Toward the Removal of Model Dependency in Soil Moisture Climate Data Records by Using an *L*-Band Scaling Reference

Remi Madelon, Nemesio Rodriguez-Fernandez, Robin van der Schalie, Tracy Scanlon, Ahmad Al Bitar, Yann Kerr, Richard de Jeu, Wouter Dorigo

Toward the Removal of Model Dependency in Soil Moisture Climate Data Records by Using an L-Band Scaling Reference

Rémi Madelon, Nemesio J. Rodríguez-Fernández, Robin van der Schalie, Tracy Scanlon, Ahmad Al Bitar, Yann H. Kerr, *Fellow, IEEE*, Richard de Jeu, and Wouter Dorigo

***Abstract*—Building climate data records of soil moisture (SM) requires computing long time series by merging retrievals from sensors on-board different satellites, which implies to perform a bias correction or rescaling on the original time series. Due to their long time span and high temporal frequency, model data could be used as a common reference for the rescaling. However, avoiding model dependence in observational climate data records is needed for some applications. In this article, the possibility of using as reference remote sensing data from one of the L-band sensors specifically designed to measure SM is discussed. Advanced Microwave Scanning Radiometer 2 SM time series were rescaled by matching their cumulative distribution functions (CDFs) to those of Soil Moisture and Ocean Salinity (SMOS), Soil Moisture Active Passive (SMAP), and Global Land Data Assimilation System (GLDAS) NOAH model time series. The CDF computation was investigated as a function of the time series length, finding significant differences from four to nine years. Replacing temporal by spatial variance does not allow us to compute better CDFs from short time series. The rescaled time series show a high correlation ($R > 0.8$) to the original ones and a low bias with respect to the reference ($<0.03\,\mathrm{m}^3 \cdot \mathrm{m}^{-3}$). The time series rescaled using several SMOS or SMAP datasets were also evaluated against *in situ* measurements and show performances similar to or slightly better than those rescaled using the model GLDAS. The impact of random errors and gaps of the observational data into the rescaling was evaluated. These results show that it is actually possible to use L-band data as reference to rescale time series from other sensors to build long time series of SM.**

***Index Terms*—Advanced microwave scanning radiometer 2 (AMSR2), cumulative distribution function (CDF) matching, L-band, long time series, soil moisture (SM), soil moisture active passive (SMAP), soil moisture and ocean salinity (SMOS).**

Manuscript received September 7, 2021; revised December 9, 2021; accepted December 14, 2021. Date of publication December 21, 2021; date of current version January 12, 2022. This work was supported by the European Space Agency Climate Change Initiative program. The work of Rémi Madelon, Nemesio J. Rodríguez-Fernández, Ahmad Al Bitar, and Yann H. Kerr was supported by the Centre National d'Etudes Spatiales TOSCA program. *(Corresponding author: Rémi Madelon.)*

Rémi Madelon, Nemesio J. Rodríguez-Fernández, Ahmad Al Bitar, and Yann H. Kerr are with the Centre National d'Etudes Spatiales, Centre National de la Recherche Scientifique, Institut de Recherche pour le Développement, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), Université de Toulouse, Centre d'Etudes Spatiales de la Biosphère, 31401 Toulouse, France (e-mail: remi.madelon@cesbio.cnes.fr; nemesio.rodriguez-fernandez@cesbio.cnes.fr; ahmad.albitar@cesbio.cnes.fr; yann.kerr@cesbio.cnes.fr).

Robin van der Schalie and Richard de Jeu are with the VanderSat B.V., 2011VK Haarlem, The Netherlands (e-mail: rvanderschalie@vandersat.com; rdejeu@vandersat.com).

Tracy Scanlon and Wouter Dorigo are with the Research Group for Climate and Environmental Remote Sensing, Department of Geodesy and Geoinformation, Vienna University of Technology, 1040 Vienna, Austria (e-mail: tracy.scanlon@geo.tuwien.ac.at; wouter.dorigo@geo.tuwien.ac.at).

Digital Object Identifier 10.1109/JSTARS.2021.3137008

I. Introduction

SOIL moisture (SM) was identified as one of the 50 "essential climate variables" (ECVs) by the Global Climate Observing System in the context of the United Nations Framework Convention on Climate Change [1]. It interacts in many hydrological processes, such as infiltration, runoff, precipitation, and evaporation [2], and SM estimates are needed in weather forecasting [3]–[5], in agriculture applications [6], and to monitor extreme climate events, such as floods and droughts [7]–[11]. Long time series of SM are crucial to monitor the Earth's climate evolution, and this is the goal of initiatives such as the European Space Agency's Climate Change Initiative (ESA CCI) program [12]. Since 2012, this project has been aiming at developing consistent satellite-based long-term climate data records (CDRs) of SM from active, passive, and combined instruments [13]–[18].

Remote sensing from space allows a global monitoring of SM, but due to the relatively short operational life of space missions, merging data from different instruments is required. To effectively detect SM trends and changes, it is crucial to benefit from the longest time series possible because natural short-term variations can hide the longer ones. For instance, the strategy developed by the ESA CCI [17] consists of merging different SM datasets obtained with different algorithms applied to measurements from active and passive sensors, on-board satellites, that have different characteristics (frequency, spatial resolution, temporal and spatial coverage, polarization, revisit time, etc.). It is then necessary to apply a bias correction (commonly referred to as scaling or rescaling) on the time series obtained with different sensors to provide a consistent merged dataset.

Different methods can be used [19], [20], but bias correction is frequently performed by matching the cumulative distribution function (CDF) of the source time series to that of a reference time series. This is, for example, the approach used for debiasing satellite data before data assimilation into numerical weather prediction models [4], [21], [22] or into hydrological models [10], [23]. In these cases, the model time series are naturally used as the reference to individually rescale the different remote sensing datasets. CDF matching is also the method selected to

rescale SM time series from different microwaves sensors within the ESA CCI SM project.

First, passive and active products are rescaled using the time series from a passive and an active sensor, respectively, as reference. Finally, a combined product is computed by rescaling the passive and the active using the NOAH Global Land Data Assimilation System (GLDAS) land surface model as reference [13], [17], [24].

CDRs of ECVs have to be extensively validated both using direct (comparison to measurements) and indirect methods (comparison to other CDRs) [25]. For example, observational CDRs of ECVs are needed to evaluate modeled CDRs and to perform data assimilation into models [26], [27]. Therefore, observational CDRs should be as independent to models as possible. In the case of the ESA CCI SM dataset, using the GLDAS model data to rescale active and passive time series introduces, in a statistical sense, information of a model product in the final ESA CCI merged dataset. This drawback is actually discussed by users of the ESA CCI SM for different applications, such as surface and root zone SM monitoring [28], [29], large-scale long-term SM variability studies [17], [30], model evaluation, and data assimilation [26], [31], [32]. The growing need for model-free observational CDRs of ECVs, in particular for SM, is discussed in [16] and in the ESA SM CCI user requirement document, which summarizes the ESA CCI SM user requirement surveys.

As an alternative to rescaling the data using the GLDAS model, it has been proposed to use L-band data from one of the two instruments specifically designed to measure SM, i.e., on-board ESA soil moisture and ocean salinity (SMOS) and National Aeronautics and Space Administration (NASA) Soil Moisture Active Passive (SMAP) satellites, as the reference to rescale other time series [33].

To investigate this approach, in the current article, the Advanced Microwave Scanning Radiometer 2 (AMSR2) SM time series were rescaled using as the reference different datasets from SMAP, SMOS, and the GLDAS NOAH model. A trend and seasonality decomposition of the rescaled time series was performed, and they were evaluated against *in situ* measurements. The effects of the length of the time series used as the reference to compute the CDF were evaluated using the longest coherent L-band SM time series, those based on SMOS, and by comparing the CDFs in different time periods. Finally, the presence of random error and a reduced temporal sampling in satellite-based products was also studied.

The rest of this article is organized as follows. Section II presents the different remotely sensed and model- and ground-based data that were used in this article. Section III describes the different CDF matching approaches applied on the AMSR2 time series and how data were selected for the evaluation. Section IV shows the performances of the reference and rescaled time series with respect to *in situ* measurements. The results of the time-series decomposition and the random error propagation assessment are also presented. Section V discusses the pros and cons of using an L-band dataset as the reference for the rescaling and the impact of having long time series on the CDF computation. Finally, Section VI concludes this article.

II. Data

A. Japan Aerospace Exploration Agency (JAXA) AMSR2

The AMSR2 instrument is a conical scanning passive microwave radiometer built by the JAXA and launched on board the Global Change Observation Mission 1–Water satellite (GCOM–W1) on May 18, 2012 [34], [35]. The instrument still operates, and the GCOM–W1 satellite provides a complete coverage of the Earth every two days since May 2012 (see Fig. 1). Ascending and descending orbits cross the equator at 1:30 P.M. and 1:30 A.M., respectively. AMSR2 follows on the objective of the NASA AMSR-E on board the Aqua satellite and observes brightness temperatures in vertical and horizontal polarizations at seven frequencies: 6.9 (C_1), 7.3 (C_2), 10.7 (X), 18.7 (Ku), 23.8 (K), 36.5 (Ka), and 89.0 (W) GHz, of which only the first four are used as a base for SM retrievals and only the first three (C_1, C_2, and X) were considered in this article, as only these three bands are used in the ESA CCI SM products [36]. The JAXA Earth Observation Research Center provides SM datasets retrieved from the AMSR2 X-band using a radiative transfer model [37] and the JAXA algorithm [38]. However, in the current article, the AMSR2 SM obtained within the framework of the ESA CCI project were used. These retrievals are computed using the Land Parameter Retrieval Model (LPRM) algorithm version 6 [39], and they are available in a 0.25° regular latitude–longitude grid as daily files. Only descending overpasses (morning orbits) from July 2012 to June 2020 were considered in this article. A comparison between the JAXA and LPRM algorithms can be found in [34]. Hereafter, $AM2$ refers to the AMSR2 SM time series and data.

B. ESA SMOS

The SMOS mission is part of the Earth Explorer program from the ESA, with contributions from the Centre National d'Etudes Spatiales, France, and the Centro Para el Desarrollo Tecnológico Industrial, Spain. The satellite was launched on November 2, 2009 to measure globally and frequently the surface SM over land and sea surface salinity over the oceans [40], with a revisit period of one to three days and Equator crossings at 6:00 A.M. and 6:00 P.M. for ascending and descending overpasses, respectively. The SMOS mission relies on a 2-D interferometric instrument operating at the L-band (21 cm, 1.4 GHz) able to measure brightness temperatures in both vertical and horizontal polarizations, at incidence angles from 0° to 60° and with a spatial resolution of 40 km on average. SMOS still operates and provides the longest homogeneous L-band record of SM. To be consistent with the $AM2$ overpasses considered in this article, only the ascending overpasses of SMOS were taken into account as they correspond to the morning orbits. The following four SM datasets derived from different retrieval algorithms were used for the evaluation and as the reference for the rescaling of $AM2$.

1) Centre Aval de Traitement des Données (CATDS) SMOS-L3 SM: The daily CATDS SMOS Level-3 V300 SM product [41] is a multiorbit Level-3 SM product provided by the CATDS. The SM retrieval process is based on the algorithm used for the SMOS Level-2 product [42] and focuses on the iterative

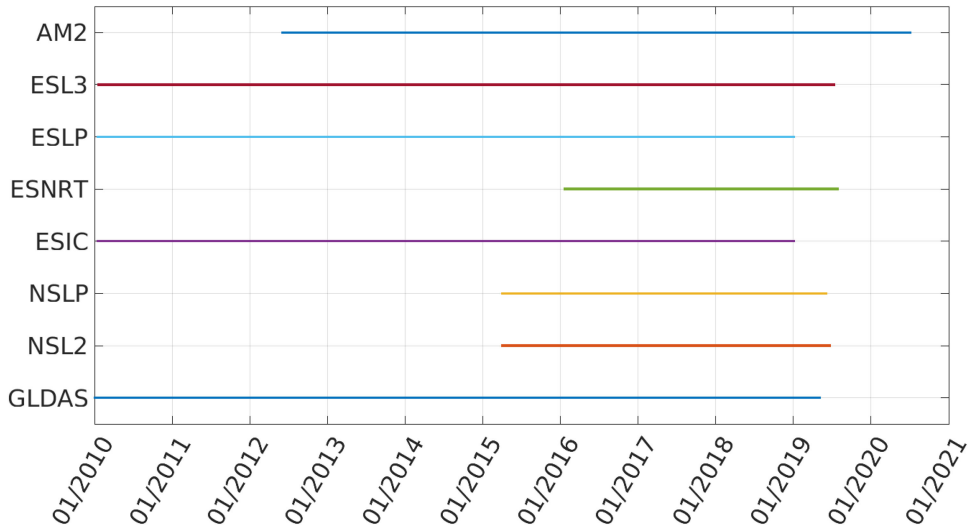

Fig. 1. Temporal coverage of the different datasets used in the article. Acronyms are explained in Section II.

minimization of the difference between a forward model and the brightness temperatures. The model uses the τ–ω (optical depth-single scattering albedo) approach to take into account the effect of vegetation. The data are provided in the 25-km EASE-Grid v2.0 [43] as NetCDF files, and only SM retrievals from January 2010 to July 2019 (see Fig. 1) were considered in the different computations and analysis. Hereafter, $ESL3$ refers to the ESA SMOS CATDS L3 SM time series and data.

