THE IMPROVEMENT OF SCIENCE QUALITY AND UTILITY OF SOIL MOISTURE ESTIMATIONS FROM SATELLITE-BASED PASSIVE MICROWAVE REMOTE SENSING

by

Runze Zhang

Bachelor of Science
China Agricultural University, 2014

Master of Engineering
University of New South Wales, 2018

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Department of Engineering Systems and Environment

School of Engineering and Applied Science

University of Virginia

2023

Accepted by:

Venkataraman Lakshmi, Thesis Supervisor

James Smith, Committee Chair

Steven Chan, Committee Member

Julianne Quinn, Committee Member

Xi Yang, Committee Member

Jonathan Goodall, Committee Member

Acknowledgement

I wish to express my heartfelt gratitude to my advisor, Dr. Venkataraman Lakshmi, for his invaluable mentorship and support throughout my doctoral studies. His unwavering support and encouragement helped me to stay focused and motivated in achieving my research goals.

I am deeply grateful to Dr. Steven Chan for his invaluable insights and rigorous academic attitude, which were critical to the success of my research and advanced my understanding of a qualified research scientist. I would like to extend my sincere appreciation to Dr. Seokhyeon Kim who played an instrumental role in shaping my interests and guiding me towards this moment. I would also like to thank Dr. Ashish Sharma, Dr. Rajat Bindlish, and the members of my dissertation committee, Dr. James Smith, Dr. Julianne Quinn, Dr. Jonathan Goodall, and Dr. Xi Yang, for their feedback and constructive criticism that helped me refine my research and enhance the quality of my dissertation.

I also wish to acknowledge the generous support and intelligent advice from Dr. Bin Fang. In addition, I want to thank my lab colleagues and friends, who have provided me with a vibrant and collaborative academic community. Their insightful comments, constructive feedback, and stimulating discussions have enriched my research and broadened my perspectives.

I want to express my deep appreciation to my parents, Mrs. Guangxiang Zhang, and Mr. Xingqing Liu, who have been my steadfast source of support throughout my life. Their unwavering love, encouragement, and sacrifice have been the bedrock of my success. I could not have accomplished this without them, and I am forever grateful for their unparallel dedication to my well-being and achievements. Last but not least, I am deeply grateful to my fiancée, Ms. Nguyễn Thị Thiên Thanh, whose unwavering love, devotion, and patience have sustained me

through the many challenges and obstacles of this journey. Her unwavering belief in me, her support, and her understanding have been a constant source of inspiration and motivation. I am grateful for her companionship, her encouragement, and her profound impact on my life.

Abstract

Soil moisture is an important measure of the exchange of water and energy between the land and the atmosphere. Passive microwave remote sensors onboard Earth observation satellites have served as the most promising tool for quantifying the water content stored in the surface soil layer at a quasi-global scale. In recent decades, the utilization of satellite-based soil moisture retrievals has benefited a variety of applications, including the detection of extreme climate events, water resource and irrigation management, and numerical weather predictions. However, the available passive microwave soil moisture datasets are still unable to entirely satisfy the needs of climatological studies and applications due to the insufficient retrieval quality over certain land surface conditions, the coarse spatiotemporal resolution, the absence of information associated with nonuniform vertical moisture gradients, and the unavailability of a consistent long-term satellite-based soil moisture data record. Given these limitations, this dissertation aims to enhance the quality and utility of current passive microwave-based surface soil moisture data by attempting to resolve the above-mentioned issues.

This dissertation first outlines the analyses for soil moisture retrievals that are impacted by water bodies (i.e., lakes) and soil organic matter to provide clues for refining the operational algorithm of deriving soil moisture from observed brightness temperatures at L-band (1.41 GHz). Specifically, this dissertation identified the lake mix-layer temperature from ERA5 Land and the dielectric mixing model of Mironov 2019 as the preferred options to mitigate water contamination and the effects of SOM in the passive microwave remote sensing retrieval of soil moisture. Their utilization will greatly improve the accuracy of the next-generation L-band soil moisture dataset.

Subsequently, the thesis describes the study that fills temporal gaps in Soil Moisture Active Passive (SMAP) data over the conterminous United States (CONUS) by incorporating a satellite-based precipitation product and a data-driven loss function approach. Validation of this SMAP-based 12-hourly soil moisture product not only exhibited great accuracy but also successfully captured most soil moisture peaks caused by heavy rainfall. The proposed loss-function approach could quantitatively characterize local-scale hydrologic losses near the land surface and can be used for back-extension and forecasts of soil moisture estimates through the incorporation of precipitation measurements.

After that, a global-scale comparison of three advanced satellite-based products was conducted to identify their relative strengths in capturing temporal variability of regional-scale soil moisture. As a result, a global complementarity of the areas was observed where each satellite-based soil moisture product showed its respective advantage in capturing soil moisture variations. Such an evaluation can serve as a guideline for data users to select proper soil moisture datasets for their research and applications. In the appendix (p.141-164), the formulations of a layered radiative transfer model have been presented, which inversions can be used to infer vertically heterogeneous moisture profiles from passive microwave observations.

In summary, this dissertation is dedicated to improving the scientific quality and utility of state-of-the-art SMAP soil moisture retrievals by addressing several identified drawbacks. The outcomes of this dissertation hold great promise for changing how radiometer observations are interpreted in the future while the newly yielded soil moisture data with temporal continuity and higher accuracy will help researchers quantitatively understand the linkages between water balance components and deepen the understanding of the terrestrial-atmosphere interactions in the context of climate change.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

Soil moisture is a vital component of Earth's system as it plays a crucial role in controlling energy fluxes at the land-atmosphere interface and in regulating precipitation into surface runoff and storage as part of the water budget (Koster et al. 2004; Petropoulos 2013; Wagner et al. 2003). Of late, soil moisture has been acknowledged as an Essential Climate Variable (ECV) by the Global Climate Observing System (GCOS) (Dorigo et al. 2015). An accurate understanding of soil moisture distribution and dynamics could advance our comprehension of the water, energy, and carbon cycles. Additionally, detailed knowledge about the state of soil moisture, including its spatial and temporal dynamics, is of exceptionally beneficial for various meteorological, climatological, and ecohydrological applications, such as precipitation estimations, agricultural management, flood prediction, drought monitoring, and improved weather forecast skills (Koster et al. 2010; Robock and Li 2006; Seneviratne et al. 2010; Vereecken et al. 2014).

In general, gravimetric weighing of soil samples and installation of in-situ electromagnetic probes are regarded as the most reliable means to quantify soil water content (Babaeian et al. 2019). However, in-situ measurements can be subject to uncertainties due to a variety of factors, such as high salinity level and inadequate calibration (Babaeian et al. 2019). In addition, implementation of such methods is invasive (e.g., requiring the extraction of soil samples), costly and laborious. Moreover, in-situ sensors typically provide only point-scale measurements and may not accurately represent soil moisture variations over a larger area. As a result, they are often limited to small-scale experiments or local monitoring sites, making it challenging to obtain a comprehensive view of soil moisture dynamics over large areas.

Remote sensing techniques that adopt sensors mounted on aircraft and satellite platforms have been identified as a feasible alternative to quantifying and mapping soil moisture across the globe (Petropoulos et al. 2015). Compared to sensors operated in visible and infrared spectra, microwave remote sensing has several remarkable advantages, including the ability to detect changes in dielectric properties of the soil-water medium, the negligible influence of the atmosphere, and independence from solar illumination (Jackson 1993; Karthikeyan et al. 2017). As such, a variety of spaceborne passive and active microwave sensors have been developed and utilized to analyze soil moisture patterns at various scales (Karthikeyan et al. 2017). Passive sensors (radiometer) detect the soil moisture variations by measuring the naturally thermal emissions from the land surface in the form of brightness temperature. On the other hand, active microwave sensors (radar) capture the backscattered coefficients from the Earth's surface. Over the sparsely vegetated areas, soil moisture retrievals from passive microwave remote sensing are prone to higher accuracy relative to those from the observations of active sensors as the former measurements are of less disturbance from the surface roughness. In practice, the exceptional performance of soil moisture retrievals from the L-band (0.5 – 1.5 GHz) spaceborne microwave radiometers has been confirmed by a number of validation studies, e.g., Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP) (Al-Yaari et al. 2017; Chan et al. 2018; Chan et al. 2016).

However, the available passive microwave soil moisture datasets are still unable to entirely satisfy the needs of climatological studies and applications due to 1) the insufficient retrieval quality over particular land surface conditions, 2) the absence of information associated with vertical moisture gradients, 3) the coarse spatiotemporal resolution, and 4) the unavailability of a consistent long-term satellite-based soil moisture data record. Specifically, there are various

regions diagnosed as unfavorable for retrieving soil moisture. Densely vegetated areas, for example, could significantly attenuate the signals transmitting from the underlying soil, causing a severe deviation from the true soil moisture magnitude (Karthikeyan et al. 2017). In addition to those naturally adverse conditions, anthropogenic processes like radio frequency interference (RFI) could also contaminate the radiometer measurements and degrade the retrieval quality (Kerr et al. 2012). Notably, soil moisture data derived from the SMOS and SMAP missions are assumed to be an average moisture content within a 5-cm soil layer (O'Neill et al. 2021a). Although such an assumption could largely simplify the retrieval procedure, the alignment of the assumed and actual emission depths, and the applicability of this 5cm-depth assumption over a global range remain to be further investigated. Moreover, the constant soil moisture over the tens-of-kilometers scale sometimes cannot fully reflect the spatial variability of fine-scale soil wetness, while a revisit frequency of 2 – 3 days might fail to record the fluctuation of surface soil moisture responses to short-term intensive precipitation events (Peng et al. 2021; Zhang et al. 2021a). Furthermore, the period of available passive microwave remote sensing soil moisture datasets from the mentioned missions is shorter than 15 years, thereby not meeting the needs of climatology studies and related applications that require consistent long-term soil moisture observations.

Therefore, the objective of this dissertation is three-fold: 1) to advance and improve the current retrieval algorithm to enhance the accuracy of soil moisture estimations, 2) to increase the temporal availability of soil water variations into a 12-hourly scale, and 3) to investigate and construct a layered radiative transfer model capable of accurately modeling the incoherent microwave emissions from the natural soil media, which is characterized by heterogenous moisture and temperature profiles. These objectives will result in soil moisture retrievals that

more accurate, complete and immune to certain confounding factors (e.g., soil organic matter). Beyond the water balance perspective, the outcomes from this thesis will build the foundation for advancing the interpretation of future radiometer observations, and deepen the understanding of the linkages between water, energy, and carbon fluxes in the context of climate change. It should be noted that the scope of this thesis is primarily focused on the state-of-the-art soil moisture retrievals derived from the National Aeronautics and Space Administration (NASA) SMAP mission (Entekhabi et al. 2010a).

The structure of this dissertation is organized as follows. Works related to the water contamination and the degraded performance over organic-rich soils of the existing SMAP soil moisture retrievals have been investigated in Section 2 and Section 3. Section 4 illustrates the topic of temporally gap-filling SMAP soil moisture product through water balance budgeting. Section 5 describes the construction of a layered radiative transfer model for use in non-uniform subsurface soil medium. In Section 6, the adequacy of three advanced satellite-based soil moisture products in capturing temporal dynamics is evaluated and discussed. Finally, conclusions followed by a summary present in Section 7.

Chapter 2. Evaluation of Global Surface Water Temperature Data Sets for Use in Passive Remote Sensing of Soil Moisture

R. Zhang, S. Chan, R. Bindlish, and V. Lakshmi (2021). Evaluation of Global Surface Water Temperature Data Sets for Use in Passive Remote Sensing of Soil Moisture. *Remote Sens.* **2021**, 13, 1872. https://doi.org/10.3390/rs13101872

Abstract

Inland open water bodies often pose a systematic error source in the passive remote sensing retrievals of soil moisture. Water temperature is a necessary variable used to compute water emissions that is required to be subtracted from satellite observation to yield actual emissions from the land portion, which in turn generates accurate soil moisture retrievals. Therefore, overestimation of soil moisture can often be corrected using concurrent water temperature data in the overall mitigation procedure. In recent years, several data sets of lake water temperature have become available, but their specifications and accuracy have rarely been investigated in the context of passive soil moisture remote sensing on a global scale. For this reason, three lake temperature products were evaluated against in-situ measurements from 2007 to 2011. The data sets include the lake surface water temperature (LSWT) from Global Observatory of Lake Responses to Environmental Change (GloboLakes), the Copernicus Global Land Operations Cryosphere and Water (C-GLOPS), as well as the lake mix-layer temperature (LMLT) from the European Centers for Medium-Range Weather Forecast (ECMWF) ERA5 Land Reanalysis. GloboLakes, C-GLOPS, and ERA5 Land have overall comparable performance with Pearson correlations (R) of 0.87, 0.92 and 0.88 in comparison with in-situ measurements. LSWT products exhibit negative median biases of −0.27 K (GloboLakes) and −0.31 K (C-GLOPS), whereas the

median bias of LMLT is 1.56 K. When mapped from their respective native resolutions to a common 9 km Equal-Area Scalable Earth (EASE) Grid 2.0 projection, similar relative performance was observed. LMLT and LSWT data are closer in performance over the 9 km grid cells that exhibit a small range of lake cover fractions (0.05–0.5). Despite comparable relative performance, ERA5 Land shows great advantages in spatial coverage and temporal resolution. In summary, an integrated evaluation on data accuracy, long-term availability, global coverage, temporal resolution, and regular forward processing with modest data latency led us to conclude that LMLT from the ERA5 Land Reanalysis product represents the most optimal path for use in the development of a long-term soil moisture product.

Keywords: Lake Mix-Layer Temperature (LMLT); Lake Surface Water Temperature (LSWT); ERA5 Land; Global Observatory of Lake Responses to Environmental Change (GloboLakes); Copernicus Global Land Operations Cryosphere and Water (C-GLOPS)

2.1. Introduction

Soil moisture is a critical component of the Earth's systems, mainly because of its capability to control surface energy fluxes, potentially exchange with the atmosphere, and partitioning precipitation into infiltration and surface runoff in terms of water budget (Dai et al. 2004; Entekhabi et al. 2010a; Koster et al. 2004; Petropoulos 2013). Given its relatively slow variance, soil moisture has been recognized as an essential variable in climatic studies and numerical weather predictions (Entekhabi et al. 2010a; Koster et al. 2004). Knowledge of accurate soil moisture measurements could benefit a variety of applications, ranging from drought monitoring, flood and landslide prevention, agricultural productivity improvements, and weather forecasts (Dai et al. 2004; Entekhabi et al. 2010a; Koster et al. 2004). However, a global coverage of long-term soil moisture monitoring by in-situ measurements is impractical only from the perspectives of expenditure and manpower required by the operation and maintenance of the associated facilities.

In recent decades, satellite-based surface soil moisture products have created extensive opportunities to study terrestrial-atmosphere interactions and hydrological cycles at a global scale. Compared to optical and active microwave sensors, passive microwave sensors are more sensitive to surface soil moisture in the presence of the same confounding factors (e.g., clouds, vegetation, surface roughness, etc.) (Al-Yaari et al. 2014; Kim et al. 2019) while offering more frequent repeat global coverage of about 2–3 days, as compared to more than 10 days by active sensors. These advantages of passive microwave remote sensing of soil moisture have been the principal driving force behind the application of various spaceborne radiometers (e.g., Special Sensor Microwave Imager (SSMI) (Lakshmi et al. 1997), Advanced Microwave Scanning

Radiometer (AMSR-E) (Njoku et al. 2003), Soil Moisture and Ocean Salinity (SMOS) (Kerr et al. 2010) and Soil Moisture Active Passive (SMAP) (Entekhabi et al. 2010a)) for soil moisture retrieval over the last decade. Lately, there has been tremendous progress in the area of improving the spatial resolution of satellite derived soil moisture to 1 km (Fang et al. 2018; Fang et al. 2013; Fang et al. 2020) and 400 m (Fang et al. 2021b).

However, the accuracy of soil moisture retrieved from satellite observations can be influenced by several factors (O'Neill et al. 2021a), such as vegetation, topography, surface roughness etc. Inland open water bodies (e.g., lakes, rivers, wetlands, etc.) are also an important error source for passive remote sensing of soil moisture (Fernandez-Moran et al. 2017b; Kerr et al. 2012; O'Neill et al. 2021a). Specifically, signals detected by satellite sensors include both emissions by land areas as well as water bodies. Without proper correction procedures, the contribution of the microwave emissions of water can add to microwave emissions from adjacent land, resulting a systematic wet bias in soil moisture estimates (O'Neill et al. 2021a). To mitigate this contamination, it is necessary to separate the mixture of land-water brightness temperature and remove the partial emissions contributed by water bodies. This can be achieved through subtraction of water emissions from the total brightness temperatures. Water temperature representative of inland water bodies is thereby a required variable in the estimation of water radiation within the field-of-view (FOV) of the radiometer upon antenna gain pattern correction, in addition to the fractions of water cover.

