large discrepancy with the estimated SMC. Also, the methodology is not useful for applications where bulk information about soil moisture is required.
2) Reflectance is affected by atmospheric effects, such as gases and aerosols. To mitigate these effects, the PRISMA images that have been used are processed at L2D level. These are the images that have been geometrically and atmospherically corrected. Moreover, error pixels are removed from the image to produce more accurate spectra.
3) Cloud cover can lead to missing information in images. For this reason, only PRISMA images were selected with less than 20% cloud cover. However, we observed that the majority of the applied images have less than 1% cloud covered.
4) Unlike laboratory experiments, where environmental and acquisition conditions are controlled, in remote sensing these conditions can vary significantly between different acquisitions. The PRISMA images cover a wide range and are acquired over a large time span, at different times of the day and different seasons, leading to a broad distribution of different solar elevations and azimuth angles. These varying illumination conditions along with topographic variations cause large spectral variability. As most of these variations cause scaling of the spectra, these effects are mitigated by the normalization step in the proposed procedure.
5) Due to the limited spatial resolution, pure soil pixels are rare, and most pixels are mixtures of soil and vegetation. Because vegetation contains much more moisture than soil, the chosen wet endmember spectrum may not correspond well with the ground measured SMC, and SMC is typically overestimated. This is observed in the generated SMC map (see Figs. 9 and 10). NDVI was used to remove vegetation pixels as much as possible from the image. The threshold of 0.9 for the NDVI value is chosen such that pure vegetation pixels are removed. A lower value would remove too many mixed soil–vegetation pixels, leaving not enough soil pixels for the averaging procedure. Many of the remaining pixels are mixed soil–vegetation pixels. To limit the effect of mixed pixels, the averaging procedure within a 3 km ratio around the station is applied. From Table II, it is clear that even then, the SMC is regularly overestimated, as a significant amount of spectra are projected onto the wet endmember.
6) Due to the limited spatial resolution, an image can be expected to contain different soil types, each with its own water-bearing properties. Even though this effect can significantly affect the spectral reflectance, it can be mitigated by calibrating all spectra corresponding to a station against a distinct wet endmember of that station (Step 3 of the procedure). It has previously been shown in laboratory experiments that this strategy led to SMC estimations that are less sensitive to soil type.
7) The fact that a significant amount of spectra are projected onto the wet endmember emphasizes that the choice of the wet endmember is crucial. The endmember spectrum is chosen as the spectrum corresponding to the highest SMC ground station measurement. Given the limited spatial resolution and the averaging procedure, that spectrum may not correctly reflect the corresponding SMC measurement, e.g., due to the influence of vegetation.
8) As a consequence of all these effects, there may be large deviations between the estimated SMC and the ground station SMC. However, the results confirm the correlation between remote sensing and ground station measurements, with a global error of less than 10%. Moreover, the data allow for a spatial mapping of the SMC.

VI. Conclusion

In this work, we have explored the potential of hyperspectral remote sensing for large-scale SMC estimation. PRISMA hyperspectral images were acquired near ground stations of the ISMN. A procedure is then developed to estimate the SMC from the images and validate the results with the measured SMC from the ground stations. The results demonstrate that SMC can be estimated from remote sensing reflectance in the SWIR range. The limited spatial resolution requires special attention to vegetation and mixed soil–vegetation pixels. The estimation approach requires an endmember spectrum of high moisture content along with its ground-truth moisture content. Results indicate that the quality of this endmember is crucial.

Acknowledgment

The authors would like to thank Agenzia Spaziale Italiana (ASI) for providing PRISMA hyperspectral remote sensing data.

References

[1] D. Datta, M. Paul, M. Murshed, S. W. Teng, and L. Schmidtke, "Soil moisture, organic carbon, and nitrogen content prediction with hyperspectral data using regression models," *Sensors (Basel, Switzerland)*, vol. 22, 2022, Art. no. 7998.
[2] R. D. S. Ansari, "Estimation of soil moisture content: A review," *Int. J. Theor. Appl. Mech.*, vol. 12, no. 3, pp. 571–577, 2017.
[3] K. Furtak and A. Wolińska, "The impact of extreme weather events as a consequence of climate change on the soil moisture and on the quality of the soil environment and agriculture–a review," *CATENA*, vol. 231, 2023, Art. no. 107378.
[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, 2010.
[5] E. Babaeian, M. Sadeghi, S. B. Jones, C. Montzka, H. Vereecken, and M. Tuller, "Ground, proximal, and satellite remote sensing of soil moisture," *Rev. Geophys.*, vol. 57, no. 2, pp. 530–616, 2019.
[6] Z.-L. Li, P. Leng, C. Zhou, K.-S. Chen, F.-C. Zhou, and G.-F. Shang, "Soil moisture retrieval from remote sensing measurements: Current knowledge and directions for the future," *Earth- Sci. Rev.*, vol. 218, 2021, Art. no. 103673.
[7] W. Liu, F. Baret, X. Gu, B. Zhang, Q. Tong, and L. Zheng, "Evaluation of methods for soil surface moisture estimation from reflectance data," *Int. J. Remote Sens.*, vol. 24, no. 10, pp. 2069–2083, 2003.
[8] L. Weidong, F. Baret, G. Xingfa, T. Qingxi, Z. Lanfen, and Z. Bing, "Relating soil surface moisture to reflectance," *Remote Sens. Environ.*, vol. 81, no. 2, pp. 238–246, 2002.
[9] D. Lobell and G. Asner, "Moisture effects on soil reflectance," *Soil Sci. Soc. Amer. J.*, vol. 66, no. 3, pp. 722–727, 2002.