2) SMOS-LPRM: The daily SMOS-LPRM SM product has been derived using LPRM version 6 and SMOS Level-3 brightness temperatures. This is the same LPRM algorithm as that used for $AM2$, but the parameters have been optimized for L-band sensors [44]. It is the SMOS dataset that is currently used in the ESA CCI project and is produced in a 0.25° regular latitude–longitude grid. All the available data from January 2010 to December 2018 (see Fig. 1) were taken into account in this article. Hereafter, $ESLP$ refers to the ESA SMOS-LPRM SM time series and data.

3) SMOS-NRT: The ESA SMOS Near Real-Time (NRT) Level-2 V100–V200 SM Neural Network product [45] provides SM estimates computed using statistical coefficients estimated by a neural network trained on the nominal ESA SMOS Level-2 SM, but in NRT for operational applications. The data are provided in the ISEA4H9-grid common to other ESA Level-1 and Level-2 SMOS products. The NRT files contain data for fractions of orbits, and they are distributed in NetCDF format. Only SM estimates from January 2016 to July 2019 (see Fig. 1) were considered in this experiment. Hereafter, $ESNRT$ refers to the ESA SMOS-NRT SM time series and data.

4) SMOS-IC: The INRA-CESBIO SMOS-IC V105 product [46] is a simplified research product that considers aggregated surface emissions at 25-km resolution. The algorithm was designed by the Institut National de la Recherche Agronomique and the Centre d'Etudes Spatiales de la Biosphère to perform global retrievals of SM and L-band VOD. It is based on the two-parameter inversion of the L-MEB model, as defined in [42] and [47], and considers the satellite footprint as homogeneous. The data are provided in the 25-km EASE-Grid v2.0, as NetCDF files. SM retrievals from January 2010 to December 2018 (see Fig. 1) were used in the different computations and analysis. Hereafter, $ESIC$ refers to the ESA SMOS-IC SM time series and data.

C. NASA SMAP

SMAP was launched on January 31, 2015 by NASA. It carries on board a passive radiometer operating at 1.4 GHz (L-band) and a synthetic aperture radar, an active instrument operating at 1.2 GHz. The respective spatial resolutions of the two instruments are 40 km and 1–3 km; however, the radar stopped working a few months after launch. SMAP provides passive measurements of the land surface SM in vertical and horizontal polarizations at a fixed incidence angle of 40° [48]. SMAP ascending and descending orbits cross the equator at 6:00 P.M. and 6:00 A.M., respectively, and the revisit period is between two and three days. To be consistent with the $AM2$ overpasses considered in this article, only the descending overpasses of SMAP were taken into account as they correspond to the morning orbits. The following two SM datasets derived from different retrieval algorithms were used for the evaluation and as the reference for the rescaling of $AM2$.

1) SMAP-LPRM: The SMAP-LPRM SM product provides daily SM estimates using LPRM version 6 and brightness temperatures. This is the same LPRM algorithm as that used for $AM2$, but the parameters have been optimized for L-band sensors [44]. It is the SMAP dataset that is currently used in the ESA CCI project, and it is produced in a 0.25° regular latitude–longitude grid. SM estimates from April 2015 to May 2019 (see Fig. 1) were considered in this experiment. Hereafter, $NSLP$ refers to the NASA SMAP-LPRM SM time series and data.

2) SMAP-L2: The daily NASA SMAP Level-2 V005 SM product [49] provides SM estimates using the single- and double-channel algorithms [50] and water-corrected brightness

TABLE I
IN SITU NETWORKS FROM THE ISMN THAT WERE USED IN THIS ARTICLE

Network	Depth (m)	Sensors	Location	Reference
FR-Aqui	0.05–0.05	4	France	[73]
HOBE	0.05–0.05	61	Denmark	[74]
iRON	0.05–0.05	6	USA	[75]
SCAN	0.05–0.05	61	USA	[76]
SMOSMANIA	0.05–0.05	17	France	[77]
SNOTEL	0.05–0.05	64	USA	[78]
SOILSCAPE	0.05–0.05	34	USA	[79], [80]
TERENO	0.05–0.05	6	Germany	[81]
USCRN	0.05–0.05	44	USA	[82]

Depths are quoted as two numbers: the first one is the upper depth, and the second one is the lower depth of the sensor. Both numbers are equal when the sensor is placed horizontally. The third column gives the number of sensors remaining for each network after applying all of the criteria used for the evaluation that are discussed in Section III-A.

temperatures measured at the L-band. The data are provided in the 36-km EASE-Grid v2.0 as NetCDF files, and only SM retrievals from the single-channel algorithm for V polarization were considered. In this article, the data used cover the time span from April 2015 to June 2019 (see Fig. 1). Hereafter, $NSL2$ refers to the NASA SMAP-L2 SM time series and data.

D. GLDAS NOAH Model

SM model simulations representative of water in the top soil layer (0–10 cm depth) from the GLDAS NOAH-3.6 version 2.1 model [51] were also used as the reference for the CDF matching and compared to remote sensing datasets. These model data result from remotely sensed and ground-based measurements combined together using advanced land surface modeling and data assimilation techniques. They are provided in a 0.25° regular latitude–longitude grid as NetCDF files. Only SM predictions from January 2000 to April 2019 (see Fig. 1) were taken into account for the comparisons and rescaling. The GLDAS NOAH model was chosen in this article because it is the one currently used as the reference to rescale the other sensors time series within the ESA CCI project for SM. Hereafter, $GLDAS$ refers to the GLDAS NOAH model SM time series and data.

E. In Situ Measurements

Ground-based measurements retrieved from the International Soil Moisture Network (ISMN) were used to assess the performances of the rescaled $AM2$ [52], [53]. Since the microwave radiation detected from space comes from the top surface of the soil, only sensors between 0 and 5 cm depth were considered. In total, there were 1740 sensors time series available, but only a few hundred were selected to be compared to remotely sensed and model data (see Section III-A). Table I presents the networks used in this article.

TABLE II
IGBP LAND COVER CLASSES OF THE LOCATIONS OF THE ISMN SENSORS USED IN THIS ARTICLE

IGBP land cover classification	Sensors
Cropland Natural Vegetation Mosaic	18
Croplands	94
Deciduous Broadleaf Forest	4
Evergreen Needleleaf Forest	16
Evergreen Broadleaf Forest	1
Grasslands	87
Mixed Forest	16
Open Shrublands	13
Savannas	34
Woody Savannas	14

Second column is the number of sensors remaining for each class after applying all of the criteria used for the evaluation that are discussed in Section III-A.

F. Land Cover and Climate Classifications

Two well-known classifications were used to assess the performances of the rescaling of $AM2$ during this article.

The first one is the land cover classification from the International Geosphere Biosphere Programme (IGBP) that is based on the Moderate Resolution Imaging Spectroradiometer data [54], [55]. It consists of classifying the Earth land surface into categories that describe the type of observed vegetation, such as croplands, forests, or barrens.

The second classification is the Köppen–Geiger climate classification [56]. This system divides the Earth land surface into fives main climate groups (tropical, dry, temperate, continental, and polar), with each group being divided based on seasonal precipitation and temperature patterns.

According to the locations of the ISMN sensors and the data selection explained in Section III-A, not all the classes from both classifications were represented (see Tables II and III).

III. METHODS

A. Data Selection and Evaluation Against In Situ Measurements

The purpose of this article is to evaluate the performances of the original $AM2$ time series as well as those rescaled using $ESL3$, $ESLP$, $ESNRT$, $ESIC$, $NSLP$, $NSL2$, and $GLDAS$. Hence, it was chosen to compare remotely sensed and model time series to independent *in situ* measurements retrieved from the ISMN.

The first step was to spatially interpolate all the datasets on the 0.25° spaced grid of $AM2$ using the nearest neighbor approach. Then, the remotely sensed and model time series at the closest grid node to each of the 1740 ISMN sensors locations (1397

TABLE III
KÖPPEN–GEIGER CLIMATE CLASSES OF THE LOCATIONS OF THE ISMN SENSORS USED IN THIS ARTICLE

Main climate	Precipitation	Temperatures	Sensors
Arid	Desert	Cold Arid	5
Arid	Steppe	Cold Arid	38
Arid	Steppe	Hot Arid	4
Snow	Fully Humid	Cool Summer	15
Snow	Fully Humid	Hot Summer	6
Snow	Fully Humid	Warm Summer	34
Snow	Summer Dry	Warm Summer	2
Snow	Winter Dry	Cool Summer	2
Warm Temperate	Fully Humid	Hot Summer	55
Warm Temperate	Fully Humid	Warm Summer	84
Warm Temperate	Summer Dry	Hot Summer	42
Warm Temperate	Summer Dry	Warm Summer	10

Last column is the number of sensors remaining for each class after applying all of the criteria used for the evaluation that are discussed in Section III-A.

sites) were retrieved. A maximal radial distance of 25 km was set as the threshold for this selection.

$AM2$ time series were rescaled following the different methods discussed in Section III-B. Hereafter, $AM2_{ESL3}$, $AM2_{ESLP}$, $AM2_{ESNRT}$, $AM2_{ESIC}$, $AM2_{NSLP}$, $AM2_{NSL2}$, and $AM2_{GLDAS}$ refer to $AM2$ rescaled using $ESL3$, $ESLP$, $ESNRT$, $ESIC$, $NSLP$, $NSL2$, and $GLDAS$ as the reference, respectively.

The original $AM2$, $ESL3$, $ESLP$, $ESNRT$, $ESIC$, $NSLP$, $NSL2$, and $GLDAS$ time series as well as the $AM2_{ESL3}$, $AM2_{ESLP}$, $AM2_{ESNRT}$, $AM2_{ESIC}$, $AM2_{NSLP}$, $AM2_{NSL2}$, and $AM2_{GLDAS}$ time series were compared to the ground-based time series by computing the Pearson correlation, the standard deviation of the difference (STDD), and the bias. A minimum of 100 SM samples in common with the *in situ* measurements for each reference and rescaled time series were required to compute those statistical metrics. A maximum difference of 1 h between the acquisition times of the remotely sensed retrievals and ground-based measurements was set as the threshold for the selection. Then, only sensor locations for which all the time series to compare fulfilled the previous condition were used. In the end, the evaluation was done at 297 sensors locations (253 sites; see Fig. 2) from the networks listed in Table I. It is worth mentioning that while these networks cover a large array of climatic zones and soil conditions (see Tables II and III), they are restricted to two continents.

The results of the evaluation with respect to *in situ* measurements were analyzed in three different ways:

1) by considering all the networks at once;
2) for each network independently;
3) as a function of the IGPB land cover and Köppen–Geiger climate classifications (see Tables II and III).

B. Rescaling by CDF Matching

The CDF is a specific way to describe the distribution of a discrete or continuous variable X. The purpose of this function is to give the probability that X will take a value less than or equal to a certain threshold. The CDF of an SM time series can be written as

$$\mathrm{CDF}_{\mathrm{SM}}(X) = P(\mathrm{SM} <= X) \tag{1}$$

where X is a given value of SM and P is the probability. The CDF of a distribution is usually plotted as a curve, where the vertical and horizontal axes show the percentage of samples and the value of the variable, respectively.

Fig. 3 presents the median difference between $ESIC$ from the SMOS launch date (nine years of data) and $ESIC$ from the SMAP launch date (four years of data). Examples of the nine- and four-year long CDF were also displayed and compared at selected locations around the world. In some cases, there are significant differences between the two CDFs, meaning that the time period can greatly affect the CDF computation. These differences are discussed in Section V.

As mentioned in [22], CDF matching consists of transforming the CDF of one variable X (source data) to mimic that of another variable Y (reference data) by using a function f. This process can serve many objectives, but in this article context, it is used to correct for systematic differences between source and reference datasets that are both supposed to describe SM dynamics. Indeed, bias between SM time series can be observed due to differences in designs, incidence angles, spatial footprints, or frequencies used between sensors. According to the nature of the data and the application, different versions of f can be used for the CDF matching. CDF matching is sometimes done as a simple linear transformation using the mean (μ) and the standard deviation (σ) of the source and reference as follows:

$$S_m = A + B \times S \tag{2}$$

where S and R are the source and the reference data, respectively, and S_m is the source data after being matched and coefficients A and B are computed as

$$\begin{aligned} B &= \sigma_R/\sigma_S \\ A &= \mu_R + B \times \mu_S. \end{aligned} \tag{3}$$

As explained by Aires *et al.* [22], this rescaling assumes that both the source and the reference distributions are Gaussian. In this article, the SM random variable distributions from the different datasets were not considered as Gaussian, and hence, this approach will not be discussed further. Empirical CDFs were considered, and the CDF matching was applied using linear piecewise and polynomial transformations. These two techniques are discussed in Sections III-B1 and III-B2.