Recently, a number of assimilation and reanalysis data sets that describe long-term variations of water temperatures have been released (Carrea and Merchant 2019, 2020b; Muñoz-Sabater et al. 2021a). These data sets include lake surface water temperature (LSWT) from the Global Observatory of Lake Responses to Environmental Change (GloboLakes) (Carrea and Merchant

2019) and the Copernicus Global Operations Cryosphere and Water (C-GLOPS) (Carrea and Merchant 2020b), as well as the lake mix-layer temperature (LMLT) from the European Centers for Medium-Range Weather Forecast (ECMWF) ERA5 Land Reanalysis (Muñoz-Sabater et al. 2021a; Sabater 2019). These recent LSWT products have adopted the algorithms used in estimating sea surface temperature but with modifications of processing procedures and additions of lake-related parameters (Carrea and Merchant 2020a; MacCallum and Merchant 2012), thus yielding more reliable estimations of water temperature specific to inland water bodies (compared to those derived from algorithm designed for sea surface temperature). In addition, the above LSWT and LMLT data sets are expected to provide data over the open water adjacent to land, given that they are projected in regular latitude-longitude grids, which is especially helpful for the water correction in passive soil moisture retrieval algorithms. However, these water temperature data sets are supplied at vastly different spatial and temporal resolutions because of diverse computational procedures and input sources (Carrea and Merchant 2019, 2020b; Muñoz-Sabater et al. 2021a).

Lake temperature data sets that incorporate long-term satellite remote sensing observations have been widely investigated, evaluated and applied, since LSWT is considered as an important indicator of climate change and highly related to the chemical and physical process within the water bodies (Lieberherr and Wunderle 2018; Samuelsson et al. 2010). In order to study lake responses to climate change, validation efforts of LSWT products are generally carried out at the lake scale or at a coarse temporal resolution (Crosman and Horel 2009; Layden et al. 2015; Lieberherr and Wunderle 2018; Njoku and Kong 1977; O'Reilly et al. 2015; Schaeffer et al. 2018; Wan et al. 2017). For example, Lieberherr and Wunderle (2018) evaluated the LSWT derived from the Advanced Very High-Resolution Radiometer (AVHRR) with in-situ

measurements over 26 European Lakes where AVHRR LSWT data have absolute biases within 2 K and Pearson Correlations (R) of more than 0.9. In addition, LSWT data provided by Landsat 5 and 7 were evaluated in 59 water bodies (Schaeffer et al. 2018), and their mean absolute error is 1.34 K for those grids away from the land areas more than 180 m. Moreover, studies of intra- and inter-annual variability of LSWT as a response to climate change normally requires long-term LSWT data with a coarse temporal resolution (Lieberherr and Wunderle 2018; Maberly et al. 2020). The authors of Maberly et al. (2020) analyzed seasonal patterns of LSWT and constructed global lake thermal regions based on the LSWT data at a half-monthly time step.

However, requirements for water temperature data of inland water bodies are greatly different in the application of water correction in the derivations of soil moisture from passive remote sensing observations. Water temperature data are expected to be representative over a large but fixed spatial scale, such as the 9 km Equal-Area Scalable Earth (EASE) Grid 2.0 (Brodzik et al. 2012) which is a typical satellite-based Level 2 passive soil moisture retrieval setting (e.g., SMAP). Data sets with a higher temporal frequency are more optimal and conform to the instantaneous satellite observations of soil moisture given that the surface temperature variations in time can greatly impact soil moisture retrieval. Assessments and inter-comparisons of lake water temperature data sets in the frame of soil moisture are insufficient.

LMLT data of ERA5 Land represent the average temperatures for the top layer of lakes, which are different from surface skin temperatures (on the orders of micrometers and millimeters) illustrated by satellite based LSWT products. A default depth of 25 m has been widely used in the derivation of ERA5 Land LMLT given the unavailability of water depths over most inland water bodies (Sabater 2019). Therefore, systematic discrepancies between LMLT and LSWT data sets are expected as they reflect the lake thermal conditions at different depths, but this is

rarely investigated. Research from Muñoz-Sabater et al. (2021a) updated the depth information and compared LMLT data with in-situ measurements (with only one station having hourly data), and concluded that the mean absolute errors of ERA5 Land are from 2.25 to 3.22 K. Nevertheless, the representative depth (about 1 m) of in-situ observed temperature is also different from GloboLakes and C-GLOPS LSWT products. Moreover, inter-comparisons between areal LSWT and LMLT data sets could also be useful to more or less indicate their quality over a wider geographical coverage.

In light of these, three lake temperature data sets were selected for evaluation and inter-comparison over a five-year period from 2007 to 2011, including LSWT from GloboLakes and C-GLOPS, and LMLT from ERA5 Land. Since the assessments of the above products emphasize their usefulness in providing correction for microwave emissions from open water for passive soil moisture retrieval, performance evaluation was conducted at their native spatial scales and on the 9 km EASE Grid as an illustration. The latter was used to demonstrate how these data sets perform in a larger spatial extent common in passive soil moisture retrieval products. In addition to data accuracy, several aspects were also examined in this study, which are long-term availability, global coverage, temporal resolution, regularity in data maintenance and extension.

2.2. Study Regions and Lake Temperature Data Sets

2.2.1. Study Regions

In this study, the buoys used to represent the in-situ measurements of lake water temperature are distributed over 11 lakes in the North America as shown in **Figure 2.1**. Most available stations are concentrated on the Great Lakes region composed of Lake Superior (47.7º N, 87.5º W), Lake Huron (44.8º N, 82.4º W), Lake Erie (42.2º N, 81.2º W), Lake Michigan (44.0º N, 87.0 W), and

Lake Ontario (43.7º N, 77.9º W). Along the border of the United States and Canada, the total surface area of these five lakes is around 244106 km^2. Additionally, the Great Lakes have abundant freshwater resources, approximately 21% of the global surface freshwater volume. There are also several stations within the Great Lakes area, including Lake Saint Clair (42.5º N, 82.7º W), Lake Nipissing (46.3º N, 79.8º W), and Lake Smicoe (44.4º N, 79.3º W). The remaining three lakes are Lake Great Slave (61.5º N, 114.0º W), Lake Winnipeg (52.1º N, 97.3º W), and Lake of the Woods (49.2º N, 94.8º W) that are located in the further north regions. According to the classification of lake thermal regions developed by Maberly et al. (2020), these lakes could be grouped into Northern Temperate and Northern Cool with average temperatures around 282.95 K and 279.25 K, respectively.

2.2.2. Lake Temperature Data Sets

Three lake temperature products consisting of two satellite-based LSWT and one-model based LMLT were examined in this study. The specifications of these data sets are summarized in **Table 2.1**. In-situ measurements from buoys available within the studying period were used as the reference in the validation.

2.2.2.1. Global Observatory of Lake Responses to Environmental Change (GloboLakes)

GloboLakes LSWT version 4.0 provides daily averages of surface water temperature for around 1000 lakes mostly extracted from the Global Lakes and Wetland Database (GLWD) Level 1 product (Carrea and Merchant 2019; Politi et al. 2016). The retrievals of LSWT were derived using the same algorithms on satellite observations from various types of sensors and platforms, including the Along Track Scanning Radiometer (ATSR-2) on European Remote Sensing Satellite (ERS-2), Advanced Along Track Scanning Radiometer (AATSR) on Envisat, and

AVHRR on MetOpA (Carrea et al. 2015). This data set contains a long-term daily LSWT data ranging from 1995 to 2016 with a spatial resolution of 0.05°.

In this study, daily files with a regular latitude-longitude grid were used. In addition to skin temperature of lakes' surface, the information of uncertainty estimations and quality level related to LSWT retrievals is also included in the data set. Quality level is considered as a flag to quantify the confidence of LSWT retrievals, such as the degree of interference caused by cloud contamination and is unable to necessarily represent the accuracy level. Data with all quality levels were retained in order to satisfy the need of high temporal frequency in water correction of soil moisture retrievals. It should be noted that a static land-water mask based on the European Space Agency Climate Change Initiative (ESA CCI) Land Cover map with a spatial resolution of 300 m for the time period 2005–2010 has been applied to identify and delineate the lake pixels, which could lead to inappropriate estimations of LSWT over water bodies with dynamic surface extensions (Carrea et al. 2015). In addition, the combined use of observations from different satellite sensors will inevitably introduce more error sources despite the increase of available samples and the harmonized processing.

2.2.2.2. Copernicus Global Land Operations (C-GLOPS)

LSWT of C-GLOPS contains gridded 10-day mean surface water temperature over more than 1000 lakes composed of the world's largest and those water bodies of particularly scientific interest (Carrea and Merchant 2020b). Specifically, the 10-day LSWT data of C-GLOPS are temporally aggregated and generated via calculating the weighted average of Level 3 gridded daytime files partitioned by the 1st to 10th, and 11th to 20th, and 21st to the end of each month. Three types of LSWT are included in this data set, which are historical (v1.0.2), reprocessed (v1.0.2), and near real-time products (v1.0.1 and v1.1.0), depending on the timeliness and

completeness of the Level 1b inputs during processing (Carrea and Merchant 2020c). Since the LSWT retrievals at a regular 1/120° resolution grid are originally derived from the visible and infrared from the AATSR and Sea and Land Surface Temperature Radiometer (SLSTR) onboard Envisat, and Sentinel 3A and 3B, the temporal coverage of this data set was separated into two periods: May 2002 to April 2012, and November 2016 to present. This data set can be freely accessed and downloaded through the Copernicus Global Land Portal (https://land.copernicus.eu/global/products/lswt, accessed on March 9, 2021). Similar to GloboLakes LSWT, the uncertainty information and quality levels are also available, and all quality-level data were adopted in this work.

According to Carrea and Merchant (2020c), the accuracy of C-GLOPS overall fulfills the uncertainty requirement of 1 K comparing against in-situ measurements. However, the utilization of LSWT with a quality level lower than 3 should be handled carefully (Carrea and Merchant 2020a). Again, one common factor that could influence the observations is the presence of clouds given that the AATSR and SLSTR operate at visible and infrared bands (Carrea and Merchant 2020a). Additionally, the application of a uniform threshold standard has risks in the identification of water and non-water grid cells (Carrea and Merchant 2020a). Contamination from land associated signals could affect the retrieval quality. As a result, LSWT retrievals in the areas near the land are more likely to have lower performance. Moreover, the numbers and observation times used in the temporal aggregation varies from place to place, possibly leading to spatially or temporally inconsistent thermal conditions, even for the same lake (Carrea and Merchant 2020a). Furthermore, a static mask representing the maximum surface water extent of lakes from 2015 to 2010 was adopted (Carrea and Merchant 2020a), which is improper for those water bodies with evident changes in areas at the 1 km^2 scale.

2.2.2.3. ERA5 Land

ERA5 Land is a reanalysis product containing a series of variables that describe the state of various land components from 1981 to the current time (Muñoz-Sabater et al. 2021a; Sabater 2019). Although ERA5 Land utilizes ERA5 atmospheric forcing as inputs to derive the land components, the ERA5 Land data set has a higher spatial resolution at 0.1° compared to its predecessor at 0.25°. Without coupling with the atmospheric module of ECMWF's Integrated Forecasting System and ocean wave models as well as data assimilation, the processes related to the computation and delivery of ERA5 Land are expected to be more efficient. The core of ERA5 Land is the Tiled ECMWF Scheme for Surface Exchanges over Land incorporating land surface hydrology (H-TESSEL) (version CY45R1).

In this study, hourly LMLT from the ERA5 Land data set was selected to reflect the thermal conditions over various inland water bodies. The ECMWF Integrated Forecasting System separates the vertical structure of inland water bodies into two levels: the upper (mix layer) and lower (thermocline layer) layers given the implementation of the Flake model (Mironov 2008; Mironov et al. 2010). In light of this, LMLT represents the average water temperature at the uppermost layer of lakes and differs from skin temperature at the water surface. Lake-related variables can be calculated for each pixel so as to incorporate the sub-grid features of the small to medium-size lakes (Manrique-Suñén et al. 2013; Mironov et al. 2010; Muñoz-Sabater et al. 2021a). ERA5 Land data set provides a spatially complete temperature map for both inland water bodies and coastal waters, but it is necessary to use an auxiliary data set that describes the water fraction within the grid cell simultaneously with LMLT. This static map of lake cover is provided by the reanalysis product of ERA5 (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview,

accessed on November 15, 2020) at 0.25° and then interpolated into 0.1° and the 9 km EASE grid. For inland water bodies where lake depths are not available, a default value of 25 m has been used (Sabater 2019), which is likely to result in unreasonable LMLT retrievals, especially for those lakes with shallow depths.

2.2.2.4. In-Situ Measurements of Lake Surface Temperature

In-situ lake surface temperatures collected by the moored buoys from fixed locations were used to benchmark the performance of assimilated and reanalysis products. Based on the stations considered in the Quality Validation Report of C-GLOPS (Carrea and Merchant 2020c), 34 buoys (**Table A1**) over lakes in Northern America were selected. These measurements are from either the National Data Buoy Center (NDBC) (https://www.ndbc.noaa.gov/, accessed on March 14, 2021) or Fisheries and Ocean Canada (FOC) (https://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/data-donnees/index-eng.asp, accessed on March 14, 2021). **Figure 2.1** describes the geographical locations of in-situ measurements. Historical files of hourly water temperature can be used in validation for remotely sensed and modelled lake temperature data sets.

Table 2.1. Summary of lake temperature data sets used in this study.

Datasets	**GloboLakes**	**C-GLOPS**	**ERA5 Land**
Variable	LSWT	LSWT	LMLT
Spatial Coverage	991 inland waters	1018 inland waters	Global inland water bodies and coastal waters
Spatial Resolution	0.05° × 0.05°	1 km × 1 km	0.1° × 0.1°
Temporal Coverage	June 1995–December 2016	May 2002–March 2012; June 2016–present	January 1981 to present
Temporal Resolution	Daily	10-day interval	Hourly

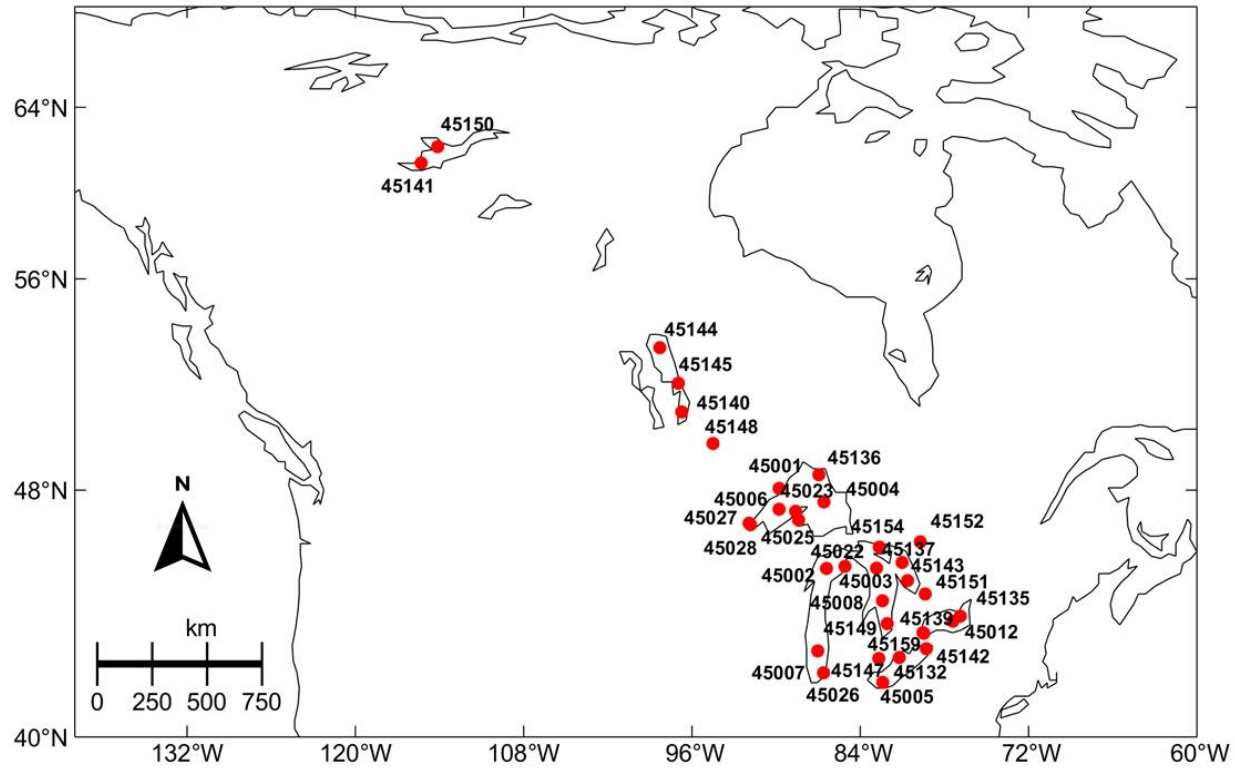


Figure 2.1. Distribution of in-situ buoys (red points).

2.3. Methodology and Assessment Metrics

2.3.1. Validation against In-Situ Measurements

Evaluation of lake temperature data sets were firstly assessed by comparing against 34 in-situ buoys distributed in the North America at their native spatial and temporal resolutions. LSWT data of all quality levels from GloboLakes and C-GLOPS were retained in the assessments to incorporate as many observations as possible. Although low-quality-level samples does not necessarily represent inferior accuracy, the overall better statistic metrics for LSWT data at quality levels of 4 or 5 have been obtained (Carrea and Merchant 2020c). Despite that, the degradation on the performance of LSWT data sets is expected to be limited as the majority of LSWT data with quality levels of 4 and 5 (Carrea and Merchant 2020c). Following a similar

manner used in Kerr et al. (2012), data with water temperature below 275.15 K (2 ºC) were screened out to avoid unreliable observations near the freezing point.

Given that in-situ data were collected from different sources (**Table A1**), quality control measures have been adopted to filter out those abnormal and suspicious observations. A pre-determined threshold (in addition to 275.15 K mentioned above) and manual inspection were used. Firstly, a threshold of 306.55 K was employed and those in-situ observations with temperatures above this value were removed. This boundary value represents the maximum temperature for lakes under Northern Temperate and Northern Cool classifications (all in-situ stations are within these areas) (Maberly et al. 2020). Then, some sudden spikes or extremely low temperatures described by the in-situ time series were manually masked.