[10] A. Bablet et al., "MARMIT: A multilayer radiative transfer model of soil reflectance to estimate surface soil moisture content in the solar domain (400–2500nm)," *Remote Sens. Environ.*, vol. 217, pp. 1–17, 2018.
[11] P. Dobriyal, A. Qureshi, R. Badola, and S. A. Hussain, "A review of the methods available for estimating soil moisture and its implications for water resource management," *J. Hydrol.*, vol. 458-459, pp. 110–117, 2012.
[12] S. Haubrock, S. Chabrillat, C. Lemmnitz, and H. Kaufmann, "Surface soil moisture quantification models from reflectance data under field conditions," *Int. J. Remote Sens.*, vol. 29, no. 1, pp. 3–29, 2008.
[13] Y. Wang et al., "Quantifying the effects of climate and vegetation on soil moisture in an arid area, China," *Water*, vol. 11, no. 4, 2019, Art. no. 767. [Online]. Available: https://www.mdpi.com/2073-4441/11/4/767
[14] S. Li and Y. Sawada, "Soil moisture-vegetation interaction from near-global in-situ soil moisture measurements," *Environ. Res. Lett.*, vol. 17, no. 11, 2022, Art. no. 114028.
[15] S. U. S. Lekshmi, D. N. Singh, and M. S. Baghini, "A critical review of soil moisture measurement," *Measurement*, vol. 54, pp. 92–105, 2014.
[16] H. Wang et al., "Humidity-sensitive PMMA fiber Bragg grating sensor probe for soil temperature and moisture measurement based on its intrinsic water affinity," *Sensors*, vol. 21, 2021, Art. no. 6946.
[17] J. Tian and W. D. Philpot, "Relationship between surface soil water content, evaporation rate, and water absorption band depths in SWIR reflectance spectra," *Remote Sens. Environ.*, vol. 169, pp. 280–289, 2015.
[18] A. Nagy, P. Riczu, B. Gálya, and J. Tamás, "Spectral estimation of soil water content in visible and near infra-red range," *Eurasian J. Soil Sci.*, vol. 3, pp. 163–171, 2014.
[19] D. Zhang and G. Zhou, "Estimation of soil moisture from optical and thermal remote sensing: A review," *Sensors (Basel, Switzerland)*, vol. 16, 2016, Art. no. 1308.
[20] N. Y. et al., "The soil moisture active passive experiments: Validation of the SMAP products in Australia," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, pp. 2922–2939, Apr. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9145804/
[21] E. G. Njoku and D. Entekhabi, "Passive microwave remote sensing of soil moisture," *J. Hydrol.*, vol. 184, no. 1, pp. 101–129, 1996.
[22] S. Khanal, J. Fulton, and S. Shearer, "An overview of current and potential applications of thermal remote sensing in precision agriculture," *Comput. Electron. Agriculture*, vol. 139, pp. 22–32, 6 2017.
[23] M. Sadeghi, S. B. Jones, and W. D. Philpot, "A linear physically-based model for remote sensing of soil moisture using short wave infrared bands," *Remote Sens. Environ.*, vol. 164, pp. 66–76, 2015.
[24] R. M. Hoffer and C. J. Johannsen, "Ecological potential in spectral signatures analysis," *Remote Sens. Ecol.*, pp. 1–16, 1969.
[25] A. Ångström, "The albedo of various surfaces of ground," *Geografiska Annaler*, vol. 7, pp. 323–342, 1925.
[26] W. G. Planet, "Some comments on reflectance measurements of wet soils," *Remote Sens. Environ.*, vol. 1, no. 2, pp. 127–129, 1970.
[27] S. A. Bowers and R. J. Hanks, "Reflection of radiant energy from soils," *Soil Sci.*, vol. 100, pp. 130–138, 1965.
[28] D. L. Neema, A. Shah, and A. N. Patel, "A statistical optical model for light reflection and penetration through sand," *Int. J. Remote Sens.*, vol. 8, no. 8, pp. 1209–1217, 1987.
[29] D. Levitt, J. Simpson, and A. Huete, "Estimates of surface soil water content using linear combinations of spectral wavebands," *Theor. Appl. Climatol.*, vol. 42, no. 4, pp. 245–252, 1990.
[30] Z. Gao, X. Xu, J. Wang, H. Yang, W. Huang, and H. Feng, "A method of estimating soil moisture based on the linear decomposition of mixture pixels," *Math. Comput. Model.*, vol. 58, no. 3, pp. 606–613, 2013.
[31] J. Tian, J. Yue, W. D. Philpot, X. Dong, and Q. Tian, "Soil moisture content estimate with drying process segmentation using shortwave infrared bands," *Remote Sens. Environ.*, vol. 263, 2021, Art. no. 112552.
[32] A. Lesaignoux, S. Fabre, and X. Briottet, "Influence of soil moisture content on spectral reflectance of bare soils in the 0.4–14 μm domain," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2268–2285, 2013.
[33] Z. Yin, T. Lei, Q. Yan, Z. Chen, and Y. Dong, "A near-infrared reflectance sensor for soil surface moisture measurement," *Comput. Electron. Agriculture*, vol. 99, pp. 101–107, 2013.
[34] M. L. Whiting, L. Li, and S. L. Ustin, "Predicting water content using Gaussian model on soil spectra," *Remote Sens. Environ.*, vol. 89, no. 4, pp. 535–552, 2004.
[35] J. Peng, H. Shen, and S. He, "Soil moisture retrieving using hyperspectral data with the application of wavelet analysis," *Environ. Earth Sci.*, vol. 69, pp. 279–288, 2013.
[36] A. M. Mouazen, R. Karoui, J. De Baerdemaeker, and H. Ramon, "Characterization of soil water content using measured visible and near infrared spectra," *Soil Sci. Soc. Amer. J.*, vol. 70, no. 4, pp. 1295–1302, 2006.
[37] W. Philpot, "Spectral reflectance of wetted soils," *Proc. ASD IEEE GRS*, vol. 2, pp. 1–12, 2010.
[38] H. Bach and W. Mauser, "Modelling and model verification of the spectral reflectance of soils under varying moisture conditions," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 1994, pp. 2354–2356.
[39] B. Koirala, Z. Zahiri, and P. Scheunders, "A robust supervised method for estimating soil moisture content from spectral reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5539813.
[40] K. V. Jambhali, B. Koirala, Z. Bnoulkacem, and P. Scheunders, "Soil moisture content estimation from hyperspectral remote sensing data," in *Proc. IEEE 13th Workshop Hyperspectral Imag. Signal Process.: Evol. Remote Sens.*, 2023, pp. 1–5.
[41] W. D. et al., "The international soil moisture network: Serving Earth system science for over a decade," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 11, pp. 5749–5804, 2021. [Online]. Available: https://hess.copernicus.org/articles/25/5749/2021/
[42] N. Earth, "1:110m cultural vectors," (n.d.). [Online]. Available: https://www.naturalearthdata.com/downloads/110m-cultural-vectors/