$AM2$ at C_1-, C_2-, and X-bands (source data) were rescaled following those techniques on a per-pixel basis, using alternatively all the datasets presented in Section II as the reference data. In addition, the CDF matching process was performed at different time scales.

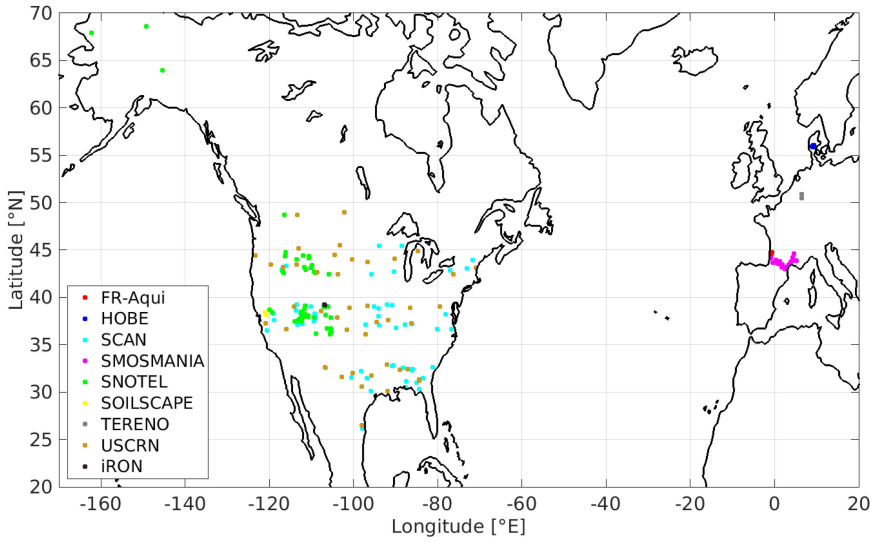

Fig. 2. Locations of the *in situ* sensors and networks from the ISMN that were used in this article.

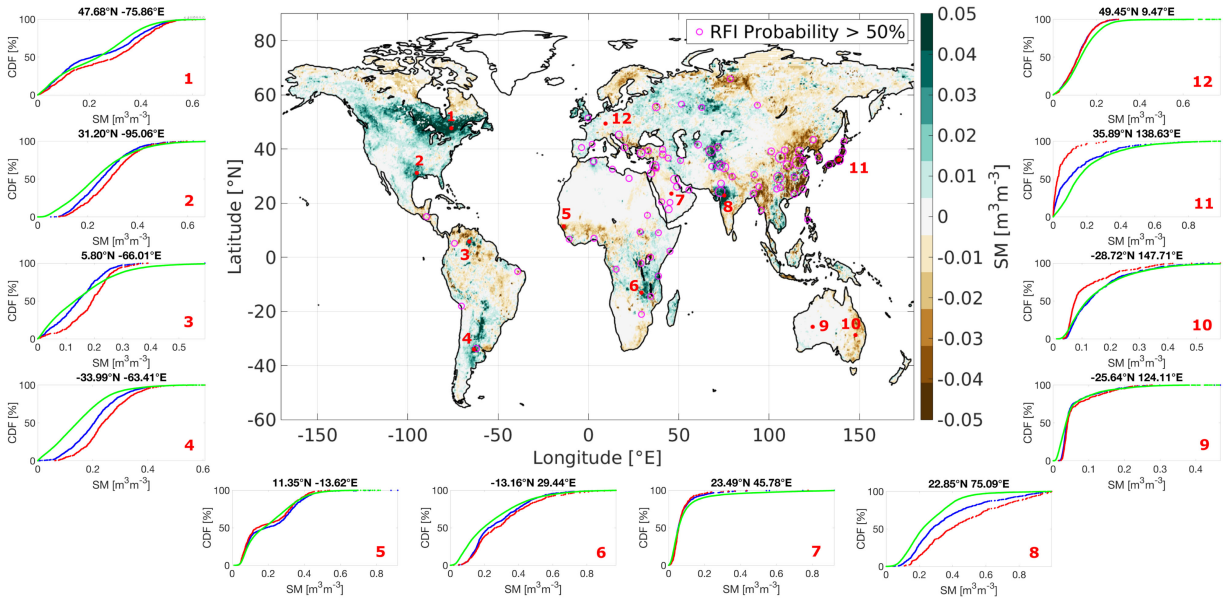

Fig. 3. Map of the mean of the $ESIC$ SM computed in the 2015–2018 period minus the mean computed in the 2010–2018 period. The surrounding panels show a comparison between the $ESIC$ CDF computed in the 2015–2018 period (red line) with respect to that computed in the 2010–2018 period (blue line) for a selected set of locations shown as red dots in the map. The $ESIC$ CDF computed in the 2015–2018 period using all data in a 2° radius is also shown in green line.

1) The source CDF is matched to the reference by computing only one set of scaling parameters for all the days of the year.
2) The source CDF is matched to the reference on a seasonal basis. A specific set of scaling parameters is computed independently for each of the following monthly sets: January, February, March; April, May, June; July, August, September; and October, November, December.
3) The source CDF is matched to the reference on a monthly basis. A specific set of scaling parameters is computed for each month of the year.

One must bear in mind that $AM2$ were rescaled from 2012 to 2020 in all cases, even when the reference time series did not cover this entire period of time. In other words, this means that the scaling parameters are applied to this entire period even though they may be computed from shorter periods.

1) Linear Piecewise CDF Matching: The piecewise approach focuses on the matching of the empirical CDF of the source data with those of the reference data [13], [21], [57]. In this case, both the source and reference data are sorted in increasing order, and a binning process is performed on them according to a predefined set of percentiles. Once the empirical CDF are built, there are many ways to apply the transformation on the source data.

The chosen method here is the one currently used in the ESA CCI project and consists of a piecewise linear interpolation [58].

The idea is to compute a specific slope and intercept between the source and reference data for each percentile. Then, the rescaling of the source data is performed individually for each percentile by using the corresponding linear coefficients. An additional linear least-squares regression is applied between the source and reference data by using the values smaller to the second percentile with an intercept through the second percentile [58]. This is also undertaken for values above the second latest percentile. The computed slope and intercept are then used to rescale the source data below the first and above the last percentile, respectively. The aim of this process is mainly to reduce the impact of potential outliers during the rescaling. Following [58], in this article, a set of 12 percentiles bins is used as default: [0–5, 5–10, 10–20, 20–30, 30–40, 40–50, 50–60, 60–70, 70–80, 80–90, 90–95, 95–100]. Each percentile bin was required to include at least 20 SM observations. If this is not the case for any of the percentile bins, the number of bins is reduced until there are at least 20 SM samples in each of them.

Hereafter, $ILPW$ CDF matching refers to the improved linear piecewise CDF matching technique.

2) Polynomial Fitting CDF Matching: The polynomial fitting CDF matching, explained in [23], uses a function approximation technique to find the relationships between the source and the reference data CDF. First, the source and reference CDFs are computed with the number of percentiles equal to the number of SM observations in their respective time series. Second, the source CDF is linearly interpolated on the percentiles of the reference CDF, and the difference between the two is computed. Then, the difference is plotted against the interpolated source CDF, and a polynomial fit is performed. Finally, the polynomial function is used to compute the correction to apply to the source data. As mentioned in [22], this method has the advantage to provide a continuous transformation, but it is more complex to control. Indeed, it can be subjected to instabilities, and extreme events are not sure to be well preserved. A polynomial fit of degree 5 was used to rescale $AM2$ in this article. In the case there was fewer than 15 percentiles in the reference CDF, the data were not rescaled and were discarded.

Hereafter, PF CDF matching refers to the polynomial fitting CDF matching technique.

C. Trend and Seasonality Decomposition

For climate analysis and applications, it is essential that the CDF matching does not alter the long-term SM trends of the source time series. The dynamic range is subjected to change to become consistent with the reference data, but the trend and seasonality must remain the same between the source and rescaled time series. Hence, a time-series decomposition was performed on $AM2$ and $AM2_{NSL2}$. The time series were deseasonalized following an additive decomposition method explained in [59] and implemented in MATLAB under the name of *Seasonal Adjustment Using a Stable Seasonal Filter*. This method consists of a trend and seasonality parametric decomposition. A convolution is performed on the time series to extract the long-term trend. Then, the trend is subtracted from the original time series, and the detrended data are averaged on a monthly basis to provide an estimation of the seasonality.

D. Random Error and Temporal Frequency of the Observational Data

In contrast to models, for which errors are mostly of systematic nature, errors in observational data are mostly of random nature. In addition, while several global land surface models run with a few hours time steps, microwave sensors on-board polar-orbiting satellites have a revisit frequency of one to three days. Therefore, it is important to study the impact of the random errors and the revisit frequency when observational data are used as reference for the rescaling process. Both effects were studied with a numerical experiment using synthetic data. Starting with GLDAS time series, they were altered in two different ways to simulate the observational effects mentioned above.

1) The temporal frequency of $GLDAS$ time series was decreased from one observation per day to one observation every three days to simulate the usual revisit of satellites (hereafter, referred as $GLDAS_{f=1/3}$).
2) A Gaussian noise with mean $\mu = 0$ and sigma $\sigma = 0.02$ $\mathrm{m}^3/\mathrm{m}^3$ was added to the $GLDAS$ time series for each time step to simulate random errors in L-band data (hereafter, referred as $GLDAS_{\sigma_{0.02}}$). With a 3σ distribution of 0.06 $\mathrm{m}^3/\mathrm{m}^3$, this noise is representative of SM retrievals by the current generation of microwave sensors. However, Gaussian noise with a larger σ of 0.04 $\mathrm{m}^3/\mathrm{m}^3$ was also tested ($GLDAS_{\sigma_{0.04}}$). A total of 100 draws were computed both for $GLDAS_{\sigma_{0.02}}$) and $GLDAS_{\sigma_{0.04}}$) time series.

Finally, $AM2$ time series over North America were rescaled using the perturbed time series $GLDAS_{f=_{1/3,}}$ $GLDAS_{\sigma_{0.02}}$, and $GLDAS_{\sigma_{0.04}}$ as reference and compared to $AM2$ rescaled with the original $GLDAS$ ($AM2_{GLDAS}$) in terms of STDD. In all cases, the monthly PF CDF matching was used, and the periods considered for the $AM2$ and $GLDAS$ time series were those shown in Fig. 1.

IV. Results

This section describes the performances of $AM2$, the rescaled $AM2$, as well as the reference time series with respect to the *in situ* measurements from the networks shown in Table I. Since the results are similar for the C_1-, C_2-, and X-bands of $AM2$, only those of the C_1-band are shown here. The results obtained by using all the networks together are very consistent with those obtained for each network independently. Consequently, only the results considering all the networks at once are presented, and conclusions drawn in the next sections remain valid at a network scale. The performances according to the IGBP land cover classes and Köppen–Geiger climate classification (see Tables II and III) are also addressed in this section.

A. Assessment of the Source and Reference Data

Fig. 4 presents the performances of $AM2$ (source), $ESL3$, $ESLP$, $ESNRT$, $ESIC$, $NSLP$, $NSL2$, and $GLDAS$ with