Since the temporal resolutions of in-situ observation, LMLT and LSWT data sets vary from sub-hourly to 10-day intervals, temporal averaging procedures were re-quired before assessments. Similar to the steps applied in Carrea and Merchant (2020c), for example, 10-day averages of in-situ observations were computed for evaluating C-GLOPS LSWT. In addition, it should be noted that data from in-situ observations, ERA5 Land, GloboLakes, and C-GLOPS reflect water temperature at different depths. Generally, LSWT data measured by satellite instruments could represent the water temperature from many micrometers to a few millimeters depending on the instrument frequencies, whereas the in-situ buoys usually detect the water temperature at around 1 meter (as described in **Table A1**) under the water surface and without significant effects from diurnal cycles. In terms of ERA5 Land, LMLT reflects the average temperature for the top-most layers of lakes and its corresponding depth is dependent on lake depth used in the Flake model (Mironov et al. 2010). However, a default depth of 25 meters was used in the derivations of LMLT over most inland water bodies due to the common unavailability of depth information

(Sabater 2019). Therefore, discrepancies between LMLT and LSWT are expected to exist but have not been widely investigated yet. Regarding the LSWT and in-situ observations, skin effect could account for the -0.2 K error mainly generated by the different observing depths between in-situ measurements and satellite sensors (Carrea and Merchant 2020c). The remaining residuals could be contributed by other unquantified factors, such as the near-surface stratification, underestimating atmospheric attenuation or overestimating surface emissivity (Carrea and Merchant 2020c; Crosman and Horel 2009).

Subsequently, ERA5 Land, GloboLakes, and C-GLOPS were resampled into the EASE 9km scale by bilinear interpolation in order to quantify the performances of these data sets in the context of soil moisture retrievals as well as analyze the influences on their data accuracy by rescaling processing. Again, in-situ measurements were used as reference to assess the performances of three lake temperature products at the 9km EASE grid after temporal averaging. It should be noted that statistical metrics were considered to be effective and calculated only if there are at least 30 paired data between two data sets (Hogg et al. 2010; Pallant 2011; Woolway and Merchant 2019). In addition to the uncertainty of products, seasonal trends captured by lake temperature data sets were also illustrated and examined by time series over Lake Superior and Lake Huron. Furthermore, dependency of errors between the above data sets and in-situ measurements on temperature ranges were also investigated.

2.3.2. Inter-comparisons among Lake Temperature Products at the 9km Scale

Due to the fact that in-situ measurements are only available in the limited geographical regions, a global-scale assessment based on in-situ observations is difficult to achieve. Therefore, inter-comparisons among different lake temperature products could be an effective alternative to corroborate their relative performances. It is particularly feasible when the LSWT retrieval

algorithm is found on physics (Carrea and Merchant 2020c) to guarantee the consistency of derived LSWT values. The EASE 9km grid cells with lake fractions smaller than 0.05 were not considered in accordance the requirements of passive remote sensing soil moisture retrieval algorithm.

Following similar steps adopted in the Section 2.3.1, overlapped pixels among three lake temperature data sets were firstly determined and then temporal averaging was separately applied before three groups of pairwise comparisons (**Figure 2.2**). For example, EAR5 Land hourly LMLT data were temporally aggregated at a 10-day scale prior to comparing with C-GLOPS LSWT. The general workflow adopted for the assessments and comparisons of performances of considered water temperature data sets has been described in **Figure 2.2**, and different filling colors correspond to various temporal scales. Additionally, differences and correlations among LMLT and LSWT products were studied conditioned by lake cover fractions. More importantly, the numbers of available pixels and data during the studying period were investigated and compared given that the requirements of high temporal resolution and wide spatial coverage in the soil moisture retrievals.

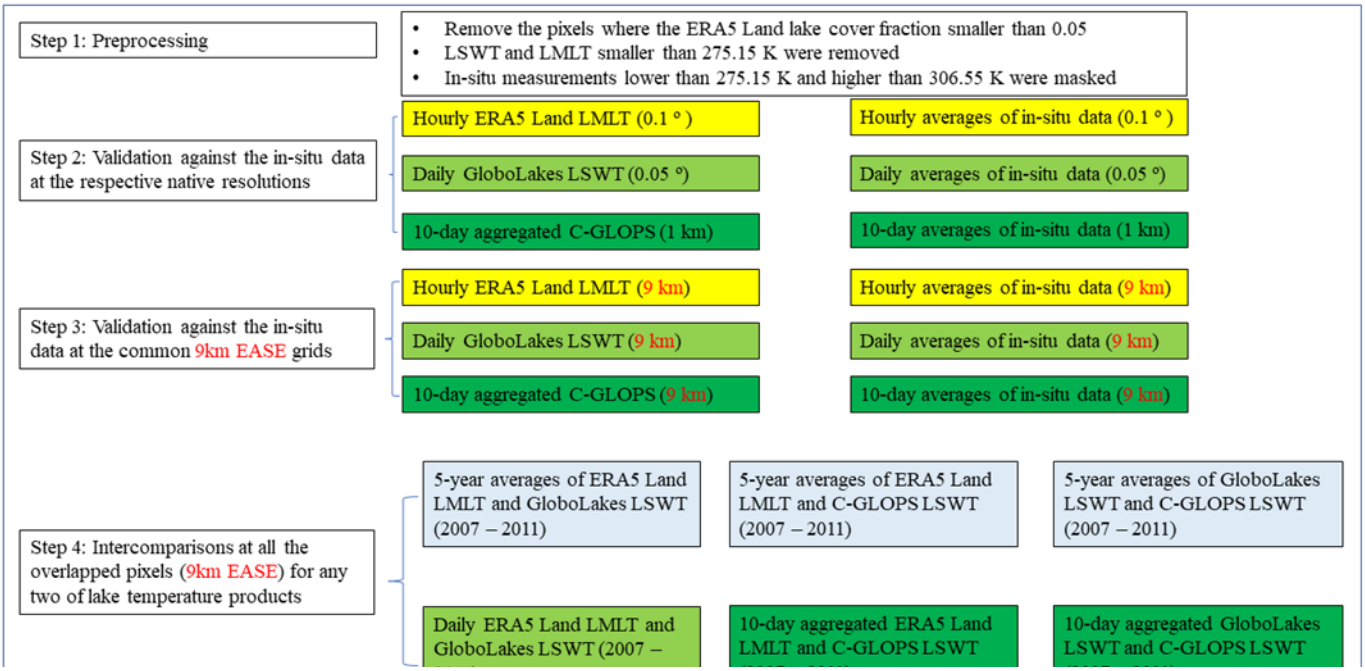


Figure 2.2. Schematic diagram that describes the workflow adopted in this study for evaluating the performance of lake temperature data sets. Different colors of the boxes indicate different temporal scales of data sets.

2.3.3. Assessment Metrics

Six statistical metrics were adopted to reflect the accuracy of assessed products, which are mean bias, standard deviations of bias, root mean square error (RMSE), the Pearson correlation coefficient (R), median bias, and robust standard deviation (RSD). Standard deviation and RSD describe the dispersion of mean bias and median bias, respectively. Robust statistical metrics (i.e., median bias and RSD) were considered because they are less susceptible to outliers due to extreme observations (Carrea and Merchant 2020c). However, unreliable estimations of water temperature over inland water bodies are inevitable, possibly due to the undetected clouds or horizontal variability at water surface caused by winds (Carrea and Merchant 2020c; MacCallum and Merchant 2012). RSD is calculated as the following steps: (1) calculate the absolute differences between the biases and median bias (2) determine the median value of those prior absolute differences (3) multiply a factor (1.5) with the median value obtained in Step (2) as

RSD (Carrea and Merchant 2020c). Here, water temperatures from the evaluated data sets and in-situ measurements are denoted as T_{est} and T_{true}. The formulas of those statistical metrics are given in Equations (1) to (6), where E […], M […], σ, μ, and N represent the arithmetic mean, median, standard deviation, mean bias, and the number of samples, respectively.

$$mean\ bias = E[T_{est}] - E[T_{true}] \tag{1}$$

$$\sigma = \sqrt{\frac{(x_i - \mu)^2}{N}} \tag{2}$$

$$RMSE = \sqrt{E[(T_{est} - T_{true})^2]} \tag{3}$$

$$R = \frac{E[(T_{est} - E[T_{est}])(T_{true} - \mathrm{E}[T_{true}])]}{\sigma_{est}\sigma_{true}} \tag{4}$$

$$median\ bias = M[T_{est} - T_{true}] \tag{5}$$

$$RSD = M[|(x_i - M[x_i])|] \times 1.5 \tag{6}$$

2.4. Results

In this section, assessment results of LMLT from ERA5 Land and LSWT from GloboLakes and C-GLOPS are presented. Firstly, validations against their corresponding temporal averaged in-situ measurements were conducted to reflect their innate performances at native scales. Time series and scatter plots were applied to describe the consistency in terms of the seasonal trends and proximity of absolute values among different products. In the framework of passive remote sensing soil moisture retrievals, the performances of ERA5 Land, GloboLakes and C-GLOPS at the EASE 9 km pixels were further analyzed, and the effects of the errors along the increased temperatures were also investigated. Additionally, spatial coverages, temporal resolutions, and overlapped pixels of LMLT and LSWT data sets were studied and prepared for inter-comparisons

over wider areas where in-situ measurements are scarce or unavailable. Discrepancies between LMLT and LSWT products were then examined across various lake cover fractions.

2.4.1. Overall Performance of Lake Temperature Products at Their Original Resolutions

Figure 2.3 describes the boxplots of median bias, RMSE and R for hourly LMLT of ERA5 Land, daily LSWT of GloboLakes, and 10-day aggregated LSWT from C-GLOPS, relative to the lake temperature from in-situ observations. In terms of median bias and RMSE, LSWT from GloboLakes and C-GLOPS is closer to in-situ measurements than LMLT from ERA5 Land. LSWT products tend to underestimate while ERA5 Land product is inclined to overestimate. Additionally, all three data sets have exhibited a strong capacity to capture the temporal variations because their R values are consistently higher than 0.8. In general, marginally inferior accuracy of ERA5 Land is expected because LSWT data from GloboLakes and C-GLOPS are able to reflect the thermal conditions over smaller areas by mapping at finer spatial resolutions and thereby correspond better to point in-situ measurements.

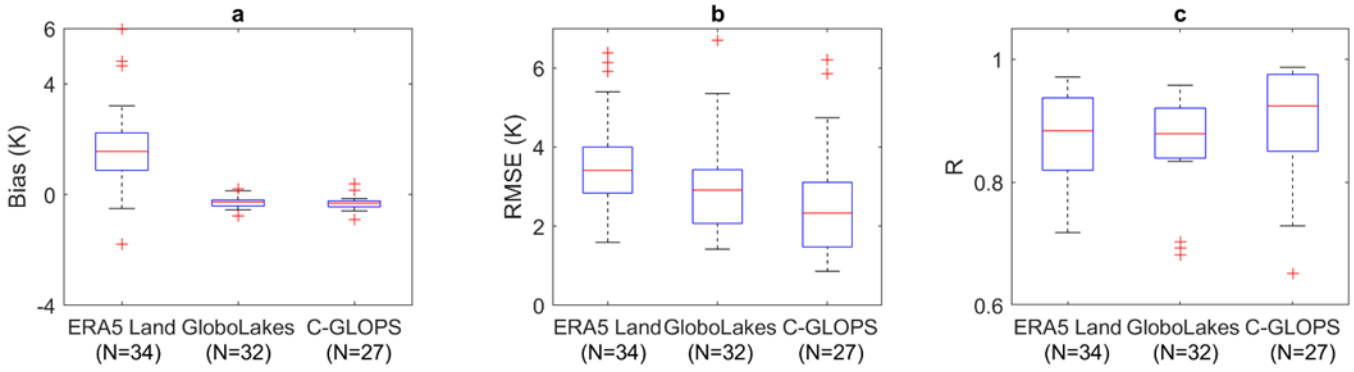


Figure 2.3. Boxplots of median bias (a), RMSE (b), and R (c) for hourly LMLT of ERA5 Land, daily LSWT of GloboLakes, and 10-day LSWT of C-GLOPS at their native spatial resolutions. N represents the number of in-situ stations used to calculate the metrics.

As mentioned in Section 2.3.3 median bias and RSD are considered to be more reliable in the assessments of water temperature data sets by reducing the impacts of possibly contaminated observations, compared to mean bias and its standard deviation (Carrea and Merchant 2020c). According to **Table 2.2**, differences between the conventional and robust statistical metrics are straightforward. RSD values of LSWT products are around half of their standard deviations whereas the discrepancies of two types of metrics are minor for LMLT from model based ERA5 Land data set.

Comparisons of time series have been carried out over Lake Superior and Lake Huron, where several buoys are available during the studying period to further assess the seasonal consistency between in-situ measurements and lake temperature data sets. As shown in **Figure 2.4** and **Figure 2.5**, seasonal trends of lake water temperature are relatively stable, and the maximum and minimum water temperatures occur in late summer (August or September (O'Reilly et al. 2015)) and spring (April or May) regardless of ice cover periods. Overall, seasonal patterns and averages of lake water temperatures in Lake Superior and Lake Huron are highly similar. The ERA5 Land product has shown noticeable overestimation, especially during summer seasons compared to GloboLakes and C-GLOPS.

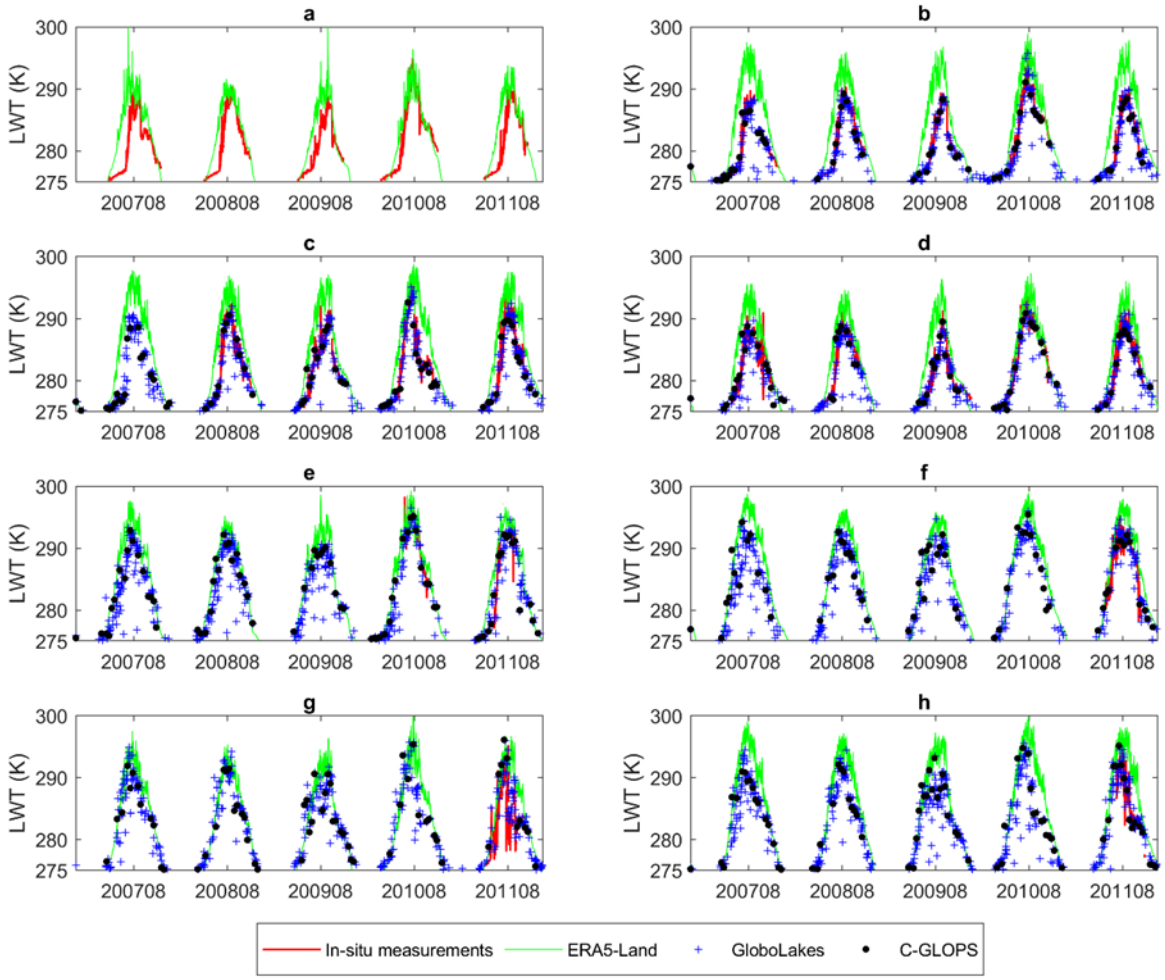


Figure 2.4. Variations of lake water temperature on Lake Superior from 2007 to 2011 (**a**: Station 45001 48.06ºN, 89.79ºW; **b**: Station 45004 47.59ºN, 86.59ºW; **c**: Station 45006 47.34ºN, 89.79ºW; **d**: Station 45136 48.54ºN, 86.95ºW; **e**: Station 45023 47.27ºN, 88.61ºW; **f**: Station 45025 46.97ºN, 88.40ºW; **g**: Station 45027 46.86ºN, 91.93ºW; **h**: Station 45028 46.81ºN, 91.83ºW).