Ketaki Vinay Jambhali (Graduate Student Member, IEEE) received the joint M.Sc. degree in geoinformation science and Earth observation from the Indian Institute of Remote Sensing, Dehradun, India, in 2022, and ITC and University of Twente, Enschede, The Netherlands.

She is currently a Ph.D. Researcher with the Vision Lab, Department of Physics, University of Antwerp, Antwerp, Belgium. Her research focuses on the quantitative characterization of water in materials using hyperspectral remote sensing data.

Bikram Koirala (Member, IEEE) received the Ph.D. degree in physics from the University of Antwerp, Antwerp, Belgium, in 2021, with a dissertation entitled "Development of Advanced Hyperspectral Unmixing Methods."

He is currently a Postdoctoral Researcher with the Research Foundation—Flanders (FWO), Brussels, Belgium. His research interests include machine learning, deep learning, radiative transfer modeling, and hyperspectral image processing.

Zakaria Bnoulkacem (Graduate Student Member, IEEE) received the M.Sc. degree in surveying engineering and geomatics from the Agronomic and Veterinary Institute Hassan II (IAV Hassan II), Rabat, Morocco, in 2021.

He is currently a Ph.D. Researcher with the VisionLab, Department of Physics, University of Antwerp, Antwerp, Belgium. His research focuses on material characterization using spectral reflectance for intimate mixtures.

Paul Scheunders (Senior Member, IEEE) received the M.S. and the Ph.D. degrees in physics, with work in the field of statistical mechanics, from the University of Antwerp, Antwerp, Belgium, in 1986 and 1990, respectively.

In 1991, he was a Research Associate with the Vision Lab, Department of Physics, University of Antwerp, where he is currently a Full Professor. He has authored or coauthored more than 200 papers in international journals and proceedings in the field of image processing, pattern recognition, and remote sensing. His current research interests include remote sensing and hyperspectral image processing.

Dr. Scheunders is an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and was a Program Committee Member in numerous international conferences. He is also a Senior Member of IEEE Geoscience and Remote Sensing Society.