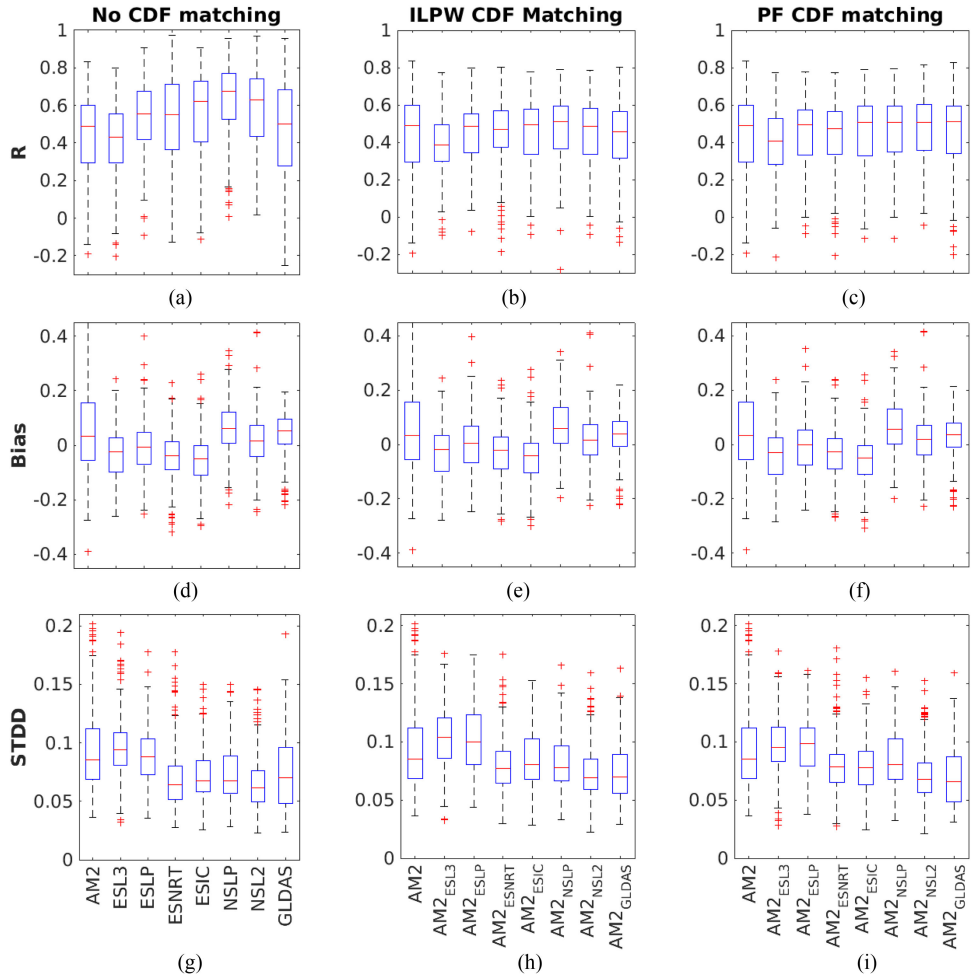

Fig. 4. Evaluation of the source, reference, and rescaled time series with respect to *in situ* measurements by considering all the networks at once. The Pearson correlation (R), bias (m^3/m^3), and STDD (in m^3/m^3) are computed for the 297 ISMN sensors locations and plotted as box plots. The first column presents the performances of $AM2$ C_1-band, $ESL3$, $ESLP$, $ESNRT$, $ESIC$, $NSLP$, $NSL2$, and $GLDAS$. The second and third columns present the performances of the rescaled $AM2$ using the $ILPW$ and PF CDF matching, respectively (see Sections III-B1 and III-B2). Acronyms of the datasets are explained in Sections II and III. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th (q_{25}) and 75th (q_{75}) percentiles, respectively. Points are considered as outliers if they are greater than $q_{75} + 1.5 \times (q_{75} - q_{25})$ or less than $q_{25} - 1.5 \times (q_{75} - q_{25})$. The whiskers extend to the most extreme data points not considered outliers. (a), (d), and (g) No CDF matching. (b), (e), and (h) ILPW CDF matching. (c), (f), and (i) PF CDF matching.

respect to *in situ* measurements through three statistical metrics: the Pearson correlation [R; see Fig. 4(a)], the bias [see Fig. 4(d)], and the STDD [see Fig. 4(g)]. As shown in Fig. 4(a), $ESIC$, $NSL2$, and $NSLP$ show the highest correlations among all the datasets, while $AM2$ and $ESL3$ give the lowest ones. $ESLP$, $ESNRT$, and $GLDAS$ show intermediate values. Fig. 4(d) shows that $ESLP$ gives the lowest bias, while $NSLP$ shows the highest one. In general, the SMOS datasets have negative bias with respect to *in situ* measurements, while $AM2$, $GLDAS$, $NSLP$, and $NSL2$ have positive bias. Regarding STDD, $AM2$, $ESL3$, $ESLP$, and $GLDAS$ show higher values than the other datasets [see Fig. 4(g)]. The STDDs are similar for $ESNRT$, $ESIC$, $NSLP$, and $NSL2$.

In Fig. 5, the results are analyzed as a function of the IGBP land cover classes that are the most represented by the ISMN sensors used in this article. Overall, the source and reference time series show correlations close to 0.5 or above except for the Evergreen Needleleaf Forest class [see Fig. 5(e)]. In addition, the performances of the reference time series are similar to or better than those of $AM2$ for almost all the classes. The correlations of the reference data are above 0.70 and significantly higher than those of $AM2$ for the Savannah class [see Fig. 5(c)]. Due to the change in the number of sensors available for the different classes (see Table II), results shown in Fig. 5(a)–(c) might be more significant than those shown in Fig. 5(d)–(f).

In Fig. 6, the results are analyzed as a function of the Köppen–Geiger climate classes that are the most represented by the ISMN sensors used in this article. All the source and reference time series show correlations close to 0.5 or above except for the two classes with Snow as main climate [see Fig. 6(e) and (f)], which are between about 0.25 and 0.5. Moreover, the performances of the reference time series are similar to or better than those of $AM2$ for almost all the classes. However, $ESL3$ and $ESIC$ show the lowest correlations in Fig. 5(e). The correlations of the reference data are above 0.70 and significantly higher than those of $AM2$ for the climate class shown in Fig. 6(c). After further

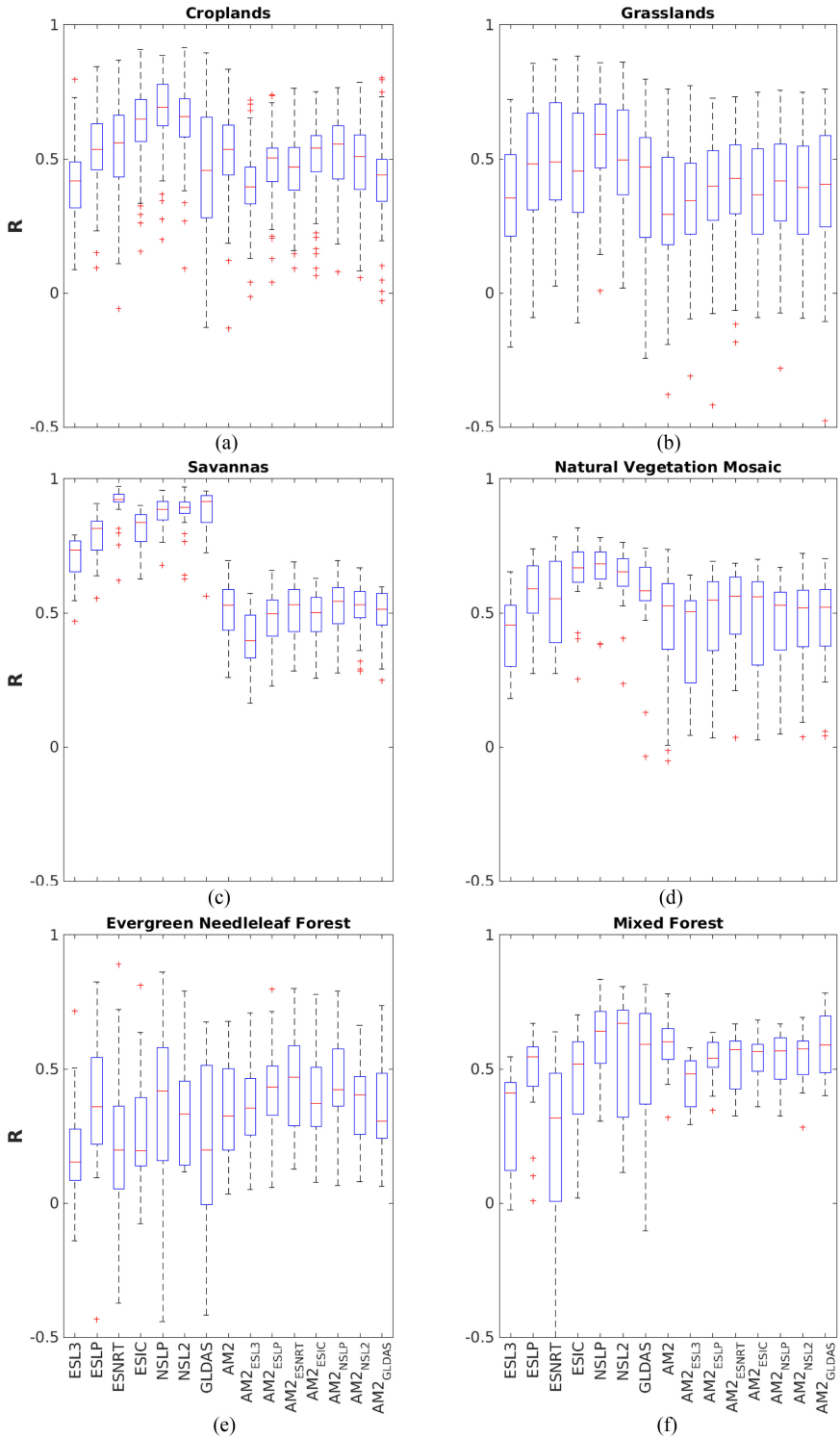

Fig. 5. (a)–(f) Evaluation of $AM2$ C_1-band, $ESL3$, $ESLP$, $ESNRT$, $ESIC$, $NSLP$, $NSL2$, $GLDAS$, as well as the rescaled $AM2$ with respect to *in situ* measurements according to the IGBP land cover classification (see Table II). The Pearson correlation (R) is shown as box plots for the six classes that are the most represented by the ISMN sensors locations used in this article. The $ILPW$ CDF matching is considered in this figure, and acronyms of the datasets are explained in Sections II and III. The structure of the box plots is described in the caption of Fig. 4.

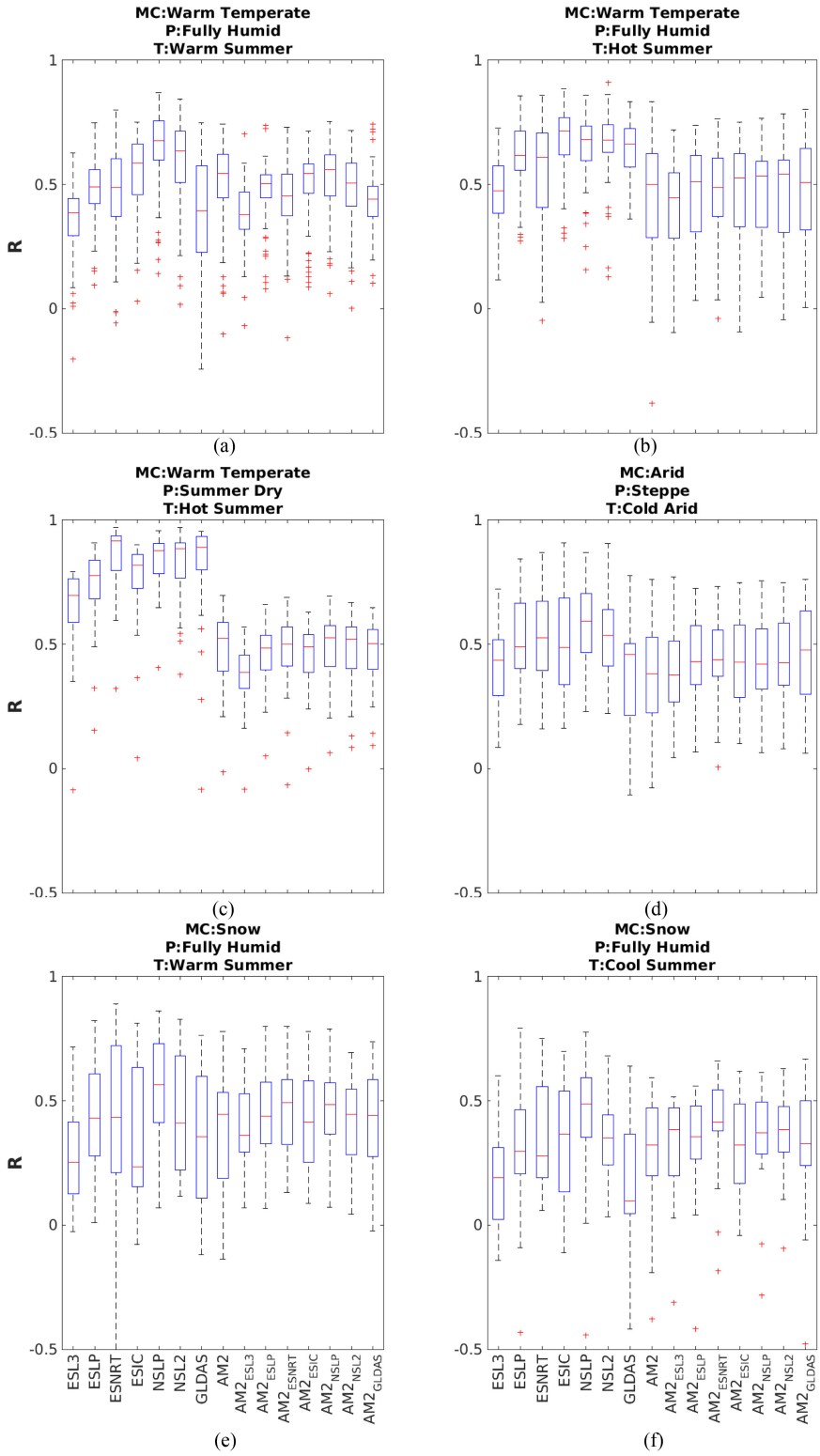

Fig. 6. (a)–(f) Evaluation of $AM2$ C_1-band, $ESL3$, $ESLP$, $ESNRT$, $ESIC$, $NSLP$, $NSL2$, $GLDAS$, as well as the rescaled $AM2$ with respect to *in situ* measurements according to the Köppen–Geiger climate classification (see Table III). The Pearson correlation (R) is shown as box plots for the six classes that are the most represented by the ISMN sensors locations used in this article. The $ILPW$ CDF matching is considered in this figure, and acronyms of the datasets are explained in Sections II and III. The structure of the box plots is described in the caption of Fig. 4.