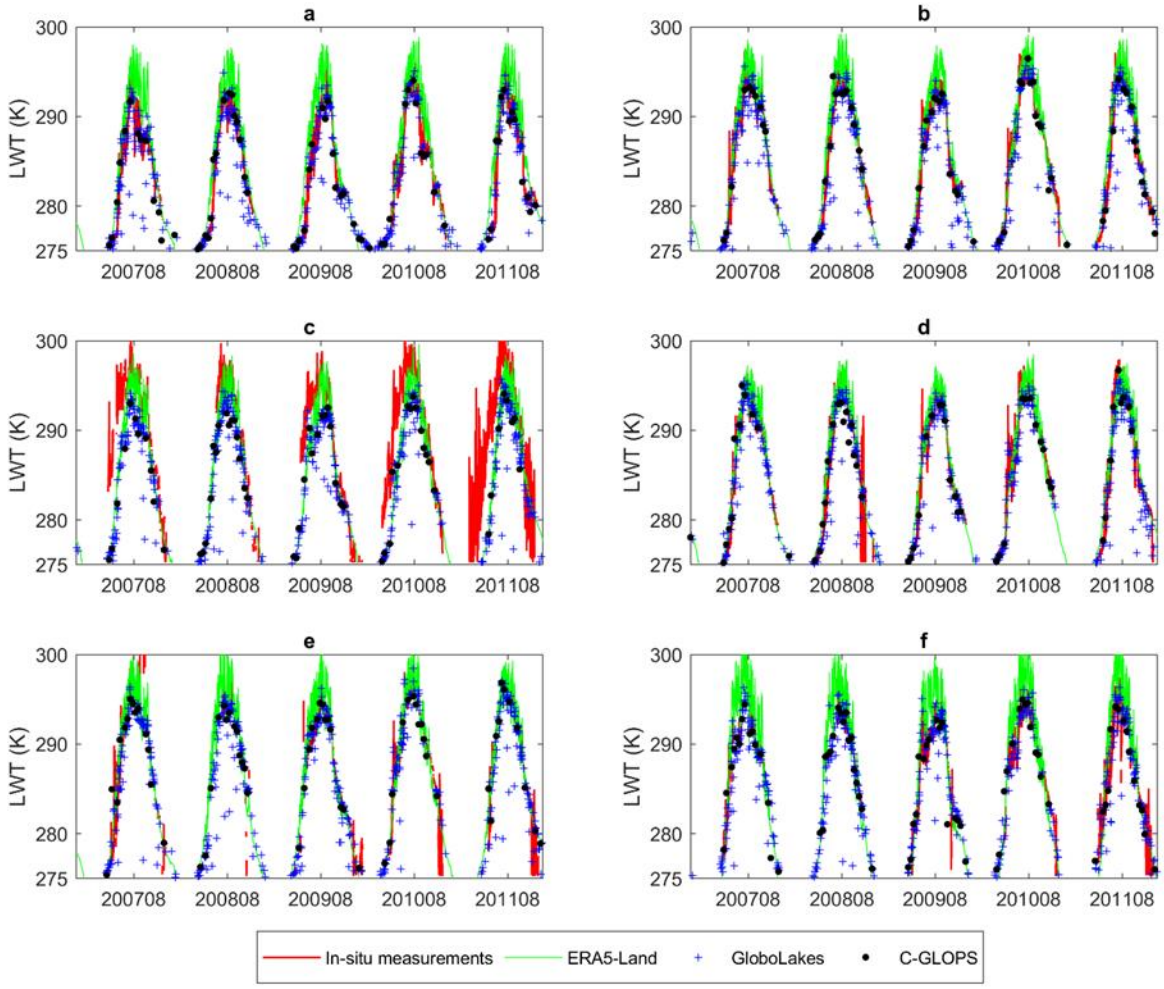


Figure 2.5. Variations of lake water temperature on Lake Huron from 2007 to 2011 (**a**: Station 45003 45.53ºN, 82.84ºW; **b**: Station 45008 44.28ºN, 82.42ºW; **c**: Station 45137 45.54ºN 81.02ºW; **d**: Station 45143 44.94ºN, 80.63ºW; **e**: Station 45149 43.54ºN, 82.08ºW; **f**: Station 45154 46.05ºN, 82.64ºW).

According to **Figure 2.6**a, c, and e, most data from lake temperature data sets and in-situ measurements are consistent and distributed along the 1:1 line. Again, ERA5 Land tends to overestimate, and LSWT values from GloboLakes and C-GLOPS have smaller biases relative to in-situ measurements. Over the locations where in-situ stations are considered in this study, lake water temperature data are mostly concentrated on the interval from 290 to 295 K. Deviation extents between in-situ observations and lake temperature data sets are relatively stable with the increase of water temperature (**Figure 2.7**a, c, and e). Moreover, the ranges of differences seem to become larger under warmer conditions.

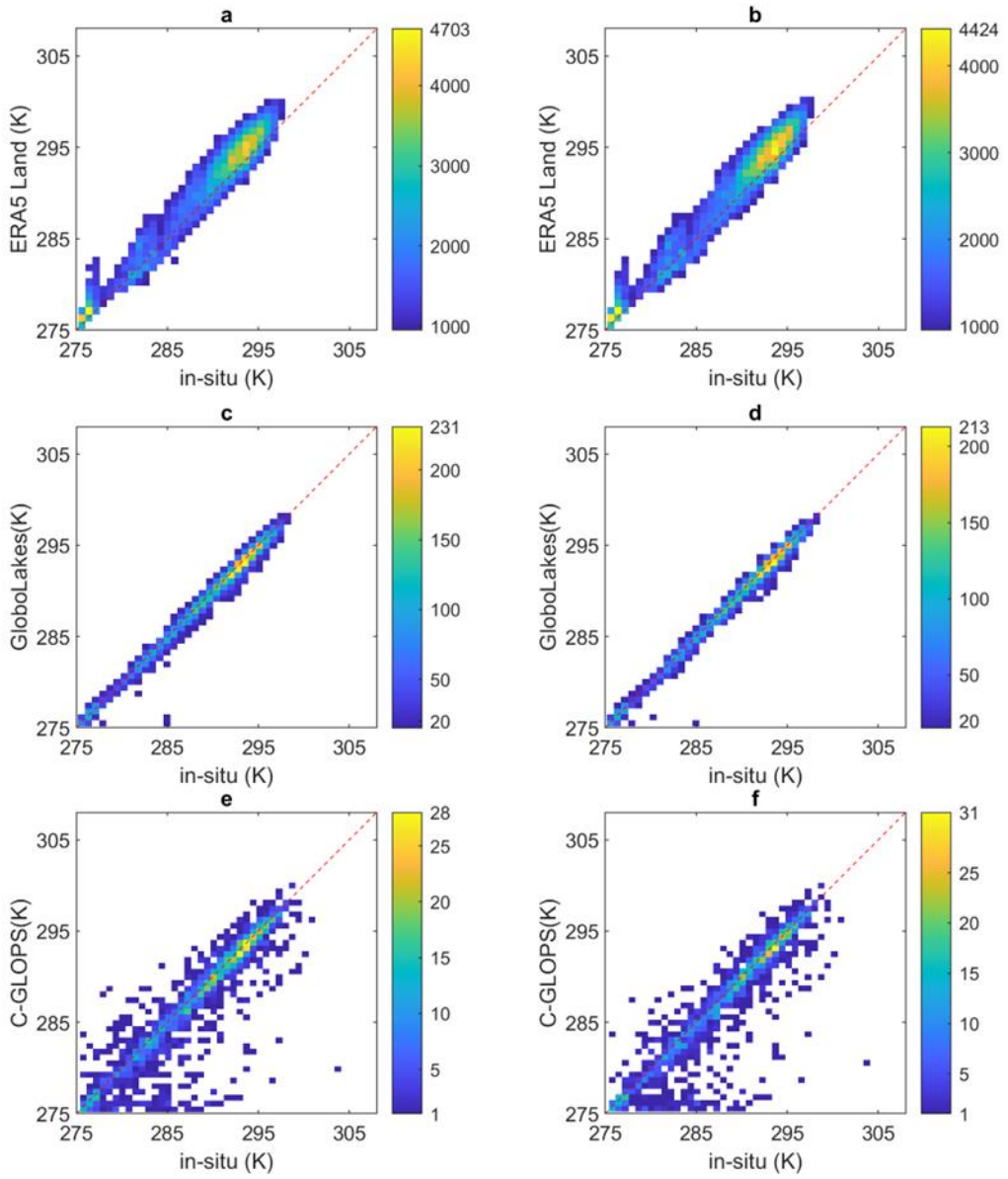


Figure 2.6. Scatter plots of data from lake temperature data sets at their original resolutions (**a**, **c**, **e**) and at the 9 km EASE grid (**b**, **d**, **f**) compared to in-situ measurements. The number on the color bar represents the number of available data samples within each assigned temperature interval. Points closer to yellow mean more samples lie in that temperature range whereas these closer to blue mean less samples lie in that temperature range.

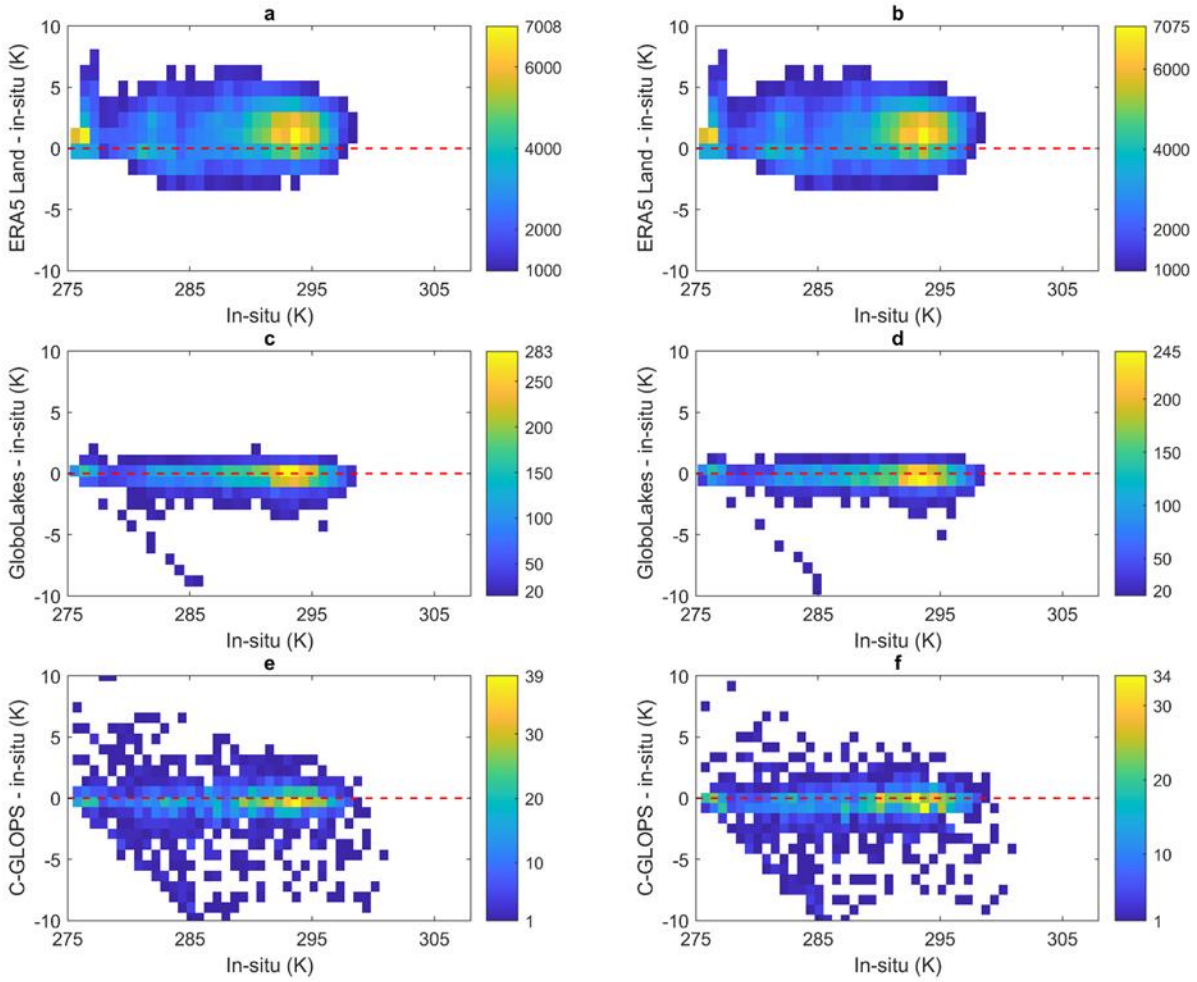


Figure 2.7. Variations of errors between lake temperature data sets and in-situ measurements with the increase of water temperature at their original resolutions (**a**, **c**, **e**) and at the 9 km EASE grid (**b**, **d**, **f**). The number on the color bar represents the number of available data samples within each assigned temperature interval. Points closer to yellow mean more samples lie in that temperature range whereas these closer to blue mean less samples lie in that temperature range.

2.4.2. Overall Performance of Lake Temperature Products at the 9km EASE Resolution

Lake temperature data sets were resampled to the 9 km EASE resolution and compared with in-situ measurements to measure the effects of changing spatial resolution on the estimations of lake temperature. Similar to the results obtained at their original resolutions, LSWT values from the GloboLakes and C-GLOPS have smaller bias and RMSE than LMLT from ERA5 Land

(**Figure 2.8**). In terms of temporal correlations, all three data sets have comparable performances with R values of 0.89, 0.88, and 0.94 (**Figure 2.8** and **Table 2.2**).

According to **Table 2.2**, the median bias (mean bias) of ERA5 Land, GloboLakes, and C-GLOPS are 1.36 K (1.40 K), −0.29 K (−0.90 K), and −0.36 K (−0.69 K). RSD values of LSWT products are smaller than the standard deviations of mean bias. Lake water temperatures provided by the resampled products were matched and compared with in-situ measurements at various corresponding temporal resolutions (**Figure 2.6**b, d, f). LSWT data of GloboLakes and C-GLOPS at the 9 km EASE resolution still have a good agreement with in-situ observations but the numbers of paired samples decrease compared to those at their native resolutions. As shown in **Figure 2.7**d, f, the errors between LSWT products at the coarser resolution and in-situ measurements are negatively extended.

Table 2.2. Statistical metrics between in-situ measurements and lake temperature products at their native resolutions and the 9 km EASE grid.

Resolution	Native Resolution						9 km EASE Grid					
Data sets/Metrics	Median Bias	RSD	Mean Bias	STD*	RMSE	R	Median Bias	RSD	Mean Bias	STD *	RMSE	R
ERA5 Land	1.56	2.76	1.64	2.76	3.41	0.88	1.36	2.68	1.40	2.68	3.34	0.89
GloboLakes	−0.27	0.86	−0.97	2.77	2.91	0.88	−0.29	0.83	−0.90	2.82	2.97	0.88
C-GLOPS	−0.31	0.93	−0.60	2.18	2.33	0.92	−0.36	0.92	−0.69	2.15	2.24	0.94

** STD: Standard Deviation.*

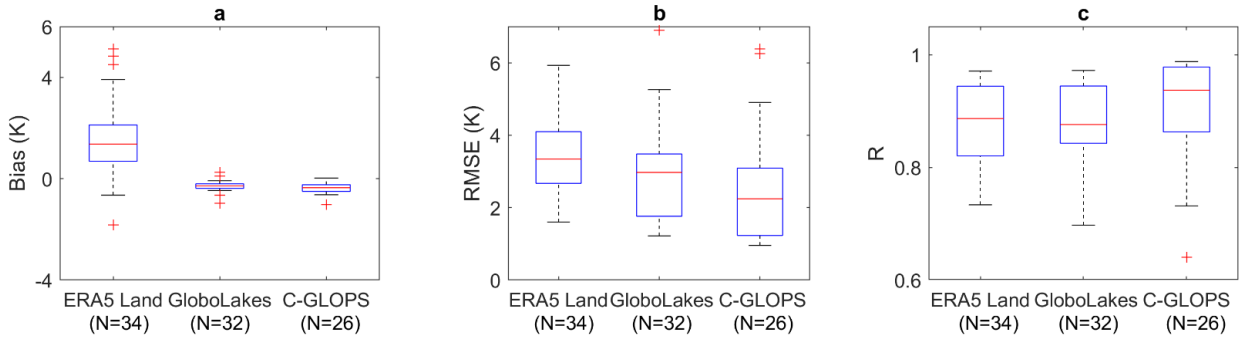


Figure 2.8. Boxplots of median bias (a), RMSE (b), and R (c) for hourly LMLT of ERA5 Land, daily LSWT of GloboLakes, and 10-day LSWT of C-GLOPS at the 9 km EASE grid. N represents the number of in-situ stations used to calculate the metrics.

2.4.3. Matchup Intercomparisons of Lake Temperature Products

Comparisons of resampled lake temperature data sets have been performed on all their overlapping regions across the world to comprehensively assess the performance of ERA5 Land LMLT, GloboLakes LSWT, and C-GLOPS LSWT. Since all three lake temperature products gridded on the 9 km EASE map still have different temporal resolutions, temporal averaging is required before their inter-comparisons. As mentioned earlier, a lake cover fraction threshold of 0.05 was used to screen out some land grids. On one hand, this percentage of lake cover conforms to the threshold of static water fraction considered in SMAP in which the quality of soil moisture retrieved in areas with a water fraction of more than 5% may be unreliable (O'Neill et al. 2021a). On the other hand, compared to 0.5, a smaller threshold is conducive to involving as many inland water bodies as possible, especially for narrow bodies and minor water areas. The inclusion of such water bodies is essential and critical in soil moisture retrievals because small-scale shallower lakes commonly have more distinct diurnal variations in temperature than sea water or deeper lakes (Mironov et al. 2010).

The number of grids with ERA5 Land product exceeds, by an order of magnitude, the GloboLakes and C-GLOPS products, which have 12781 and 11722 available pixels, respectively (**Table 2.3**). Although the temporal resolutions of GloboLakes and C-GLOPS are nominally daily and 10 days, they actually have effective LSWT data around every 4 days and 26 days, respectively. However, the ERA5 Land product has the ability to continuously update LMLT as long as the input data are available. There are a total of 10111 grids where all three products have available temperature data that are necessarily obtained in coincidence. It should be noted that there may be only one effective LSWT in certain overlapped grids. Those pixels are distributed

over various locations spanning from the Arctic region to Africa. Long-term frozen conditions could be one possible reason leading to very few observations for high-latitude grids.

Table 2.3. Spatial and temporal characteristics of lake temperature data sets.

Datasets	ERA5 Land	GloboLakes	C-GLOPS
Available Pixels	171,969	12,781	11,722
Average Temporal Resolution per available pixel (day)	1/24	~4	~26
Overlapped pixels		10,111	

Based on the available overlapping pixels between any two lake temperature products, 5-year averages were calculated to represent the grid-scale water temperatures. According to **Figure 2.9**, lake water temperatures from all three products are close to each. 5-year averages of LMLT are overall higher than 5-year averages of LSWT, with the results comparing consistently against the in-situ measurements. At the range from 280 to 285 K, 5-year averages of LSWT from C-GLOPS are slightly higher than those from GloboLakes, partially due to a small number of C-GLOPS LSWT samples over some pixels.

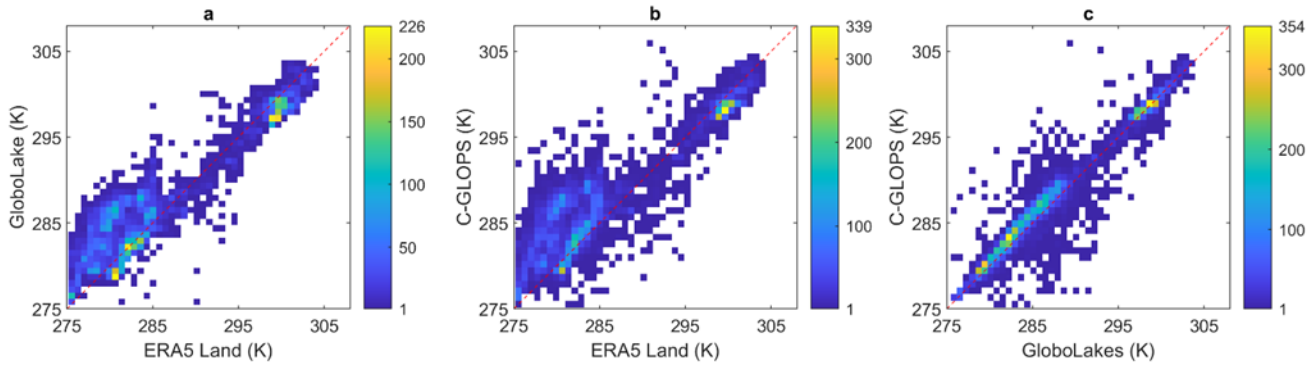


Figure 2.9. Scatter plots of 5-year averages lake temperature data. (**a**): GloboLakes versus ERA5 Land (N = 10111); (**b**). C-GLOPS versus ERA5 Land (N = 10111); (**c**). GloboLakes versus C-GLOPS (N = 10111). N represents the number of pixels with paired data. The number on the color bar represents the number of available data samples within each assigned temperature interval. Points closer to yellow mean more samples lie in that temperature range whereas these closer to blue mean less samples lie in that temperature range.

As shown in **Figure 2.10**, those statistical metrics were computed based on **Figure 2.10**a, d: daily ERA5 Land LMLT and daily GloboLakes LSWT, **Figure 2.10**b, e: 10-day ERA5 Land

LMLT and 10-day C-GLOPS, and **Figure 2.10**c, f: 10-day GloboLakes LSWT and 10-day C-GLOPS. In addition, their discrepancies were further conditioned by the lake cover fractions. Again, the composite 10-day LSWT from GloboLakes and C-GLOPS have shown great consistency in both absolute values and temporal variations because of the utilization of observations from AATSR (**Figure 2.10**c, f). In addition, ERA5 Land products have also exhibited high correlations with GloboLakes and C-GLOPS on R values near or even more than 0.9 (**Figure 2.10**d, e). LMLT tend to be lower than LSWT in the regions with low water coverage (0.05–0.25) whereas the scenarios are opposite over areas with more water bodies (**Figure 2.10**a, b). In terms of differences and R, data of ERA5 Land are closer to LSWT products when the lake cover fraction ranges from 0.05 to 0.5.

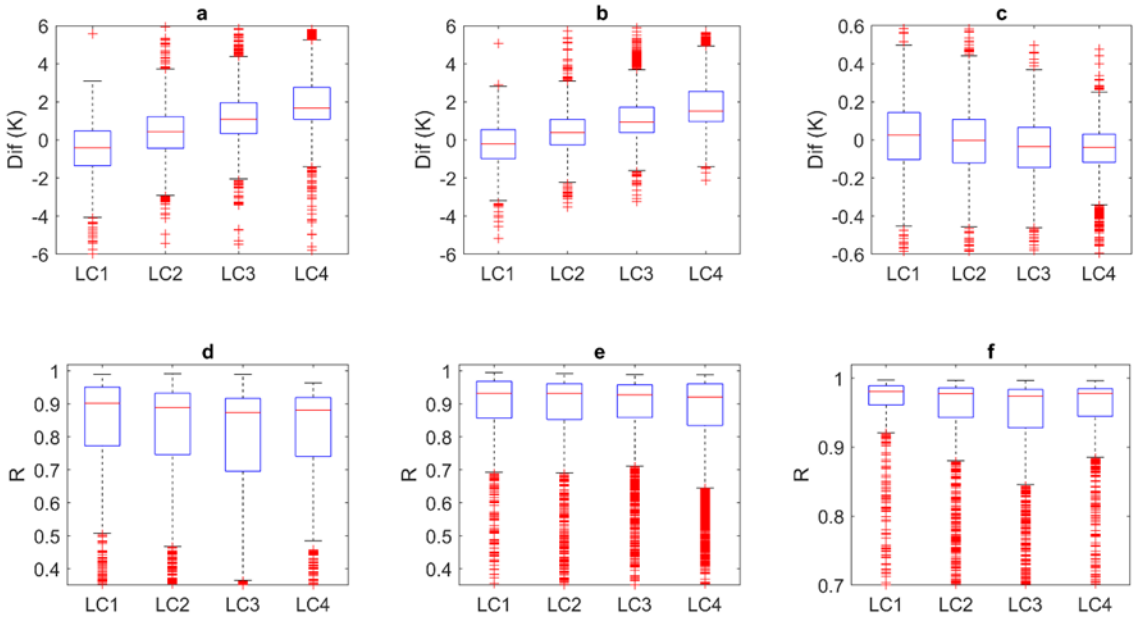


Figure 2.10. Boxplots of differences and R between lake temperature data sets at the 9 km EASE grid conditioned by lake cover (LC) fractions. 'Dif' denotes differences between two lake temperature data sets. (**a**): Dif = daily ERA5 Land LMLT–daily GloboLakes LSWT (N = 9925); (**b**): Dif = 10-day ERA5 Land LMLT–10-day C-GLOPS LSWT (N = 9044); (**c**): Dif = 10-day GloboLakes LSWT–10-day C-GLOPS LSWT (N = 8969); (**d**): daily ERA5 Land LMLT versus daily GloboLakes LSWT; (**e**): 10-day ERA5 Land LMLT versus 10-day C-GLOPS LSWT; (**f**): 10-day GloboLakes LSWT versus 10-day C-GLOPS LSWT. The intervals of lake cover percentage represented by LC1 (0.05–0.25), LC2 (0.25–0.5),

LC3 (0.5–0.75), and LC4 (0.75 -1). N represents the number of pixels in which paired data from two different data sets are more than 30.

2.5. Discussion

The impacts of varying spatial resolutions on the performances of LSWT and LMLT products can be studied by comparisons of their statistical metrics at their original resolutions and the 9 km EASE grids. Generally, metric values at the coarser resolution are within the same magnitude and comparable to those at their original spatial resolutions. Therefore, it is expected that any given resampling procedure only has limited effects on degrading quality of lake temperature data sets, especially for ERA5 Land whose native resolution is close to the 9 km EASE grid. Despite that, absolute values of median bias of C-GLOPS and GloboLakes at the 9 km EASE grid have slightly increased. This is partly because the resampled LSWT products have a coarser resolution, representing the LSWT over broader water areas, and become more deviated from in-situ data collected at a point scale.

Underestimations of LSWT products compared to the in-situ measurements conform to previous validation results (Carrea and Merchant 2020c; Crosman and Horel 2009; Lieberherr and Wunderle 2018; Schaeffer et al. 2018; Zhao et al. 2020). According to Carrea and Merchant (2020c), the differences between C-GLOPS with a quality level higher than 3 and in-situ measurements are −0.24 K ± 0.88 K, which is comparable to the results shown in **Table 2.2** (−0.31 K ± 0.93 K). A slightly larger bias could be partially caused by the consideration of all quality-level data here. In addition, all the lakes with positive biases presented in Carrea and Merchant (2020c) are not included in this study, possibly leading to a higher negative median bias. A negative 0.2 K error between the LSWT and in-situ data is considered to be normal due to the cool skin effect of surface water temperature relative to in-situ measurements of the sub-

surface (Carrea and Merchant 2020c). Similarly, systematic discrepancies among LMLT and LSWT products cannot be ignored since the water depths modeled by ERA5 Land (default 25 m (Sabater 2019)), and LSWT products (skin temperature) are different. Moreover, the mean absolute error of the ERA5 Land data set (0.1°) here is 2.83 K, close to the results shown in Muñoz-Sabater et al. (2021a) (2.25 to 3.22 K).

Despite LSWT being closer to in-situ observations in terms of absolute values, the spatial coverage of ERA5 Land is greater than the LSWT products mainly focusing on larger water bodies (Carrea and Merchant 2019, 2020b) given the differences in the available pixel numbers. Additionally, as observed in **Figure 2.4** and **Figure 2.5**, the provision of discrete or even sporadic satellite observations from GloboLakes and C-GLOPS could be insufficient to continuously reflect instantaneous thermal variations of inland water bodies required in the frame of soil moisture. As mentioned earlier, the quality of LMLT is less disturbed by the resampling procedures as well as the local weather conditions at a certain time. Moreover, ERA5 Land LMLT data are available from 1981 to present, compared to GloboLakes (1995–2016) or C-GLOPS with a gap period. Furthermore, consistently high temporal correlations of hourly LMLT and hourly in-situ measurements, daily LMLT and daily LSWT of GloboLakes, and 10-day LMLT and 10-day LSWT of C-GLOPS provide confidence in making LMLT closer to in-situ measurements by adopting proper rescaling approaches in the future. In light of these, ERA5 Land LMLT could be the optimal water temperature product used for water correction in soil moisture retrievals.

It should be noted that there are several limitations to this study. Firstly, the obtained evaluation results are based on lake temperature products from 2007 to 2011, which is only a partial portion of the temporal extent for each data set. In particular, the C-GLOPS product is separated into two

intervals using observations from different satellite sensors. There might be a slight underestimation related to the quality of the C-GLOPS product, given that the newly reprocessed C-GLOPS LSWT is aggregated using the observations from SLSTR instruments with a higher temporal frequency (Carrea and Merchant 2020c).

In addition, in-situ measurements considered here are all distributed in North America and thus unable to fully represent lake temperatures globally across various climatic and geophysical conditions. However, those areas with in-situ measurements pertain to Northern Cool in terms of lake thermal regions, representing more than 40% of total lake areas (Maberly et al. 2020). The evaluation results are thereby sufficient to indicate some aspects of the quality of considered LMLT and LSWT products. The retrieval method of LSWT products is based on physics and their stable performances are expected (Carrea and Merchant 2020a), and therefore the analyses of inter-comparison results between LMLT and LSWT data sets are of more importance. Nevertheless, it is still challenging to assess those grids with inland water bodies beyond the scopes of GloboLakes and C-GLOPS. Further-more, some areas contiguous to oceans have been excluded from the ERA5 Land data set. However, those pixels are also critical in the retrieval of global soil moisture. Therefore, plenty of products associated with sea surface temperature may be evaluated and compared with ERA5 Land LMLT in order to complement those coastal grids in future studies.

2.6. Conclusions

The accuracy of land surface emissions governs the quality of retrieved soil moisture products, and reasonable partitioning of water and land emissions from satellite-based observations requires accurate estimations of water temperature. In light of this, three newly released lake

temperature products, ERA5 Land, GloboLakes, and C-GLOPS, have been evaluated by comparing with in-situ observations as well as inter-comparisons among them from 2007 to 2011. Six statistical metrics have been selected to reflect the performance in aspects of temporal correlations and the proximity of absolute values. Overall, the LMLT of the ECMWF ERA5 Land product has been considered as the optimal option to be used in correction procedures of passive remote sensing soil moisture retrievals due to its wide spatial coverage, long-term consistent performance, less interference from resampling procedures, and the continuous provision of hourly updated data.

Generally, the three lake temperature data sets have comparable performances and adequate capacity to capture the dynamic variations of water temperatures (indicated by R values more than 0.8). The yielded differences between lake temperature products and in-situ measurements are 1.56 K ± 2.76 K, −0.27 K ± 0.86 K, and −0.31 K ± 0.93 K in the original spatial resolutions, and 1.36 K ± 2.68 K, −0.29 K ± 0.83 K, and −0.36 ± 0.92 K in the EASE 9-km scale for ERA5 Land, GloboLakes, and C-GLOPS, respectively. In light of these, the transfer of spatial resolution from their native scales to the 9-km EASE grids has not largely affected the assessment results. Moreover, the effects of temperatures on biases between lake temperature data sets and in-situ measurements are limited. Furthermore, median bias and RSD could be more appropriate to represent the quality of lake temperature products compared to the conventional metrics.

Evidently, the ERA5 Land product has advantages in both spatial coverage and temporal resolution for satisfying the requirements for soil moisture retrievals in which lake water temperatures (for example) at 6 a.m. and 6 p.m. are needed (to coincide with the SMAP overpassing time). In terms of 5-year averages over the studying period, LMLT values are overall

higher than LSWT data while the estimations of LSWT of GloboLakes and C-GLOPS are closer to each other, partially because of the utilization of observations from the same satellite sensor. Furthermore, the temporal variations of LMLT and LSWT products are highly correlated while their absolute values are closer over pixels with small water fractions in a range of 0.05 to 0.5.

Although different water depths are considered in LMLT and LSWT products as well as in-situ measurements, they exhibit similar patterns in illustrating the seasonal patterns and close values within 1.6 K to demonstrate the consistency of various considered data sets. Given those, the extensive spatial coverage, hourly updated lake temperature, and long-term availability, ERA5 Land based on the ECMWF H-TESSEL model is expected to be the best candidate for water correction in soil moisture retrievals. Additionally, this study could provide useful information related to lake temperature products for data users who are interested in the investigations of the thermal conditions of inland water bodies in the context of climate change.

2.7. Appendix

Table A1. Summary of in-situ measurements used in this study.