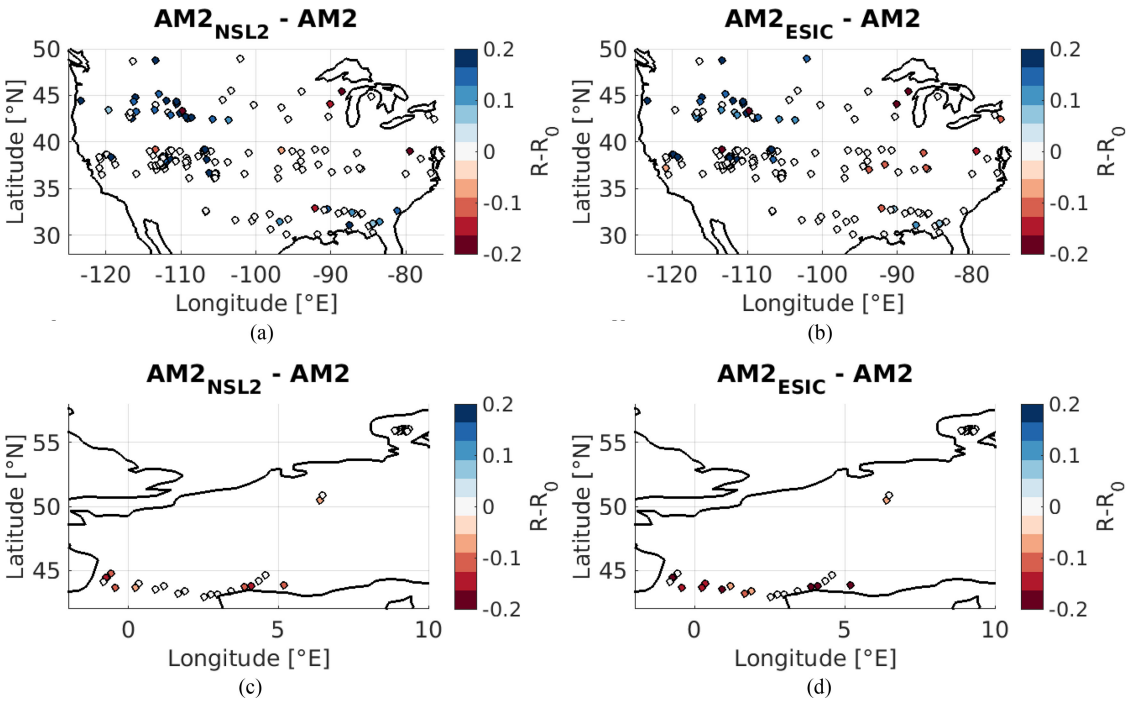

Fig. 7. (a)–(d) Different maps show the Pearson correlation with respect to *in situ* measurements of $AM2$ after rescaling (R) minus the Pearson correlation of $AM2$ (R_0). All the locations for which the difference in correlation is not significant are plotted in white. Left panels: $AM2_{NSL2}$. Right panels: $AM2_{ESIC}$. In both cases, the PF CDF matching at a monthly time scale was used (see Section III-B2). Acronyms of the datasets are explained in Sections II and III.

analysis, it was found that all the sensors locations representing the IGBP Savannas class [see Fig. 5(c)] are also included in this climate class. Taking into account the number of sensors available for each of those classes (see Table III), the results shown in Fig. 6(f) might be less significant than those shown in the other box plots.

B. Assessment of the Rescaled Data

1) CDF Matching Techniques and Time Scale Comparison: Fig. 4(b), (e), (h), and 4(c), (f), (i) present the performances of the rescaled $AM2$ with respect to *in situ* measurements using the $ILPW$ and PF CDF matching at a monthly time scale, respectively. The results of the CDF matching performed on a seasonal basis and using the full length of the time series are not presented here because the values of the different metrics do not change significantly according to the used time scale. Besides, it is interesting to note that there is almost no difference in terms of correlation, bias, and STDD between the two CDF matching approaches.

From Fig. 4(b) and (c), the correlation distributions of all the rescaled $AM2$ and $AM2$ are quite similar. Only $AM2_{ESL3}$ give slightly lower values. $AM2_{NSLP}$, $AM2_{NSL2}$, and $AM2_{GLDAS}$ show a positive bias, while $AM2_{ESL3}$, $AM2_{ESLP}$, $AM2_{ESNRT}$, and $AM2_{ESIC}$ give negative bias [see Fig. 4(e) and (f)], as expected from the results discussed in Section IV-A and plotted in Fig. 4(d). The STDD is a bit higher for $AM2_{ESL3}$ and $AM2_{ESLP}$ than for the other rescaled $AM2$, but, again, it was expected from the results in Fig. 4(g).

2) Effect of the Reference Dataset on the CDF Matching: The best performances of the rescaled time series with respect to the *in situ* measurements were found for $AM2_{ESNRT}$, $AM2_{ESLP}$, $AM2_{ESIC}$, $AM2_{NSLP}$, and $AM2_{NSL2}$, which show overall performances similar or somewhat better than those obtained for $AM2_{GLDAS}$ (see Fig. 4).

The same conclusion holds when the rescaled $AM2$ are analyzed as a function of the IGBP land cover and Köppen–Geiger climate classifications (see Figs. 5 and 6).

In addition to the box plots showing the distribution of different metrics for all the sites, the results were also analyzed site by site. Fig. 7 shows the differences between the correlation of $AM2_{ESIC}$ or $AM2_{NSL2}$ with respect to *in situ* measurements and those of the source $AM2$. The PF CDF matching at a monthly time scale was considered in this example, and all the differences that are not significant are plotted in white.

According to Fig. 7(a) and (b), there is an increase in correlation in the west part of the USA. However, a decrease in correlation is observed in the south of France [see Fig. 7(c) and (d)].

Taking into account the confidence intervals of the correlations, with respect to $AM2$, $AM2_{ESIC}$ ($AM2_{NSL2}$) shows a significant increase in correlation for 17% (18%) of the sites, while a significant decrease in correlation is observed for 9% (8%) of the locations. In summary, for most of the sites (74%), the correlation difference is not significant when the confidence intervals are taken into account. For the sites where the difference is significant, most of them (66–68%) show an increase in correlation with respect to the *in situ* measurements.

Most of the significant increases in correlation (54% for $AM2_{ESIC}$ and 50% for $AM2_{NSL2}$) are observed for the grasslands IGBP class (see Fig. 5). The significant decreases (42% for $AM2_{ESIC}$ and 38% for $AM2_{NSL2}$) mainly belong to the croplands class [see Fig. 5(a)]. Most of the significant increases

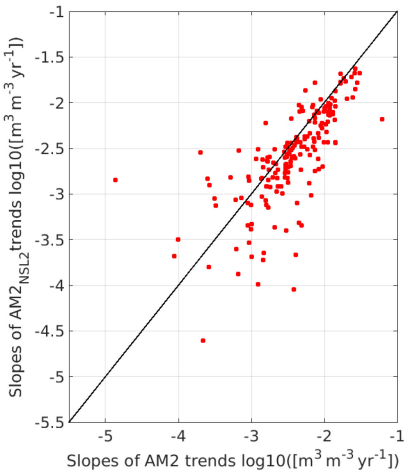

Fig. 8. Slopes of the trends of $AM2$ and $AM2_{NSL2}$ using the $ILPW$ CDF matching are plotted against each other for the 297 ISMN sensor locations (see Sections III-B1 and III-C). Acronyms of the datasets are explained in Sections II and III.

in correlation (26%, 22%, and 18% for $AM2_{ESIC}$ and 25%, 23%, and 17% for $AM2_{NSL2}$) are observed for the Köppen–Geiger classes shown in Fig. 5(b), (d), and (e), respectively. The significant decreases (15% and 38% for $AM2_{ESIC}$ and 25% and 21% for $AM2_{NSL2}$) mainly belong to the classes shown in Fig. 5(a) and (b), respectively.

3) Effect of the CDF Matching on the Time Series Trend: For applications such as building long time data records, the goal of the rescaling is just to perform a bias correction. It is important to ensure that possible trends in the source data are conserved in the rescaled data. Therefore, a trend and seasonality decomposition, as explained in Section III-C, was applied to $AM2$ and to one of the rescaled time series ($AM2_{NSL2}$) using the $ILPW$ CDF matching at a monthly scale. After the decomposition, a linear regression was computed to fit the trends. The slopes of the trends of the source and of the rescaled time series were compared site by site and plotted in logarithmic scale (see Fig. 8). The scatter plot shows that all the points are close to the 1:1 line, implying that trends are preserved. The bigger differences are found for sites that show very low trends, being very low in absolute value (less than 0.001 m $^3\cdot$ m$^{-3}\cdot$yr^{-1}) with respect to geophysical SM trends found in other studies [8], [14], [30], [60].

4) Comparison of the Source, Reference, and Rescaled Time Series at the Globe Scale: The first part of this article was devoted to comparing the source, reference, and rescaled time series only at locations for which independent *in situ* measurements are available. However, it is also necessary to get more insight into the rescaled $AMS2$ at a global scale. Therefore, a global rescaling of $AM2$ was done using $ESIC$ as the reference and applying both the $ILPW$ and PF CDF matching at a monthly time scale. Fig. 9(a) shows the bias of the source data with respect to the reference data. There is significant positive bias in the northern regions and a negative bias in the equatorial region. These biases are in agreement with those found between

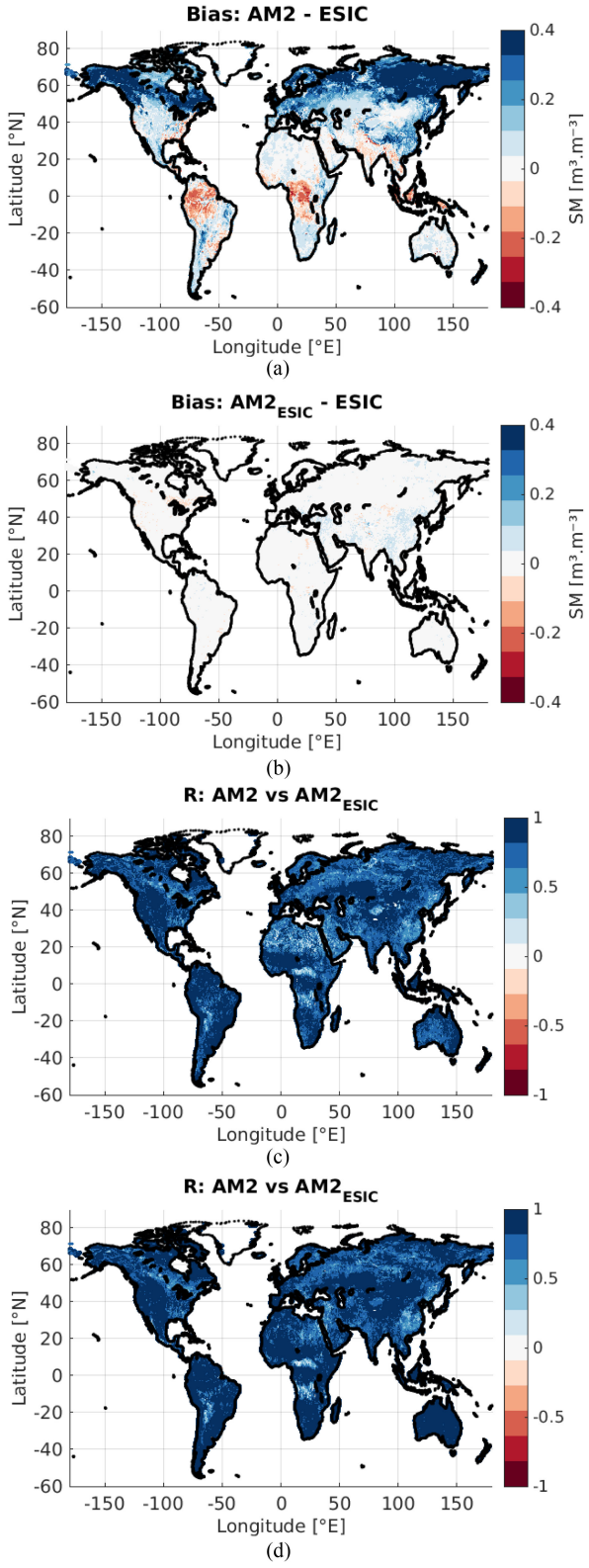

Fig. 9. Comparison of $AM2$, $ESIC$, and $AM2_{ESIC}$ at the scale of the Globe. Map (a) shows the bias between $AM2$ and $ESIC$. Map (b) shows the bias between $AM2_{ESIC}$ and $ESIC$. Maps (c) and (d) show the Pearson correlation between $AM2$ and $AM2_{ESIC}$. The $ILPW$ CDF matching was used to compute $AM2_{ESIC}$ in maps (b) and (c), while the PF CDF matching was used in map (d) (see Sections III-B1 and III-B2). Acronyms of the datasets are explained in Sections II and III.

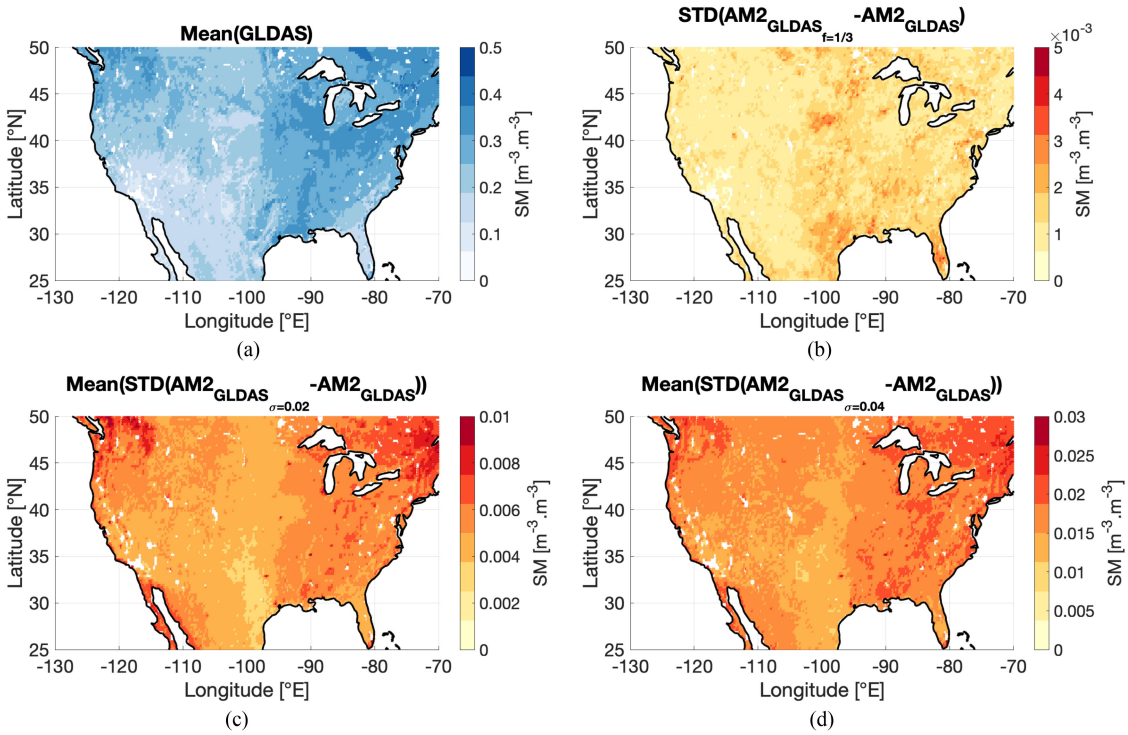

Fig. 10. (a)–(d) Impacts of using a reference dataset with a reduced temporal frequency and random errors for the CDF matching. Top left: average SM value of $GLDAS$. Top right: STDD between $AM2_{GLDAS_{f=1/3}}$ and $AM2_{GLDAS}$. Bottom left: Average of 100 STDD values between $GLDAS_{\sigma_{0.02}}$ and $AM2_{GLDAS}$. Bottom right: Average of 100 STDD values between $GLDAS_{\sigma_{0.04}}$ and $AM2_{GLDAS}$. Acronyms of the datasets are explained in Sections II and III.

ASMR-E and SMOS SM in other studies [61]. After the rescaling, the bias drops to ± 0.04 m $^3\cdot$ m^{-3} for more than 90% of the regions [see Fig. 9(b)], while the dynamics of the source data are preserved in the rescaled dataset as shown by the Pearson correlation maps [see Fig. 9(c) and (d)]. It is interesting to note that $AM2_{ESIC}$ are better correlated to $AM2$ in some regions such as Sahara and in Australia when the PF CDF matching is used for the rescaling. The map of the biases between $ESIC$ and $AM2_{ESIC}$ using the PF CDF matching is not shown here because it is quite similar to Fig. 9(b) ($ILPW$ CDF matching).

5) Impacts of the Temporal Sampling and Random Errors of Observational Data: The effect of random errors and of the reduced temporal sampling of observational data with respect to models on the rescaled time series was studied following the methodology presented in Section III-D. The numerical experiments were computed over the part of North America displayed in Fig. 10. Fig. 10(a) shows the mean SM in the 2010–2019 period. The effect of the temporal sampling of remotely sensed SM time series is shown in Fig. 10(b). The STDD between $AM2_{GLDAS_{f=1/3}}$ and $AM2_{GLDAS}$ is very low, below 0.005 m^3/m^3. This result is expected since the revisit time of microwave sensors used to estimate SM is studied to be able to capture the SM dynamics with a higher temporal frequency than the characteristic drying times after precipitation events. However, it is interesting to note that the regions in Fig. 10(b), with slightly higher STDD, correspond to sandy soil regions. This is also expected as the soil surface memory to precipitation events is shorter than for other types of soil. In summary, using observational time series as reference to compute the CDF matching does not affect significantly the results with respect to using a model time series with a denser temporal sampling.

Regarding the effect of random errors on the observational time series, they have been modeled as explained in Section III-D. The GLDAS time series were perturbed with Gaussian noise representative of the error of SM time series computed from remotely sensed data, with a σ representative of retrieved time series $GLDAS_{\sigma_{0.02}}$ and a higher value $GLDAS_{\sigma_{0.04}}$.

Fig. 10(c) shows the average of the STDD between $AM2_{GLDAS_{\sigma_{0.02}}}$ and $AM2_{GLDAS}$ for the 100 draws computed. A gradient is observed from West to East. The difference of the time series after rescaling against the noisy GLDAS time series is higher in wetter areas with more vegetation in the East (0.006–0.008 m^3/m^3) than in dry areas with less vegetation in the center and the West [0.002–0.004 m^3/m^3; see Fig. 10(a)]. The same behavior but with higher absolute values of the differences is found for $AM2_{GLDAS_{\sigma_{0.04}}}$ [see Fig. 10(d)]. The STDD values are from around 0.01 m^3/m^3 in the dry areas in the center and up to 0.025 m^3/m^3 in the West. Modeled and remotely sensed SM series are in better agreement in dry regions with low vegetation as shown by different evaluation studies [44], [61]–[64], and this is also the case for the rescaled time series even in the presence of noise in the reference data. For both σ values, the impact of adding noise to the GLDAS time series is larger than the impact of reducing the temporal sampling, and it becomes significant when the noise in the reference time series has a distribution with $\sigma = 0.04$ m^3/m^3. In this case, the

error introduced in the rescaled time series should be taken into account in the total error budget of the SM CDR.

V. Discussion

In contrast to models such as GLDAS, remote sensing retrievals can be relatively noisy and can present gaps due to a limited revisit frequency of the satellite and possibly due to low-quality retrievals that have been flagged out; this latter effect depends on the selected algorithm even when applied to the same sensor.

Therefore, in the context of removing model dependence from CDRs by using an observational dataset as the reference for the rescaling of other sensors time series, it is necessary to evaluate the time series rescaled using different L-band products.

Taking into account the different L-band SM datasets available, it has been shown in Section IV-A that, in general, SM time series computed from L-band data show better performances with respect to *in situ* measurements than $AM2$ time series. Overall, $NSL2$ and $NSLP$ followed by $ESLP$, $ESNRT$, and $ESIC$ show the best performances, similar to or better than $GLDAS$. The good results obtained for $ESLP$ confirm previous evaluations [39], [61], and the results for $ESNRT$ and $ESIC$ are in agreement with other studies [45], [46].

The $AM2$ time series rescaled using different L-band SM datasets were evaluated against *in situ* measurements (see Section IV-B). The best performances were found when the products used as the reference are those that show the best performances against *in situ* measurements. It is noteworthy that the $AM2$ time series rescaled using GLDAS also show good performances, confirming the pertinence of using this model within the ESA SM CCI [17]. However, removing all possible model dependence in CDRs is necessary for some applications such as data assimilation in climate models. In this context, van der Vliet *et al.* [65] discussed how to homogenize in flagging strategies in between sensors without using external datasets coming from models and van der Schalie *et al.* [66] discussed strategies to use the 36.5-GHz band to estimate soil temperature needed as input to estimate SM from lower frequencies without requiring model estimations. The current article shows that it is actually possible to use L-band data as the reference to rescale other remotely sensed time series instead of using the GLDAS model. This is also possible independently of the approach and time scale used for the CDF matching. Despite the differences between the time series used as the reference for the CDF matching such as the time coverage or the revisit frequency, several SMOS and SMAP datasets are legitimate candidates to replace the GLDAS model.

SMAP data have a larger spatial coverage than that of SMOS in regions affected by radio frequency interference. In contrast, SMOS time series are longer than those of SMAP by a factor of 2. Long time series are necessary to capture the long-term SM variability for climate applications. This is shown in Fig. 3, where the CDF of $ESIC$ computed from the date of SMAP launch (2015) was compared to the CDF of the full length of $ESIC$ (starting in 2010). Fig. 3 shows the median bias between the "short" and "long" $ESIC$ over the entire globe. Overall, the bias is mainly in a range between -0.01 and 0.01 $\mathrm{m}^3/\mathrm{m}^3$. However, some regions show high bias values reaching ± 0.05 $\mathrm{m}^3/\mathrm{m}^3$ such as in the Great Lakes area, Argentina, North India, China, etc. Moreover, the behavior of the two CDFs is also quite different at some locations, implying that four-year-long time series are not long enough to be representative of the SM dynamics in many places of the globe. Of course, as pointed out by other studies, the time length of the reference dataset used for the CDF matching should be as long as possible to be able to compute robust results (see [67]–[69]). Reichle and Koster [21] proposed that in case of short time series, an alternative could be to replace temporal variance by spatial variance by computing the CDF, including the data in a 2° circle around each grid point. This approach was tested here computing the CDF of SMOS time series since the date of the SMAP launch using all data in a 2° radius. Fig. 3 also shows the CDFs of the short time series using the spatial variability (green lines). Actually, the approach gives CDFs similar to those of the long SMOS time series in some locations (Points 1, 5, 7, 9, 10, and 12). However, for many of them, the CDF of the short time series was already in good agreement with the CDF of the long time series (Points 1, 5, 7, 9, and 12). Finally, the CDF using the spatial variance is very different to that of the long one for many points (Points 2, 3, 4, 8, 10, and 11). Therefore, this method cannot effectively provide more robust CDFs when the time series are short. Since both SMOS and SMAP presents pros and cons, an L-band fusion product may be the best option to serve as the reference to rescale other SM time series when building an SM CDR.

The ESA CCI retrieves SM observations from active and passive microwaves sensors since the end of the 1970s. Before being merged, these data are rescaled against the GLDAS model that has the advantage to provide time series covering this period of roughly 40 years. However, L-band data are only available since the end of 2009 with the launch of the SMOS mission. Thus, replacing the GLDAS model by an L-band sensor could be an issue for the rescaling of other sensors that operated before 2010. To make up for the lack of overlapping period between L-band data with several active and passive products, an iterative backward scaling approach could be considered. SM retrievals from sensors that overlap in time with SMOS and SMAP will be CDF matched against the L-band data. The scaling reference would be updated and include the newly scaled retrievals in addition to the L-band data. This reference would then be used for the rescaling of retrievals from older sensors that share a common period with it. Nevertheless, this iterative approach could lead to a flattening of the original dynamics and to a suboptimal error transfer and should be evaluated further in subsequent studies.

VI. Conclusion

Different approaches to rescale remotely sensed SM observations were discussed and evaluated in this article. The ESA CCI currently rescales different SM datasets by CDF matching using the GLDAS land surface model as the reference to produce long and consistent SM time series. A rescaling experiment was performed on the ASMR-2 data using different SMOS and SMAP

datasets as the reference to investigate the feasibility to replace the GLDAS model by an L-band sensor. Both the reference and rescaled time series were assessed against ground-based measurements of SM retrieved from the ISMN. In addition, a linear piecewise and polynomial fitting CDF matching were compared at different time scales (computing one single set of rescaling coefficients, computing four sets of seasonal coefficients, or computing 12 sets of monthly coefficients).

The results show that SMOS, SMAP, and GLDAS time series give better performances with respect to *in situ* measurements than those of AMSR2. The SMAP-LPRM time series get the highest correlations. The performances of the rescaled data do not change significantly depending on the time scale and the approach used for the CDF matching. In general, the rescaled time series using L-band data as the reference provide performances similar to or a slightly better than those rescaled using the GLDAS model. Through this experiment, SMAP-LPRM, SMAP-L2, SMOS-IC, SMOS-LPRM, and SMOS-NRT products provide the most promising L-band datasets to replace the GLDAS model as the reference to rescale other remotely sensed time series. However, an actual implementation of this approach to compute multi-instrument long time series of SM within projects such as the ESA CCI or the Copernicus Climate Change Service would require more validation and evaluation. More specifically, the lack of overlapping period between SMOS and SMAP with the other active and passive products, that are merged into the COMBINED ESA CCI SM product, should be addressed. Using an L-band dataset as the reference will also affect the triple collocation (TC) approach that it is currently used to assign errors to individual datasets to compute a weighted mean for the COMBINED product, as the products to be used as input for the TC will not be fully independent.