Index	Name/Country	Latitude	Longitude	Start Year	Buoy	Senor Depth (Meter Below Water Line)	Organization *
1	Superior/Canada-USA	48.06	−89.79	1979	45001	1.1	NDBC
2	Superior/Canada-USA	47.59	−86.59	1980	45004	1.3	NDBC
3	Superior/Canada-USA	47.34	−89.79	1981	45006	1.3	NDBC
4	Superior/Canada-USA	48.54	−86.95	1989	45136		MTU
5	Superior/Canada-USA	47.27	−88.61	2010	45023	3.0	MTU
6	Superior/Canada-USA	46.97	−88.40	2011	45025	3.0	UMD
7	Superior/Canada-USA	46.86	−91.93	2011	45027	1.0	UMD
8	Superior/Canada-USA	46.81	−91.83	2011	45028	1.0	NDBC
9	Huron/Canada-USA	45.53	−82.84	1980	45003	0.4	NDBC
10	Huron/Canada-USA	44.28	−82.42	1981	45008	1.3	ECCC
11	Huron/Canada-USA	45.54	−81.02	1989	45137		ECCC
12	Huron/Canada-USA	44.94	−80.63	1997	45143		ECCC
13	Huron/Canada-USA	43.54	−82.08	2000	45149		ECCC
14	Huron/Canada-USA	46.05	−82.64	1999	45154		NDBC
15	Michigan/USA	45.34	−86.41	1979	45002		NDBC
16	Michigan/USA	42.67	−87.03	1981	45007	1.3	UMC
17	Michigan/USA	45.41	−85.09	2010	45022	1.0	LT
18	Michigan/USA	41.98	−86.62	2011	45026	1.0	ECCC
19	Great Slave/Canada	61.18	−115.31	1992	45141		ECCC

20	Great Slave/Canada	61.98	−144.13	2004	45150		NDBC
21	Erie/Canada	41.68	−82.40	1980	45005	1.6	ECCC
22	Erie/Canada	42.74	−79.29	1994	45142		ECCC
23	Erie/Canada	42.46	−81.22	1989	45132		ECCC
24	Winnipeg/Canada	50.80	−96.73	1999	45140		ECCC
25	Winnipeg/Canada	53.23	−98.29	2004	45144		ECCC
26	Winnipeg/Canada	51.87	−96.97	2001	45145		NDBC
27	Ontario/Canada	43.62	−77.40	2002	45012	1.3	ECCC
28	Ontario/Canada	43.78	−76.87	1989	45135		ECCC
29	Ontario/Canada	43.23	−79.53	1991	45139		ECCC
30	Ontario/Canada	43.77	−78.98	2009	45159		ECCC
31	Woods/Canada	49.64	−94.50	2000	45148		ECCC
32	Saint Clair/Canada	42.43	−82.68	2000	45147		ECCC
33	Nipissing/Canada	46.23	−79.72	1999	45152		ECCC
34	Simcoe/Canada	44.50	−79.37	1999	45151		ECCC

** Organization represents those institution to install and maintain the corresponding buoys. NDBC: National Data Buoy Center; ECCC: Environmental and Climate Change Canada; MTU: Michigan Technological University; UMD: University of Minnesota, Duluth; UMC: University of Michigan CILER; LT: Limon Tech.*

Chapter 3. A Performance Analysis of Soil Dielectric Models Over Organic Soils in Alaska for Passive Microwave Remote Sensing of Soil Moisture

R. Zhang, S. Chan, R. Bindlish, and V. Lakshmi (2023). A Performance Analysis of Soil Dielectric Models Over Organic Soils in Alaska for Passive Microwave Remote Sensing of Soil Moisture. *Remote Sens*. **2023**, 15, 1658. https://doi.org/10.3390/rs15061658

Abstract

Passive microwave remote sensing of soil moisture (SM) requires a physically based dielectric model that quantitatively converts the volumetric SM into the soil bulk dielectric constant. Mironov 2009 is the dielectric model used in the operational SM retrieval algorithms of the NASA Soil Moisture Active Passive (SMAP) and the ESA Soil Moisture and Ocean Salinity (SMOS) missions. However, Mironov 2009 suffers a challenge in deriving SM over organic soils as it does not account for the impact of soil organic matter (SOM) on the soil bulk dielectric constant. To this end, we presented a comparative performance analysis of nine advanced soil dielectric models over organic soil in Alaska, four of which incorporate SOM. In the framework of the SMAP single-channel algorithm at vertical polarization (SCA-V), SM retrievals from different dielectric models were derived using an iterative optimization scheme. The skills of the different dielectric models over organic soils were reflected by the performance of their respective SM retrievals, which was measured by four conventional statistical metrics, calculated by comparing satellite-based SM time series with in-situ benchmarks. Overall, SM retrievals of organic-soil-based dielectric models tended to overestimate, while those from mineral-soil-based models displayed dry biases. All the models showed comparable values of unbiased root-mean-

square error (ubRMSE) and Pearson Correlation (R), but Mironov 2019 exhibited a slight but consistent edge over others. An integrated consideration of the model inputs, the physical basis, and the validated accuracy indicated that the separate use of Mironov 2009 and Mironov 2019 in the SMAP SCA-V for mineral soils (SOM < 15%) and organic soils (SOM ≥ 15%) would be the preferred option.

Keywords: Soil Moisture; Dielectric Models; SMAP; Soil Organic Matter

3.1. Introduction

Passive microwave remote sensing is considered the most suitable tool to map spatial soil wetness, owing to the negligible atmospheric influence and less interference from canopy and surface roughness (De Jeu et al. 2008; Njoku and Entekhabi 1996). The remarkable performance of soil moisture (SM) retrievals from spaceborne L-band radiometers (i.e., Soil Moisture and Ocean Salinity (SMOS) (Kerr et al. 2001) and Soil Moisture Active Passive (SMAP) (Entekhabi et al. 2010a)) has been substantiated by a number of validation studies (Chan et al. 2018; Chan et al. 2016; Colliander et al. 2017; Kim et al. 2018; Zhang et al. 2019). The mechanism that physically bridges the surface emission at microwave bands and surface SM is based on the contrasting difference between the dielectric constants of liquid water (~ 80) and dry soil (~ 4) (Ulaby et al. 1986a). The dielectric model that quantitatively links the SM with the bulk dielectric constant of the soil-water-air system is therefore critical in the retrieval algorithms of SMOS and SMAP.

Recently, numerous dielectric models were developed and applied for both spaceborne microwave radiometers and in-situ electromagnetic sensors (Bircher et al. 2016). An ideal dielectric model is envisioned, to accurately account for the dielectric response of wet soils as a function of all the relevant factors, including soil compaction, soil composition, the fraction of bound and free water, salinity, soil temperature, soil particle size distribution, and observation frequency, etc. (Dobson et al. 1985). However, the practical dielectric models are often established on a limited set of soil properties and are unable to approximate proper dielectric constants for all the surface conditions. Previous studies found that applying mineral-soil-based dielectric models over organic soils could lead to a substantial underestimation of SM (Bircher et al. 2016). Zhang et al. (2019) revealed a significant drop in the SMAP retrieval quality in regions

with soil organic carbon (SOC) exceeding 8.72%. Given that Mironov 2009 (Mironov et al. 2009) currently used in the SMOS and SMAP operation algorithms, was developed exclusively on samples of mineral soils, an update on the dielectric model that incorporates the effect of soil organic matter (SOM) is pressingly required for areas with organic-rich soils.

The influence of SOM on the bulk dielectric constant of the soil-water system is often summarized in two aspects. First, organic substrates have larger specific surface areas than minerals, indicating that organic soil has a higher fraction of bound water relative to mineral soil, when they contain the same amount of water (Bircher et al. 2016; Park et al. 2019; Wigneron et al. 2017). As such, at the same moisture, the dielectric constant of organic soil tends to be lower than that of mineral soil, as the dielectric constant of bound water is much smaller than that of free water. Second, organic soil is often marked with a larger porosity than mineral soil due to its complex structure (Bircher et al. 2016; O'Neill and Jackson 1990; Park et al. 2019; Wigneron et al. 2017). Based on these principles, several organic-soil-based dielectric models have been developed in recent years.

Although model developers pointed out the potential applicability of their models in the retrieval of SM, assessment of the efficacy of these newly developed organic-soil-based dielectric models in the derivation of passive microwave remote sensing of SM, has not been widely carried out. In light of these considerations, nine advanced dielectric mixing models were selected and tested in the context of the SMAP single-channel algorithm at vertical polarization (SCA-V) (O'Neill et al. 2021a). This study has two major objectives: 1) present the differences between the available mineral- and organic-soil-based models, in describing the complex dielectric behaviors of wet soils under various SOM conditions, and 2) evaluate their performances in organic-rich soils. The latter was achieved by comparing the SCA-V SM retrievals from different models against

in-situ measurements scattered over Alaska, where the soils are identified with noticeably higher SOM (~ 25%) relative to the global average level (Figure A1). The dielectric models considered here have been classified as the mineral-soil-based dielectric models, including Wang 1980 (Wang and Schmugge 1980), the semi-empirical Dobson 1985 modified by Peplinski 1995 (Dobson et al. 1985; Peplinski et al. 1995) (hereafter Dobson 1985), the prevalent Mironov 2009 (Mironov et al. 2009), Mironov 2012 (Mironov et al. 2012), and Park 2017 (Park et al. 2017), and organic-soil-based dielectric models, including the natural log fitting model in (Bircher et al. 2016) (hereafter Bircher 2016), Mironov 2019 (Mironov et al. 2019), Park 2019 (Park et al. 2019), and Park 2021 (Park et al. 2021).

As introduced earlier, five mineral-soil-based dielectric models were selected for a comprehensive survey of diverse models in the framework of the SMAP SCA-V algorithm over organic-rich soils. Two of them, Mironov 2013 and Park 2017, have not been widely examined under the SMOS and SMAP schemes. In contrast, the other three classic models have been extensively assessed in wide domains covered by mineral soils. However, their performances over regions with high SOM proportions have not been well-studied and compared with those of dedicated organic-soil-based models. In addition to water volume, mineral-soil-based primarily focus on the influence of soil texture, commonly characterized by sand, clay, and silt. Yet, organic-soil-based models place a greater emphasis on the SOM effect. Mironov 2019, for example, describes all parameters as functions of SOM rather than the clay percentage used in Mironov 2009. Therefore, incorporating more mineral- and organic-soil-based models may also help to construct an impression of their systematic differences when describing the dielectric behaviors of organic soils.

Table 3.1. Input variables required for nine dielectric models.

Model Inputs	Mineral Soil Based Models					Organic Soil Based Models			
	Wang1980	Dobson1985	Mironov2009	Mironov2013	Park2017	Bircher2016	Mironov2019	Park2019	Park2021
Soil Moisture	Volumetric Soil Moisture (m^3/m^3)	Volumetric Soil Moisture (m^3/m^3)	Volumetric Soil Moisture (m^3/m^3)	Volumetric Soil Moisture (m^3/m^3)	Volumetric Soil Moisture (m^3/m^3)	Volumetric Soil Moisture (m^3/m^3)	Gravimetric Soil Moisture (g/g)	Volumetric Soil Moisture (m^3/m^3)	Volumetric Soil Moisture (m^3/m^3)
Soil Organic Matter	/	/	/	/	/	/	Gravimetric Soil Organic Matter (%)	Gravimetric Soil Organic Matter (%)	Gravimetric Soil Organic Matter (%)
Clay	Gravimetric Clay Fraction (0-1)	Gravimetric Clay Fraction (0-1)	Gravimetric Clay Fraction (%)	Gravimetric Clay Fraction (%)	Volumetric Clay Fraction (0-1)	/	/	Volumetric Clay Fraction (0-1)	Volumetric Clay Fraction (0-1)
Sand	Gravimetric Sand Fraction (0-1)	Gravimetric Sand Fraction (0-1)	/	/	Volumetric Sand Fraction (0-1)	/	/	Volumetric Sand Fraction (0-1)	Volumetric Sand Fraction (0-1)
Silt	/	/	/	/	Volumetric Silt Fraction (0-1)	/	/	Volumetric Silt Fraction (0-1)	Volumetric Silt Fraction (0-1)
Bulk Density	Bulk Density (g/cm^3)	Bulk Density (g/cm^3)	/	/	/	/	Bulk Density (g/cm^3)	/	/
Frequency	/	Frequency (Hz)	Frequency (Hz)	/	Frequency (Hz)	/	/	Frequency (Hz)	Frequency (Hz)
Salinity	/	/	/	/	Salinity (‰)	/	/	Salinity (‰)	Salinity (‰)
Soil Temperature	/	Soil Temperature (℃)	/	Soil Temperature (℃)	Soil Temperature (℃)	/	Soil Temperature (℃)	Soil Temperature (℃)	Soil Temperature (℃)
Total Number of Inputs	4	6	3	3	7	1	4	8	8

3.2. Data

3.2.1. SMAP L2 Radiometer Half-Orbit 36km EASE-Grid Soil Moisture, Version 8

Launched on January 31, 2015, the SMAP mission was designed to map high-resolution SM and freeze/thaw state by combining the attributes of L-band radar and radiometer. However, the SMAP SM products presently rely on the radiometer's observations alone, due to an unexpected malfunction of the SMAP radar in July 2015. With an average revisit frequency of two to three days, the SMAP sensors cross the Equator at the local solar time of 6 a.m. and 6 p.m.

SMAP L2 Radiometer Half-Orbit 36 km EASE-Grid Soil Moisture, Version 8 (SMAP V8) (O'Neill et al. 2021b) was adopted in this study. Here, we only used the descending (6 a.m.) SM retrievals derived using the SCA-V algorithm. A series of masking procedures were utilized to avoid the application of SM retrievals of low accuracy and high uncertainty. Specifically, only the retrievals flagged as the 'recommended quality' were retained and employed in the later analysis. Given Alaska, the focused region of this study, locates at the high-latitude portion with a long-term frozen duration, we only considered those qualified SM retrievals within the time intervals from June to August between 2015 and 2021.

One noticeable improvement in the SMAP V8 (relative to an older version) is the update and extension of gridded soil parameters, ranging from SOC, silt and sand fractions to bulk density. These newly added soil attributes originate from the SoilGrid 250m (Hengl et al. 2017) and replace the earlier patched version composed of the National Soil Data Canada (NSDC), the State Soil Geographic Database (STATSGO), the Australia Soil Resources Information System (ASRIS), and the Harmonized World Soil Database (HWSD) (Das and O'Neill 2020). Since

these soil attributes are often necessary inputs for dielectric models of soil, they were also extracted from the SMAP V8.

3.2.2. In-Situ Soil Moisture Measurements

Ground-based SM measurements over Alaska were employed as benchmarks to assess the skills of diverse dielectric mixing models. Historical files of soil water content observed by in-situ sensors were first downloaded from the Natural Resources Conservation Service (NRCS), the National Water and Climate Center (NWCC) homepage (https://www.nrcs.usda.gov/wps/portal/wcc/home, accessed on 7 April 2022). At present, there are more than 40 operating stations from the Snow Telemetry (SNOTEL) (Schaefer and Paetzold 2001) and Soil Climate and Analysis Network (SCAN) (Schaefer et al. 2007). These stations are able to monitor the sub-daily variations of SM and many other climatic variables in near-real time.

However, some typical errors (Schaefer and Paetzold 2001) of in-situ SM readings, such as breaks and plateaus, have been found before their application. As a response, the other authoritative data source of in-situ SM, the International Soil Moisture Network (ISMN) (Dorigo et al. 2021; Dorigo et al. 2011a), was also considered, aiming at incorporating its flag information. Given the limited stations in Alaska, it is expected that SM data from the above two sources (NWCC and ISMN) are mostly from the same set of stations. Additionally, for the same station, the observed SM time series from the NWCC and ISMN should be identical, as the ISMN only gathers data and harmonizes them in units and time steps, without extra data processing. Given the frequent abnormal SM readings (even after adopting the quality flag) and the necessity of checking the consistency of SM measurements from two different sources,

several rigorous pre-checking procedures were applied (as described in Section 3.1) to filter out those suspicious observations where possible in advance.

3.3. Methodology

3.3.1. Preliminary Examination of In-Situ Measurements

The quality of in-situ SM data is of great importance, as these ground measurements are generally seen as the benchmark for evaluating remotely sensed and/or modeled SM data sets (Chan et al. 2018; Chan et al. 2016; Colliander et al. 2017). However, monitoring SM dynamics over high-latitude regions is still challenging due to the long-term frozen periods and harsh environments. Such difficulties have been reflected by the flat limbs and breaks frequently occurring in the SM time series from the Alaskan stations. Given those, a careful examination of in-situ SM measurements is necessary.

The general workflow of the preliminary examination steps is delineated in **Figure 3.1**. Specifically, the in-situ SM data measured at the local time of 6 a.m. and 6 p.m. (temporally align with the SMAP overpass time) were first extracted from the NWCC and ISMN stations. SM measurements with the corresponding land surface temperature below 4 ℃ were excluded, as Colliander et al. (2017) demonstrates that some sensors begin to behave abnormally under this temperature. Meanwhile, the utilization of such a threshold would also be helpful to filter out those SM measurements likely obtained during a period of active thawing and re-freezing, where SM fluctuations are excessively unstable (e.g., Figure 3c in Dorigo et al. (2013)). Additionally, stations with a distance shorter than 36 km to large water bodies or oceans were also masked, as the SMAP SM over those regions is likely influenced by water contamination. The flag information from the ISMN was also incorporated to filter in-situ data of low quality.

The matched SM data of the overlapped stations from the NWCC and ISMN are anticipated, and the greater consistency further enhances the reliability of these benchmarks. Therefore, an automatic consistency checking procedure constrained by three requirements was applied. Since breaks and plateaus still appeared on the SM time series after consistency checking, a manual visual inspection was then performed to screen those suspicious measurements. After those, there are 21 qualified stations left, and we assume that their SM data from the NWCC and ISMN are interchangeable. Furthermore, pairing with the SMAP observations removed 9 stations, and the remaining 12 stations (**Figure S1**) would be used in the later validation steps.

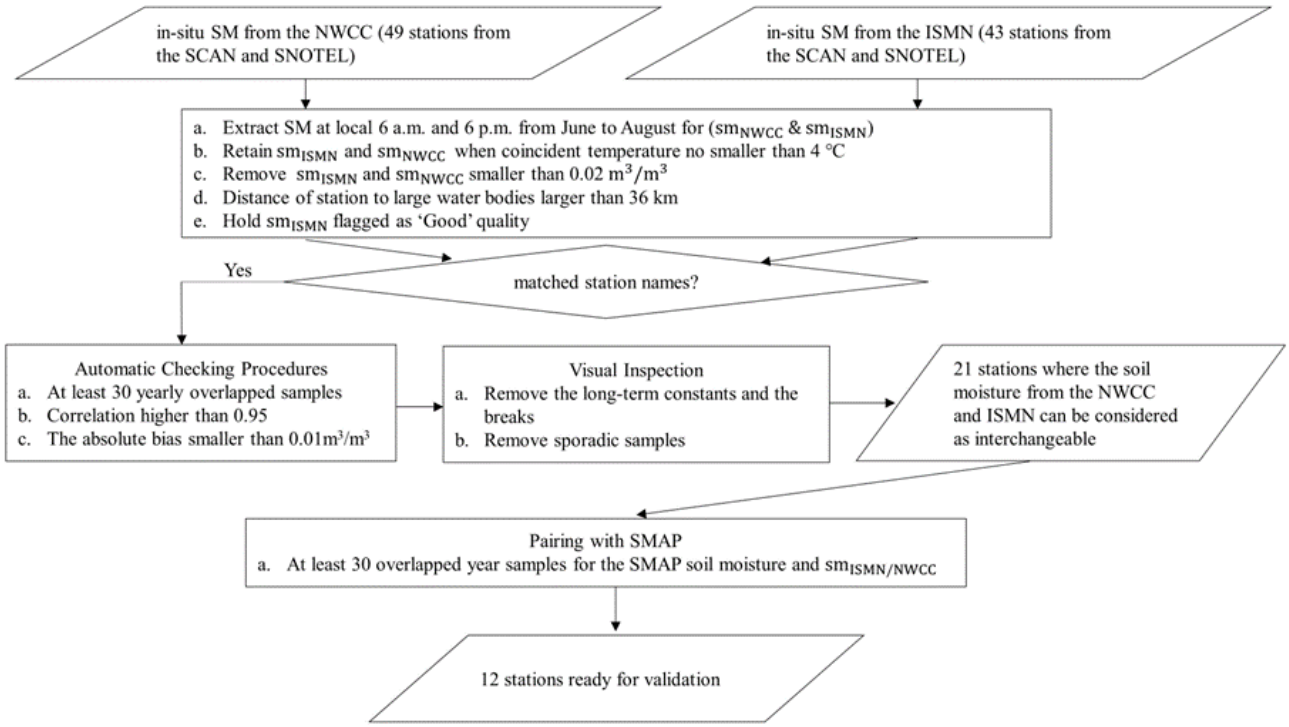


Figure 3.1. The flow chart of the preliminary examination on the Alaskan in-situ soil moisture obtained from the NWCC and ISMN

3.3.2. Derivation of Soil Moisture from Various Dielectric Models

In the SCA-V algorithm, the SMAP SM value is determined when there is a minimized difference between the simulated and the observed reflectivity (r_{smap}) (reflectivity = 1 –

emissivity) of smooth soil. At each temporal step, the value of r_{smap} over a pixel is fixed as the SMAP SCA algorithm have determined the radiative contribution from the canopy layer and the impact of surface roughness before subtracting them from SMAP observed surface brightness temperature (T_B). Hence, the influence of adopting different dielectric constant models on SM retrievals can be examined using the iterative feedback-loop procedure, to minimize the difference between the simulated reflectivity (r_{est}) and r_{smap}, and without the need to construct the whole process from SM to T_B, in consideration of simplicity.

However, r_{smap} is an intermediate product and unavailable in the original SMAP data set. Given this, the values of r_{smap} were first estimated leveraging SMAP SM and Mironov 2009. With these benchmarks, the SM retrievals of other dielectric models were then acquired based on the optimization flow described in **Figure 3.2**. Notably, the SM retrieval at a given time point is reproducible while the identical r_{smap} and model are used.

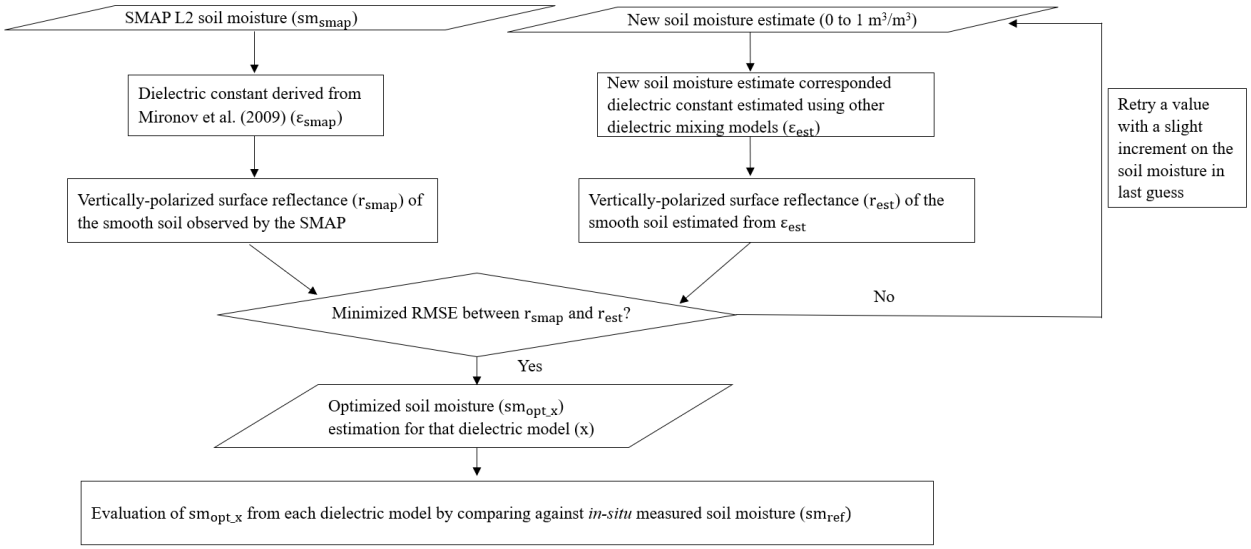


Figure 3.2. Flow chart that describes the retrieval of soil moisture using different dielectric models based on the identical SMAP observations.

3.3.3. Performance Metrics

The skill of the remote sensing SM data set has been described by four conventional metrics, which are bias, root-mean-square error (RMSE), unbiased root-mean-square error (ubRMSE), and the Pearson correlation (R) (Entekhabi et al. 2010b). These metrics could effectively reflect the discrepancies in terms of magnitudes as well as the links of the temporal evolutions between the SM estimations and the ground truth. The formulas used to compute these metrics are shown from Eq 1 to Eq 4 where E […] represents the arithmetic mean; σ_{opt} and σ_{ref} denote the standard deviations of SM retrievals of the respective dielectric model and in-situ measured SM.

$$\text{bias} = \text{E}[\text{sm}_{\text{ret}}] - \text{E}[\text{sm}_{\text{ref}}] \tag{1}$$

$$\text{RMSE} = \sqrt{\text{E}[(\text{sm}_{\text{ret}} - \text{sm}_{\text{ref}})^2]} \tag{2}$$

$$\text{ubRMSE} = \sqrt{\text{RMSE}^2 - \text{bias}^2} \tag{3}$$

$$\text{R} = \frac{\text{E}[(\text{sm}_{\text{ret}} - \text{E}[\text{sm}_{\text{ret}}])(\text{sm}_{\text{ref}} - \text{E}[\text{sm}_{\text{ref}}])]}{\sigma_{\text{ret}}\sigma_{\text{ref}}} \tag{4}$$

3.4. Results and Discussion

3.4.1. Simulated Brightness Temperature of Smooth Soil through Synthetic Experiments

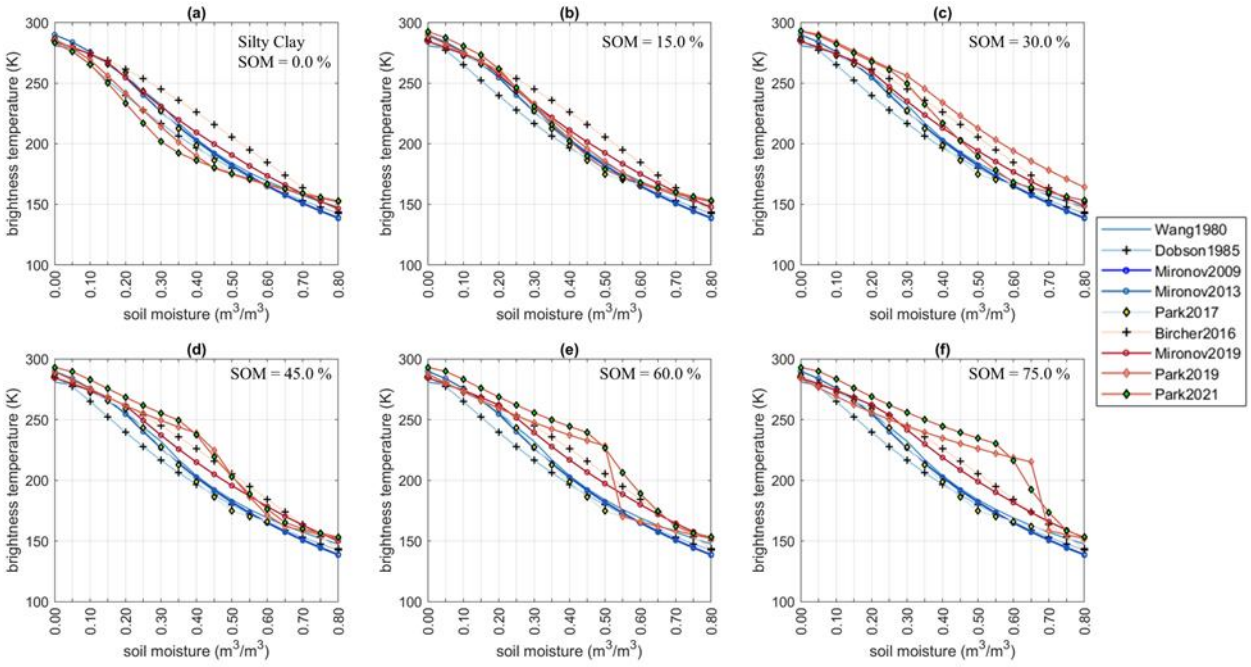
Synthetic experiments have the capability to afford complete dielectric responses to a whole SM range, by artificially controlling all the inputs required for the dielectric models (**Table 3.1**). With SOM increasing from 0% to 75% at a step of 15%, the differences between the dielectric constants estimated by mineral- and organic-soil-based dielectric models were explored. These various dielectric responses were further transferred to their corresponding thermal radiations of smooth soils, represented by the vertically polarized T_B.

Figure 3.3 presents the T_B curves derived using different dielectric models across the range of SM from 0 to 0.8 m^3/m^3. Generally, the T_B values estimated using organic-soil-based models are greater than those derived using the mineral-soil-based models particularly when SOM exceeds 15% and SM is higher than 0.1 m^3/m^3. In other words, the SM retrievals from organic-soil-based models tend to be wetter than the SM retrievals from mineral-soil-based models (e.g., Mironov 2009) given the same surface reflectivity (or T_B) of bare, smooth soil. The discrepancies between the simulated T_B magnitudes from mineral- and organic-soil-based models further grow with the increase of SOM (**Figure 3.3**). However, it should be noted that the estimated dielectric constants and their subsequent T_B values from mineral-soil-based models do not vary with SOM. The higher SM estimations of organic-soil-based models relative to mineral-soil-based models could be attributed to the fact that those organic-soil-based models assume a higher volumetric proportion of bound water (Bircher et al. 2016; Park et al. 2019; Wigneron et al. 2017). When SOM is at 15% (and below), the simulated T_B curves from all the considered models are

clustered together, bounded by Dobson 1985 and Bircher 2016 (**Figure 3.3b**). Therefore, SOM of 15% might be treated as an appropriate demarcation point for the separate use of mineral- and organic-soil-based dielectric models over mineral soils and organic soils.

Moreover, similar features of the T_B curves of those considered dielectric models have been observed while a sandy sample is tested (**Figure S2**). Such a stable-magnitude discrepancy between the red curves (organic models) and the blue curves (mineral models) under contrasting textures (sandy and clay soils) can be attributed to the insensitivity of the organic-soil-based dielectric models to soil texture. For example, Mironov 2019 only accounts for the effects of soil moisture, SOM, and soil temperature on the dielectric permittivity of organic soils. Although Park 2019 and Park 2021 incorporate both textural and SOM information, the differences in their estimated T_B values from sandy and clay samples seem insignificant under the same SOM level.

Compared to Mironov 2019, the influence of organic content on the simulated T_B magnitude seems more pronounced for Park 2019 and Park 2021. When SOM increases from 0% to 75% and SM values are smaller than 0.5 m^3/m^3, the T_B curve of Park 2021 jumps from the bottom one to the top line, with a varying amplitude on the order of tens of Kelvins (**Figure 3.3**). In contrast, as a response to growing SOM, the estimations from Mironov 2019 slowly move upward approaching the T_B curve of Bircher 2016. According to **Figure 3.3**e and f, there is a rapidly dropping segment on the T_B curve of Park 2019. Such abnormal dielectric behavior can be attributed to the improper formulas used to calculate the wilting point and porosity, with a detailed explanation in Section 3.4.4.



Model Inputs	Gravimetric Sand Fraction (%)	Gravimetric Clay Fraction (%)	Gravimetric Silt Fraction (%)	Bulk Density (g/cm^3)	Salinity (‰)	Soil Physical Temperature (°C)	Frequency (GHz)	Soil Organic Matter (%)	Volumetric Soil Moisture (m^3/m^3)	Incidence Angle (°)
Simulated Values	5.0	47.6	47.4	1.42	0.0	20.0	1.4	0.0, 15.0, 30.0, 45.0, 60.0, 75.0	0.0 – 0.8	40.0

Figure 3.3. Simulated brightness temperature of a silty clay with various soil organic matter, and the accompanied table displays all the input values where most of soil parameters are directly taken from the sample of silty clay used in Hallikainen et al. (1985).

3.4.2. Evaluation of Dielectric Models over In-Situ Sites in Alaska

Here, SM measurements from 12 sites served as benchmarks to evaluate the skills of multiple dielectric models in the setting of SMAP observations and its SCA-V algorithm. Before inter-comparison, it has been found that the assessment metrics of the satellite-based SM retrievals over the same pixel could vary a lot in different years. Using the time series in Monument Creek as an instance (**Figure 3.4**), R values range from 0.18 (2017) to 0.69 (2015). Hence, the obtained metrics (**Table 3.2**, **Table 3.3**, and **Table 3.4**) averaged over multiple years of each station might be underrated as they may be compromised by abnormal behavior in one year. Additionally, the amplitudes and frequencies of in-situ SM variations are often more pronounced relative to the SM retrievals as the latter reflects the changes over a coarse spatial extent (**Figure 3.4**). SM

variations at local scales often cannot be captured by the 36 km-scale SM retrievals, due to the omission of spatial variability within the footprint-scale area. Spatial mismatching between satellite SM retrievals and point-scale in-situ measurements could adversely impact the perceived accuracy of SMAP observations.

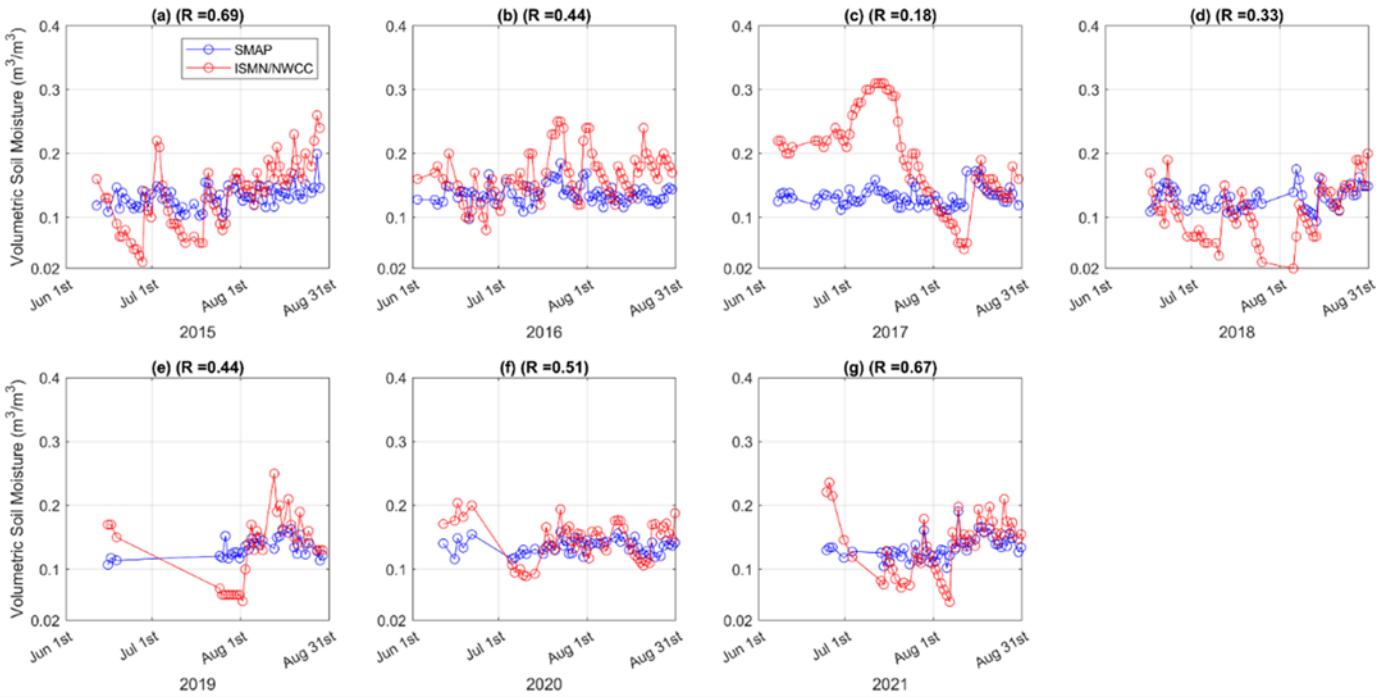


Figure 3.4. Time series of soil moisture derived from satellite observations and *in-situ* measurements at Monument Creek (65.18º N, 145.87º W). (a-g) describe the soil moisture variations of SMAP retrievals and ground measurements from 2015 to 2021.