In addition, random errors of the scaling L-band reference should be properly taken into account since they can significantly affect the CDF matching process and alter the rescaled datasets.

Finally, the effect of the time series length on the CDF was also studied as SMAP time series are shorter by a factor of 2 than those from SMOS. The CDFs of SMOS time series were computed within the SMAP dates and compared to the CDF using the full length of the SMOS time series. Significant differences were found in some regions of the globe arguing that four-year long time series are not long enough to capture the entire SM dynamics globally. These results show that the length of the time series used as the reference for the rescaling could be a crucial issue for the CDF matching to be reliable. Replacing temporal by spatial variance in the short time series does not allow, in general, to compute CDFs closer to those of the long time series.

Future studies will be dedicated to the conception of a merged L-band dataset as the reference for the rescaling to benefit from both SMAP and SMOS advantages. This could be achieved by applying the same physical-based algorithm such as LRPM to SMOS and SMAP brightness temperatures after intercalibration or by merging two SM datasets with a machine learning algorithm. In addition, as mentioned above, the use of SMOS and SMAP data to rescale SM time series from older sensors, which do not share a common operation period, needs further investigation.

ACKNOWLEDGMENT

The authors would like to thank constructive reviews from the different referees concerning the impact of random errors on the rescaling process. Their insightful comments resulted in important improvements of this manuscript. R. Madelon and N. J. Rodríguez-Fernández would like to acknowledge interesting discussions on the model independence requirement for a soil moisture climate data record with J. C. Calvet, C. Albergel, M. Hirsch, and R. Kidd. This research made use of data from the Centre Aval de Traitement des Données SMOS operated for the Centre National d'Etudes Spatiales by the Institut Français de Recherche pour l'Exploitation de la Mer, Brest, France.

REFERENCES

[1] Global Climate Observing System, "Status of the global observing system for climate," Global Climate Observing Syst., World Meteorol. Org., Geneva, Switzerland, Rep. GCOS-195, 2015.

[2] R. D. Koster *et al.*, "Regions of strong coupling between soil moisture and precipitation," *Science*, vol. 305, no. 5687, pp. 1138–1140, 2004.

[3] M. Drusch, "Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's integrated forecast system and the TMI soil moisture data set," *J. Geophys. Res., Atmos.*, vol. 112, no. D3, pp. 1984–2012, 2007.

[4] P. De Rosnay, M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel, and L. Isaksen, "A simplified extended Kalman filter for the global operational soil moisture analysis at ECMWF," *Quart. J. Roy. Meteorol. Soc.*, vol. 139, no. 674, pp. 1199–1213, 2013.

[5] N. J. Rodríguez-Fernández *et al.*, "SMOS neural network soil moisture data assimilation in a land surface model and atmospheric impact," *Remote Sens.*, vol. 11, pp. 1334–1357, 2019. [Online]. Available: ttps://doi.org/10.3390/rs11111334

[6] M. Guerif and C. L. Duke, "Adjustment procedures of a crop model to the site specific characteristics of soil and crop using remote sensing data assimilation," *Agriculture Ecosyst. Environ.*, vol. 81, no. 1, pp. 57–69, Oct. 2000.

[7] J. Keyantash and J. A. Dracup, "The quantification of drought: An evaluation of drought indices," *Amer. Meteorol. Soc.*, vol. 83, pp. 1167–1180, 2002.

[8] M. Jung *et al.*, "Recent decline in the global land evapotranspiration trend due to limited moisture supply," *Nature*, vol. 467, no. 7318, pp. 951–954, 2010.

[9] L. Brocca *et al.*, "Improving runoff prediction through the assimilation of the ASCAT soil moisture product," *Hydrol. Earth Syst. Sci.*, vol. 14, pp. 1881–1893, 2010.

[10] H. Lievens *et al.*, "SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia," *Remote Sens. Environ.*, vol. 168, pp. 146–162, 2015.

[11] T. Lopez, A. Al Bitar, S. Biancamaria, A. Güntner, and A. Jäggi, "On the use of satellite remote sensing to detect floods and droughts at large scales," *Surv. Geophys.*, vol. 41, no. 6, pp. 1461–1487, 2020.

[12] S. Plummer, P. Lecomte, and M. Doherty, "The ESA Climate Change Initiative (CCI): A European contribution to the generation of the global climate observing system," *Remote Sens. Environ.*, vol. 203, pp. 2–8, 2017.

[13] Y. Liu *et al.*, "Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals," *Hydrol. Earth Syst. Sci.*, vol. 15, pp. 425–436, 2011.

[14] W. A. Dorigo *et al.*, "Evaluating global trends (1988–2010) in harmonized multi-satellite surface soil moisture," *Geophys. Res. Lett.*, vol. 39, 2012.

[15] W. Dorigo *et al.*, "Evaluation of the ESA CCI soil moisture product using ground-based observations," *Remote Sens. Environ.*, vol. 162, pp. 380–395, 2014.

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

[17] A. Gruber, T. Scanlon, R. van der Schalie, W. Wagner, and W. Dorigo, "Evolution of the ESA CCI soil moisture climate data records and their underlying merging methodology," *Earth Syst. Sci. Data*, vol. 11, no. 2, pp. 717–739, 2019.

[18] W. Preimesberger, T. Scanlon, C. H. Su, A. Gruber, and W. Dorigo, "Homogenization of structural breaks in the global ESA CCI soil moisture multisatellite climate data record," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 2845–2862, Apr. 2021.

[19] M. T. Yilmaz and W. T. Crow, "The optimality of potential rescaling approaches in land data assimilation," *J. Hydrometeorol.*, vol. 14, pp. 650–660, 2012.

[20] M. Afshar and M. Yilmaz, "The added utility of nonlinear methods compared to linear methods in rescaling soil moisture products," *Remote Sens. Environ.*, vol. 196, pp. 224–237, 2017.

[21] R. Reichle and R. Koster, "Bias reduction in short records of satellite soil moisture," *Geophys. Res. Lett.*, vol. 31, 2004.

[22] F. Aires, P. Weston, P. de Rosnay, and D. Fairbairn, "Statistical approaches to assimilate ASCAT soil moisture information: Part I–Methodologies and first assessment," *Quart. J. Roy. Meteorol. Soc.*, vol. 147, no. 18, pp. 2–31, 2019.

[23] L. Brocca *et al.*, "Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe," *Remote Sens. Environ.*, vol. 115, pp. 3390–3408, 2011.

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

[25] Y. Zeng *et al.*, "Analysis of current validation practices in Europe for space-based climate data records of essential climate variables," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 42, pp. 150–161, 2015.

[26] G. Flato *et al.*, "Evaluation of climate models," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the 5th Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, U.K.: Cambridge Univ. Press, 2014, pp. 741–866.

[27] Climate Modelling User Group, "Technical note on CMUG ECV soil moisture assessment report," ECMWF, ECMWF, U.K.; Meteo France, Paris, France; Max-Planck-Institut für Meteorologie, Hamburg, Germany, Met Office Medley Centre, Exeter, U.K., Tech. Rep. D3.1_v1B Version 0.5, 2013.

[28] M. G. Grillakis, A. G. Koutroulis, D. D. Alexakis, C. Polykretis, and I. N. Daliakopoulos, "Regionalizing root-zone soil moisture estimates from ESA CCI soil water index using machine learning and information on soil, vegetation, and climate," *Water Resour. Res.*, vol. 57, 2021, Art. no. e2020WR029249.

[29] J. Ikonen *et al.*, "Spatially distributed evaluation of ESA CCI soil moisture products in a northern boreal forest environment," *Geosciences*, vol. 8, no. 2, 2018, Art. no. 51.

[30] M. Piles, J. Ballabrera-Poy, and J. Muñoz-Sabater, "Dominant features of global surface soil moisture variability observed by the SMOS satellite," *Remote Sens.*, vol. 11, no. 1, 2019, Art. no. 95.

[31] N. Raoult *et al.*, "Confronting soil moisture dynamics from the ORCHIDEE land surface model with the ESA-CCI product: Perspectives for data assimilation," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1786.

[32] W. A. Lahoz and G. J. M. De Lannoy, "Closing the gaps in our knowledge of the hydrological cycle over land: Conceptual problems," *Surv. Geophys.*, vol. 35, no. 3, pp. 623–660, May 2014.

[33] R. van der Schalie *et al.*, "The effect of three different data fusion approaches on the quality of soil moisture retrievals from multiple passive microwave sensors," *Remote Sens.*, vol. 10, no. 1, 2018, Art. no. 107.

[34] S. Kim, Y. Y. Liu, F. M. Johnson, R. M. Parinussa, and A. Sharma, "A global comparison of alternate AMSR2 soil moisture products: Why do they differ," *Remote Sens. Environ.*, vol. 161, no. 10, pp. 43–62, 2015.

[35] K. Imaoka, M. Kachi, M. Kasahara, N. Ito, K. Nakagawa, and T. Oki, "Instrument performance and calibration of AMSR-E and AMSR2," *Remote Sens. Spatial Inf. Sci.*, vol. 38, no. 8, pp. 13–18, 2000.

[36] T. Scanlon *et al.*, "ESA climate change initiative plus—soil moisture. Algorithm theoretical baseline document (ATBD). Supporting product version 06.1," EODC, Vienna Austria; TU Wien, Vienna; VanderSat, Haarlem, The Netherlands; CESBIO, Toulouse, France; ETH Zürich, Zürich, Switzerland, Tech. Rep. D2.1 Version 2, 2021.

[37] T. Mo, B. Choudhury, T. Schmugge, J. Wang, and T. Jackson, "A model for microwave emission from vegetation-covered fields," *J. Geophys. Res., Oceans*, vol. 87, no. C13, pp. 11229–11237, 1982.

[38] H. Fujii, T. Koike, and K. Imaoka, "Improvement of the AMSR-E algorithm for soil moisture estimation by introducing a fractional vegetation coverage dataset derived from MODIS data," *J. Meteorol. Soc. Jpn.*, vol. 29, no. 11, pp. 282–292, 2009.

[39] R. van der Schalie *et al.*, "The merging of radiative transfer based surface soil moisture data from SMOS and AMSR-E," *Remote Sens. Environ.*, vol. 189, pp. 180–193, 2017.

[40] 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.

[41] A. Al Bitar *et al.*, "The global SMOS level 3 daily soil moisture and brightness temperature maps," *Earth Syst. Sci. Data*, vol. 9, pp. 293–315, 2017.

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

[43] M. J. Brodzik, B. Billingsley, T. Haran, B. Raup, and M. H. Savoie, "EASE-Grid 2.0: Incremental but significant improvements for earth-gridded data sets," *ISPRS Int. J. Geo-Inf.*, vol. 1, no. 1, pp. 32–45, 2012.

[44] R. van der Schalie, Y. Kerr, J. Wigneron, N. Rodríguez-Fernández, A. Al-Yaari, and R. de Jeu, "Global SMOS soil moisture retrievals from the land parameter retrieval model," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 45, pp. 125–134, 2016.

[45] N. J. Rodríguez-Fernández *et al.*, "SMOS near real time soil moisture product: Processor overview and first validation results," *Hydrol. Earth Syst. Sci.*, vol. 21, pp. 5201–5216, 2017.

[46] R. Fernandez-Moran *et al.*, "SMOS-IC: An alternative SMOS soil moisture and vegetation optical depth product," *Remote Sens.*, vol. 9, no. 5, 2017, Art. no. 457.

[47] 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.

[48] D. Entekhabi *et al.*, "SMAP Handbook–soil moisture active passive: mapping soil moisture and freeze/thaw from space," Jet Propulsion Lab., NASA, Pasadena, CA, USA, Tech. Rep., 2014.

[49] P. E. O'Neill, S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish, *SMAP L2 Radiometer Half-Orbit 36 km EASE-Grid Soil Moisture, Version 5*. Boulder, CO, USA: NASA Nat. Snow Ice Data Center Distrib. Active Arch. Center, 2018.

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

[51] M. Rodell *et al.*, "The global land data assimilation system," *Bull. Amer. Meteorol. Soc.*, vol. 85, no. 3, pp. 381–394, 2004.

[52] W. Dorigo *et al.*, "The international soil moisture network: A data hosting facility for global in situ soil moisture measurements," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 5, pp. 1675–1698, 2011.

[53] W. Dorigo, I. Himmelbauer, and D. Aberer, "The international soil moisture network: Serving earth system science for over a decade," *Hydrol. Earth Syst. Sci.*, vol. 25, pp. 5749–5804, 2021.

[54] T. R. Loveland *et al.*, "Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data," *Int. J. Remote Sens.*, vol. 21, no. 6/7, pp. 1303–1330, 2000.

[55] P. D. Broxton, X. Zeng, D. Sulla-Menashe, and P. Troch, "A global land cover climatology using MODIS data," *J. Appl. Meteorol. Climatol.*, vol. 53, no. 6, pp. 1593–1605, 2014.

[56] M. Kottek, J. Grieser, C. Beck, B. Rudolf, and F. Rubel, "World map of the Köppen-Geiger climate classification updated," *Meteorologische Zeitschrift*, vol. 15, no. 3, pp. 259–263, 2006.

[57] M. Drusch, E. Wood, and H. Gao, "Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture," *Geophys. Res. Lett.*, vol. 32, 2005.

[58] L. Moesinger *et al.*, "The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA)," *Earth Syst. Sci. Data*, vol. 12, no. 20, pp. 177–196, 2020.

[59] P. Brockwell and R. Davis, *Introduction to Time Series and Forecasting*, 2nd ed. New York, NY, USA: Springer, 2002.

[60] C. Albergel *et al.*, "Skill and global trend analysis of soil moisture from reanalyses and microwave remote sensing," *J. Hydrometeorol.*, vol. 14, no. 4, pp. 1259–1277, 2013.

[61] N. J. Rodríguez-Fernández *et al.*, "Long term global surface soil moisture fields using an SMOS-trained neural network applied to AMSR-E data," *Remote Sens.*, vol. 8, no. 11, 2016, Art. no. 959.

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

[63] Y. H. Kerr *et al.*, "Overview of SMOS performance in terms of global soil moisture monitoring after six years in operation," *Remote Sens. Environ.*, vol. 180, pp. 40–63, 2016.

[64] C. Albergel, G. Balsamo, P. d. Rosnay, J. Muñoz-Sabater, and S. Boussetta, "A bare ground evaporation revision in the ECMWF land-surface scheme: Evaluation of its impact using ground soil moisture and satellite microwave data," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 10, pp. 3607–3620, 2012.

[65] M. van der Vliet *et al.*, "Reconciling flagging strategies for multi-sensor satellite soil moisture climate data records," *Remote Sens.*, vol. 12, no. 20, 2020, Art. no. 3439.

[66] R. van der Schalie *et al.*, "L-band soil moisture retrievals using microwave based temperature and filtering. Towards model-independent climate data records," *Remote Sens.*, vol. 13, no. 13, 2021, Art. no. 2480.
[67] C. Draper, J.-F. Mahfouf, and J. Walker, "An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme," *J. Geophys. Res., Atmos.*, vol. 114, no. D20, pp. 6715–6752, 2009.
[68] Y. Y. Liu, R. A. M. De Jeu, M. F. McCabe, J. P. Evans, and A. I. J. M. van Dijk, "Global long-term passive microwave satellite-based retrievals of vegetation optical depth," *Geophys. Res. Lett.*, vol. 38, pp. 3439–3462, 2011.
[69] C. B. Blankenship, J. L. Case, B. T. Zavodsky, and W. L. Crosson, "Assimilation of SMOS retrievals in the land information system," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6320–6332, Nov. 2016.
[70] A. Al-Yaari *et al.*, "The AQUI soil moisture network for satellite microwave remote sensing validation in South-Western France," *Remote Sens.*, vol. 10, 2018. Art. no. 1839.
[71] S. Bircher, J. E. Balling, N. Skou, and Y. H. Kerr, "Validation of SMOS brightness temperatures during the HOBE airborne campaign, Western Denmark," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1468–1482, Jan. 2012.
[72] E. C. Osenga, J. C. Arnott, K. A. Endsley, and J. W. Katzenberger, "Bioclimatic and soil moisture monitoring across elevation in a mountain watershed: Opportunities for research and resource management," *Water Resour. Res.*, vol. 55, pp. 2493–2503, 2019.
[73] G. Schaefer, M. Cosh, and T. Jackson, "The USDA natural resources conservation service soil climate analysis network (SCAN)," *J. Atmos. Ocean. Technol.*, vol. 24, no. 12, pp. 2073–2077, 2007.
[74] J.-C. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, and B. Piguet, "In situ soil moisture observations for the CAL/VAL of SMOS: The SMOSMANIA network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 1196–1199.
[75] G. Leavesley *et al.*, "A modeling framework for improved agricultural water supply forecasting," in *Proc. AGU Fall Meeting Abstr.*, 2008, vol. 1, Art. no. 0497.
[76] M. Moghaddam *et al.*, "A wireless soil moisture smart sensor web using physics-based optimal control: Concept and initial demonstrations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 4, pp. 522–535, Dec. 2010.
[77] M. Moghaddam *et al.*, "Soil moisture profiles and temperature data from SoilSCAPE sites, USA," Data Set Version: 1, Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, TN, USA, 2016.
[78] S. Zacharias *et al.*, "A network of terrestrial environmental observatories in Germany," *Vadose Zone J.*, vol. 10, no. 3, pp. 955–973, 2011.
[79] J. Bell *et al.*, "Climate reference network soil moisture and temperature observations," *J. Hydrometeorol.*, vol. 14, pp. 977–988, 2013.

Rémi Madelon received the Engineering degree in energy systems from the École Polytechnique de l'Université d'Orléans, Orléans, France, in 2018.

From 2018 to 2019, he was a Research Engineer with the Institut de Physique du Globe de Paris, Paris, France, for the ESA SWARM mission. In 2020, he joined the Centre d'Etudes Spatiales de la Biosphère, Toulouse, France, within the Soil Moisture and Ocean Salinity Team. His research interests include soil moisture, remote sensing from space, and machine learning.

Nemesio J. Rodríguez-Fernández received the Licentiate degree in fundamental physics and the Ph.D. degree in astrophysics from the Universidad Complutense de Madrid, Madrid, Spain, in 1996 and 2002, respectively.

From 2002 to 2004, he was a Marie Curie Fellow with the Observatoire de Paris, Paris, France. From 2004 to 2006, he was a Teaching Assistant with Paris Diderot University, Paris, and the University of Bordeaux 1, Bordeaux, France. From 2006 to 2011, he was an Astronomer with the Institut de Radio Astronomie Milimetrique, Grenoble, France. Since 2012, he has been a Centre National de la Recherche Scientifique Scientist and Research Engineer with the Centre d'Etudes Spatiales de la Biosphère, Toulouse, France, with the exception of 11 months in 2015–2016 when he joined the European Center for Medium Range Weather Forecast (ECMWF), Reading, U.K., as a Research Scientist. He designed the neural network algorithm of the European Space Agency's Soil Moisture and Ocean Salinity (SMOS) soil moisture near-real-time product and the one that it is used for operational data assimilation of SMOS soil moisture at ECMWF. He is also working in aperture synthesis algorithms and new satellites mission design (SMOS-High Resolution). His current research interests include microwave remote sensing for Earth observation, in particular machine learning approaches for the estimation of geophysical parameters, such as soil moisture and biomass and their assimilation into numerical weather prediction models.

Robin van der Schalie received the Ph.D. degree in remote sensing and hydrology from the Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, in 2017.

His Ph.D. research focused on the integration of soil moisture from European Space Agency (ESA)'s Soil Moisture and Ocean Salinity (SMOS) mission into long-term climate data records. After this, he continued doing research on algorithm improvements behind the retrieval of land surface parameters derived from passive microwave observations in frequencies ranging from 1.4 GHz (L-band) to 37 GHz (Ka-band). He is currently a Senior Remote Sensing Scientist with VanderSat, where he is responsible for the continued contribution to ESA's Climate Change Initiative and the Copernicus Climate Change Services. He has experience working with state-of-the-art L-band missions, such as SMOS and Soil Moisture Active Passive, and less frequently used sensors such as SMMR, FengYun-3, and WindSat. He also focuses on R&D activities around the integration of EO into a wide range of applications, for example, on utilizing passive microwave-based datasets for large-scale crop yield predictions.

Tracy Scanlon received the B.Sc. degree in physics from the University of Bristol, Bristol, U.K., in 2007, and the M.Sc. degree in remote sensing from Cranfield University, Cranfield, U.K., in 2014. She is currently working toward the Ph.D. degree in quality control remote sensing soil moisture products with the Vienna University of Technology, Vienna, Austria.

She is working on the ESA CCI SM and C3S SM products with the Vienna University of Technology. Her research interests include the development of improved soil moisture product merging algorithms and the quantification of uncertainties in such products.

Ahmad Al Bitar received the B.A. degree from the Lebanese University, Beirut, Lebanon, in 2001 and the M.E. degree from the Institut National des Sciences Appliquées, Lyon, France, in 2003, both in civil engineering, and the Ph.D. degree in hydrogeology from the Institut National Polythechnique de Toulouse, Toulouse, France, in 2006.

His thesis work focused on numerical modeling in stochastic porous media in the context of surface/subsuface coupling. In 2006, he joined the Centre d'Etudes Spatiales de la Biosphère (CESBIO), Toulouse, as a Research Fellow on integrated hydrological modeling. Since 2008, he has been with the Soil Moisture and Ocean Salinity (SMOS) mission team with CESBIO. He contributed to the SMOS commissioning phase, the calibration/validation activities, and the definition of L3 multiorbit SMOS products. He is working on high-end level-4 products from the SMOS mission. His main research interests include assimilation of remote sensing data more specifically from microwave sensors into integrated hydrological modeling.

Yann H. Kerr (Fellow, IEEE) received the Engineering degree in telecommunication and radar technologies from Ecole Nationale Suprieure de l'Aeronautique et de l'Espace, Toulouse, France, in 1977, the M.Sc. degree in electrical and electronic engineering from Glasgow University, Glasgow, U.K., in 1981, and the Ph.D. degree in astrophysics, geophysics and spatial techniques from University Paul Sabatier, Toulouse, in 1992.

From 1980 to 1985, he was with the Centre National d'Etudes Spatiales, Paris, France. In 1985, he joined LERTS, where he was the Director in 1993 and 1994. He spent 19 months with the Jet Propulsion Laboratory, Pasadena, CA, USA, in 1987 and 1988. Since 1995, he has been with the Centre d'Etudes Spatiales de la Biosphère, Toulouse, France, where he has been the Deputy Director and Director since 2007. He is involved with many space missions. He was an EOS Principal Investigator (interdisciplinary investigations) and Principal Investigator and precursor of the use of the SCAT over land. In 1990, he started to work on the interferometric concept applied to passive microwave earth observation and was subsequently the science lead on the MIRAS project for the European Space Agency (ESA) with Matra Marconi Space and Observatoire Midi Pyrénées. He was also a Co-Investigator on IRIS, OSIRIS, and HYDROS for NASA. He was a Science Advisor for MIMR and Coinvestigator on AMSR. In 1997, he first proposed the natural outcome of the previous MIRAS work with what was to become the SMOS Mission, which was eventually selected by ESA in 1999 with him as the SMOS mission Lead Investigator and Chair of the Science Advisory Group. He is also in-charge of the SMOS science activities coordination in France. He has organized all the SMOS Science workshops. His research interests include the theory and techniques for microwave and thermal infrared remote sensing of the earth, with emphasis on hydrology, water resources management, and vegetation monitoring.

Richard de Jeu received the M.S. degree in environmental hydrology and the Ph.D. degree in environmental hydrology from Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, in 1996 and 2003, respectively.

From 1997 to 1999, he was a Research Assistant with Vrije Universiteit Amsterdam. From 1999 to 2001, he was a Research Assistant with Goddard Space Flight Center, National Aeronautics and Space Administration (NASA), Greenbelt, MD, USA. Over the years, he has successfully developed several satellite-derived data products, including soil moisture and evaporation. These products are widely used and accessible through the official European Space Agency and NASA data portals. He is involved in many projects supported by the Dutch Organization for Scientific Research, European Framework, and the European and American Space Agencies. He is a Managing Director with Transmissivity B.V./VanderSat B.V., Noordwijk, The Netherlands. His research interests include passive microwave radiometry and the use of this technique for hydrological applications.

Wouter Dorigo received the M.Sc. degree in physical geography with an emphasis on remote sensing from Utrecht University, Utrecht, The Netherlands, in 2000, and the Ph.D. degree in remote sensing from the Technical University of Munich, Munich, Germany, in 2008.

From 2002 to 2006, he was with Imaging Spectroscopy Group, German Remote Sensing Data Center, German Aerospace Center, Oberpfaffenhofen, Germany. Since 2007, he has been a Researcher, a Lecturer, and a Scientific Project Coordinator with Research Group Remote Sensing, Department of Geodesy and Geoinformation, Vienna University of Technology (TU Wien), Vienna, Austria, where he is currently a Professor of climate and environmental remote sensing. He has authored or coauthored more than 60 publications in refereed journal articles and conference proceedings. Using earth observation data, he tries to understand and quantify the dynamics and interactions of vegetation and the water cycle in a changing climate. His research interests include remote sensing of soil moisture and vegetation, with a focus on geophysical parameter retrieval through physical and semiempirical methods, calibration and validation, process automation, up- and downscaling, image classification, data merging, and time-series analysis.

Dr. Dorigo was a recipient of the 2015 Science Award of TU Wien for the water cycle in a changing climate.