Assessment metrics of the SM retrievals derived using identical r_{smap} values and different dielectric models were computed by their temporally paired in-situ measurements. According to **Table 3.2**, SM estimates from mineral-soil-based models tend to underestimate while organic-soil-based models generally exhibit wet biases compared to ground recordings. In terms of both ubRMSE and R (**Table 3.3** and **Table 3.4**), all the models show comparable accuracy levels similar to previous results of Mialon et al. (2015) whereas Mironov 2019 displays a slight but consistent edge over other models. Compared to other dielectric models, the modest

improvement in R of Mironov 2019 is likely due to its simultaneous consideration of bulk density and SOM effects (Mironov et al. 2019)

Table 3.2. Bias of soil moisture retrievals using various dielectric models over in-situ sites in Alaska where biases from mineral- and organic-soil based models tend to underestimate and overestimate relative to in-situ measurements.

Station/Bias (m^3/m^3)	N	Mineral Soil Based Models					Organic Soil Based Models			
		Wang1980	Dobson 1985	Mironov 2009	Mironov 2013	Park 2017	Bircher 2016	Mironov 2019	Park 2019	Park 2021
Gulkana River	72	0.058	**0.025**	0.046	0.044	0.039	0.195	0.142	0.104	0.085
Spring Creek	37	-0.108	-0.153	-0.137	-0.137	-0.139	**-0.022**	-0.051	-0.105	-0.109
Atigun Pass	81	0.047	**-0.002**	0.015	0.016	0.009	0.092	0.092	0.044	0.061
Coldfoot	156	-0.085	-0.133	-0.121	-0.121	-0.124	**-0.030**	-0.036	-0.083	-0.067
Eagle Summit	320	-0.028	-0.068	-0.062	-0.061	-0.068	**0.014**	0.017	-0.033	-0.015
Gobblers Knob	262	0.031	-0.010	**-0.003**	-0.003	-0.007	0.096	0.083	0.039	0.055
Monahan Flat	121	-0.047	-0.093	-0.076	-0.077	-0.081	0.035	**0.009**	-0.029	-0.029
Monument Creek	405	0.018	-0.022	**-0.014**	-0.014	-0.016	0.091	0.073	0.029	0.041
Mt. Ryan	194	0.114	**0.078**	0.082	0.082	0.080	0.196	0.172	0.132	0.142
Munson Ridge	383	0.018	-0.019	**-0.015**	-0.015	-0.016	0.096	0.075	0.034	0.045
Tokositna Valley	253	0.014	-0.008	**-0.006**	-0.008	-0.008	0.147	0.093	0.062	0.046
Upper Nome Creek	283	-0.138	-0.180	-0.171	-0.171	-0.176	**-0.086**	-0.091	-0.138	-0.120
Mean	214	-0.009	-0.049	-0.038	-0.039	-0.042	0.069	0.048	**0.005**	0.011

Where the column of the number tagged by bold font represents the dielectric model with the smallest absolute bias in that station or mean.

Table 3.3. ubRMSE of soil moisture retrievals using various dielectric models over in-situ sites in Alaska.

Station/ubRMSE (m^3/m^3)	N	Mineral Soil Based Models					Organic Soil Based Models			
		Wang 1980	Dobson 1985	Mironov 2009	Mironov 2013	Park 2017	Bircher 2016	Mironov 2019	Park 2019	Park 2021
Gulkana River	72	**0.0132**	0.0164	0.0156	0.0154	0.0152	0.0209	0.0180	0.0169	0.0138
Spring Creek	37	0.0460	0.0457	0.0452	0.0454	0.0455	**0.0408**	0.0428	0.0446	0.0462
Atigun Pass	81	0.0311	0.0311	0.0311	0.0311	0.0311	0.0317	0.0311	0.0310	**0.0310**
Coldfoot	156	0.0736	0.0736	0.0736	0.0736	**0.0736**	0.0743	0.0737	0.0739	0.0737
Eagle Summit	320	0.0487	0.0490	0.0487	0.0487	0.0487	0.0480	**0.0477**	0.0482	0.0481
Gobblers Knob	262	0.0665	0.0663	0.0660	0.0662	0.0662	**0.0622**	0.0643	0.0628	0.0637
Monahan Flat	121	0.0722	0.0721	0.0720	0.0721	0.0721	**0.0714**	0.0718	0.0715	0.0722
Monument Creek	405	0.0510	0.0509	0.0508	0.0508	0.0508	0.0505	0.0503	0.0504	**0.0503**
Mt. Ryan	194	**0.0163**	0.0177	0.0173	0.0172	0.0173	0.0262	0.0186	0.0237	0.0187
Munson Ridge	383	0.0499	0.0492	0.0490	0.0492	0.0492	**0.0465**	0.0475	0.0467	0.0478
Tokositna Valley	253	0.1295	0.1296	0.1295	0.1295	0.1296	0.1298	**0.1294**	0.1296	0.1296
Upper Nome Creek	283	**0.0122**	0.0126	0.0124	0.0123	0.0126	0.0196	0.0129	0.0163	0.0160
Mean	214	0.0509	0.0512	0.0509	0.0510	0.0510	0.0518	**0.0507**	0.0513	0.0509

Where the column of the number tagged by bold font represents the dielectric model with the best ubRMSE in that station or mean.

Table 3.4. R of soil moisture retrievals using various dielectric models over in-situ sites in Alaska.

Station/R	N	Mineral Soil Based Models					Organic Soil Based Models			
		Wang 1980	Dobson 1985	Mironov 2009	Mironov 2013	Park 2017	Bircher 2016	Mironov 2019	Park 2019	Park 2021
Gulkana River	72	0.605	0.596	0.607	0.604	0.599	0.608	**0.621**	0.603	0.601
Spring Creek	37	0.757	0.737	0.758	0.752	0.745	0.757	**0.805**	0.752	0.746
Atigun Pass	81	0.342	**0.348**	0.344	0.344	0.344	0.341	0.333	0.347	0.347
Coldfoot	156	0.205	0.205	0.204	0.204	0.205	0.206	0.199	0.202	**0.208**
Eagle Summit	320	0.375	0.353	0.372	0.376	0.368	0.376	**0.429**	0.368	0.372
Gobblers Knob	262	0.571	0.557	0.571	0.570	0.564	0.571	**0.603**	0.575	0.577
Monahan Flat	121	0.276	0.273	0.275	0.274	0.274	0.277	0.275	**0.284**	0.276
Monument Creek	405	0.407	0.401	0.406	0.405	0.404	0.409	0.413	0.406	**0.418**
Mt. Ryan	194	0.604	0.595	0.604	0.601	0.599	0.605	**0.624**	0.604	0.601
Munson Ridge	383	0.608	0.597	0.606	0.604	0.602	0.610	**0.624**	0.611	0.611
Tokositna Valley	253	0.177	0.171	0.174	0.172	0.170	0.172	**0.176**	0.172	0.171
Upper Nome Creek	283	0.416	0.398	0.418	0.420	0.410	0.416	**0.477**	0.421	0.416
Mean	214	0.445	0.436	0.445	0.444	0.440	0.446	**0.465**	0.445	0.445

Where the column of the number tagged by bold font represents the dielectric model with the best R in that station or mean.

The other aspect that we attempted to evaluate the predictive power of various dielectric models was checking the correlations between the SM retrievals of different models and SMAP observed vertically polarized T_B. If the higher absolute R values between the time series of SM and SMAP vertically polarized T_B are assumed as a criterion that reflects the better skill of a dielectric mixing model, Mironov 2019 presents an overwhelming superiority over other models in the 765 Alaskan pixels (**Figure 3.5**). **Table S2** displays that in-situ measured SM usually has a lower correlation with SMAP vertically polarized T_B relative to correlations between satellite-based SM retrievals and SMAP T_B. However, it should be noted that such correlation-based results were inconclusive and functioned as a reference only since the impacts of vegetation disturbance and surface roughness were entirely ignored.

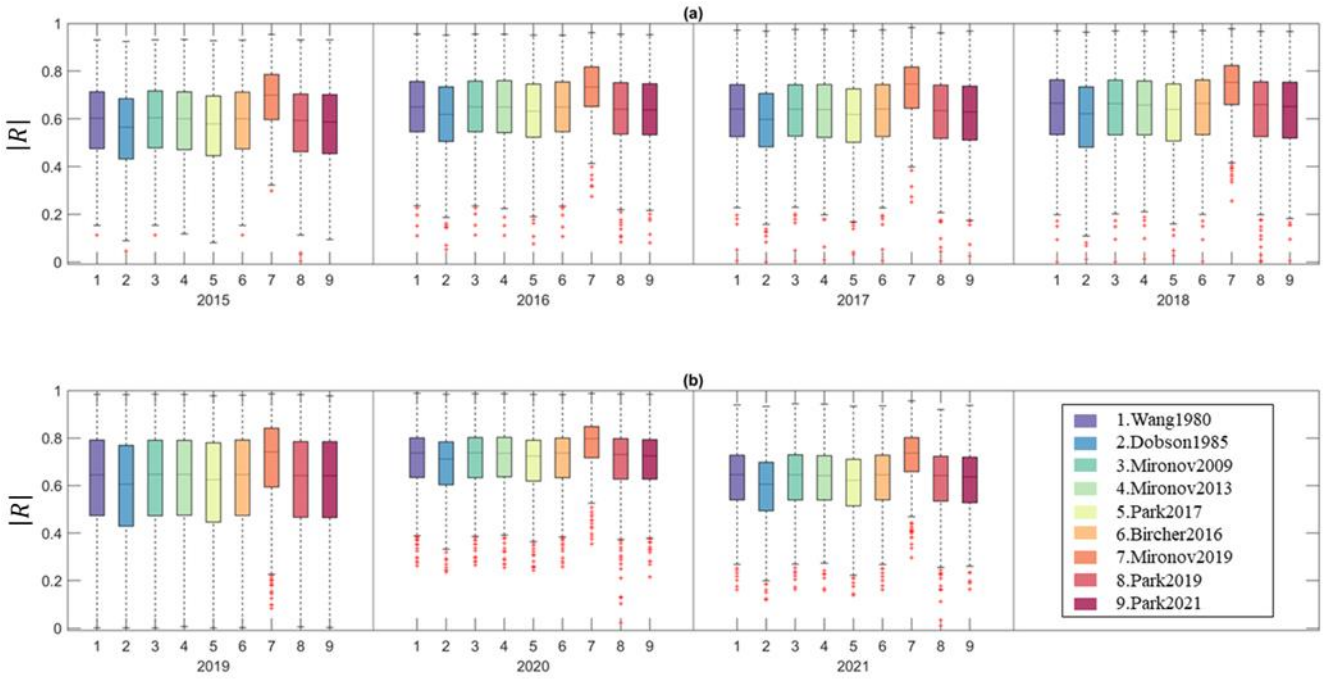


Figure 3.5. Boxplots of the absolute correlations between the soil moisture retrievals from various dielectric mixing models and the SMAP vertically polarized brightness temperature over the 765 pixels in Alaska. (a) and (b) represent the boxplots of absolute R values from 2015 to 2018 and from 2019 to 2021, respectively.

3.4.3. A Global Intercomparison between Mironov 2009 and Mironov 2019

Mironov 2009 and Mironov 2019 were selected as the representatives for mineral- and organic-soil-based dielectric models and were then compared with each other at the global scale using one-week SMAP observations from July 2, 2018, to July 8, 2018. The one-week SM retrievals of Mironov 2009 and Mironov 2019 were analyzed over more regions with abundant SOM and were also used to acquire performance clues for applying Mironov 2019 in mineral soils.

According to **Figure 3.6**a and b, satellite-based SM data are usually unavailable in many areas characterized by organic-rich soils likely owing to dense boreal forests, harsh surface roughness, as well as permanently frozen soils on the land surface (Bircher et al. 2016; Yi et al. 2019). The magnitude difference between Mironov 2009 and Mironov 2019 yielded SM retrievals are commonly above 0.05 m^3/m^3 generally when SOM is over 10% (**Figure 3.6**b and e). In the case of extreme dryness (SM < 0.1 m^3/m^3) over mineral soils (SOM < 5%), SM retrievals from

Mironov 2019 are likely lower than those from Mironov 2009. As illustrated in **Figure 3.6**d, there is a limb where SM retrievals of Mironov 2019 are nearly constant while those from Mironov 2009 vary, possibly because of soil texture.

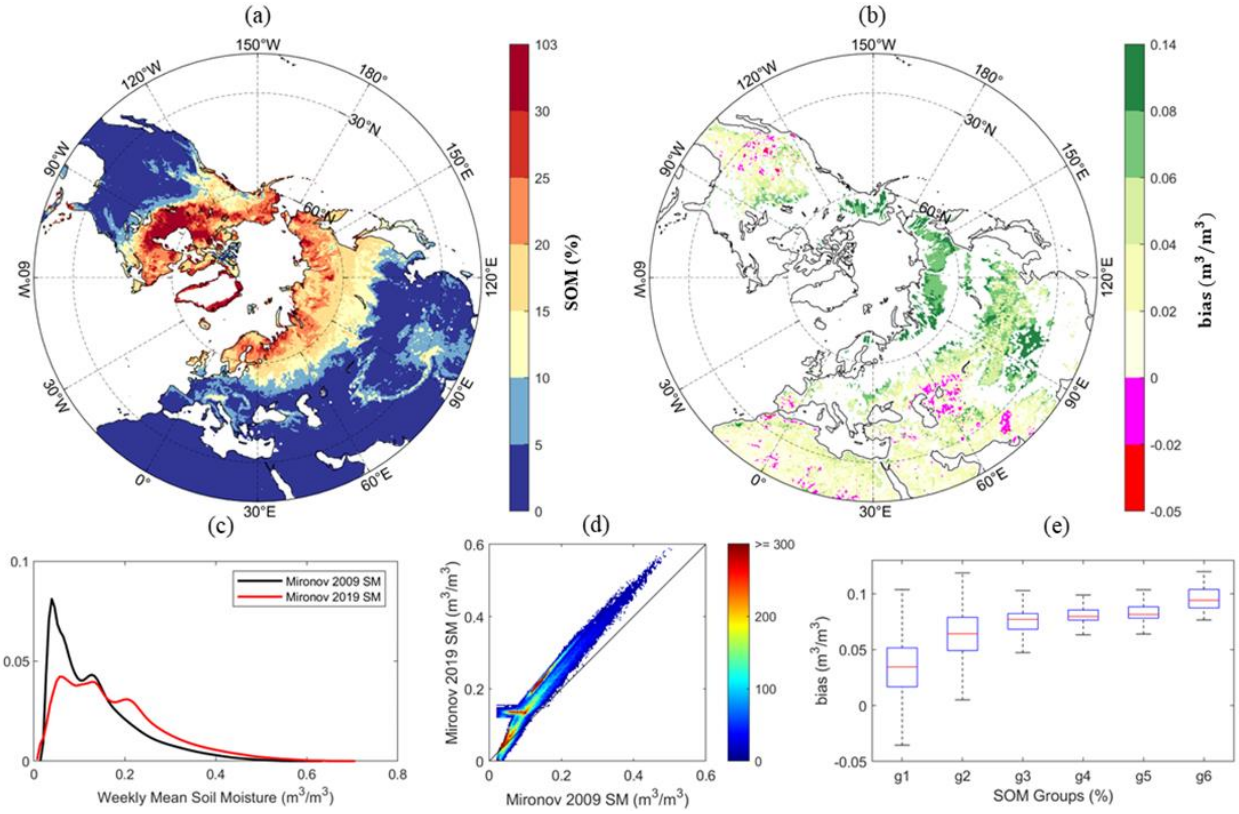


Figure 3.6. A global intercomparison of soil moisture retrievals from Mironov 2009 and Mironov 2019 where (a) the spatial distribution of soil organic matter (SOM) in a north polar view, (b) the spatial distribution of mean differences between soil moisture estimations using Mironov 2009 and Mironov 2019 (bias = $SM_{Mironov2019} - SM_{Mironov2009}$), (c) the probability distribution function of weekly mean soil moistures derived using the above two models, (d) the scatterplot of soil moisture using both models across the globe, and the color bar shows the number of pixels, and (e) the boxplot that describes the bias variations along with the increase of SOM that was already organized into 6 groups (g1 - g6). The organic range of each group is 0% - 5% (g1), 5% - 10% (g2), 10% - 15% (g3), 15% - 20% (g4), 20% - 30% (g5), and > 30% (g6).

3.4.4. Discussion

3.4.4.1. The Applicable Range of Dielectric Models

Although the above validation results over in-situ sites in Alaska demonstrated slightly better performance of Mironov 2019 over other models, it may be not the best model across all

landscapes and climatic conditions. The accuracy of a dielectric model heavily depends on its respective applicable range. A dielectric model is likely to acquire a better performance score when being applied over samples used to develop it. In other scenarios, potential degradation of the model skills can be expected. For instance, when Dobson 1985 is adopted in soils that fall beyond the prototypal soils on which Dobson 1985 was established, some unrealistic dielectric constants could be yielded (Mironov et al. 2009). According to SMAP configurations and parameters, the frequency is confined to 1.4 GHz. Most pixels in Alaska show SOM values spanning from 15% to 30%. However, it should be noted that Mironov 2019 is designed for the surface soil layer with SOM ranging from 35% to 80% (Mironov et al. 2019). Meanwhile, the natural log calibration function from Bircher et al. (2016) is proposed for highly organic soils and Decagon 5TE (in-situ sensor) which is operated at 70 MHz. Such imperfect alignments between the applicable ranges of dielectric models and the actual settings are surprisingly common, possibly leading to underestimations of the quality of these dielectric models.

3.4.4.2. Organic-Soil-Based Dielectric Models

Similar to other empirical dielectric models (Kellner and Lundin 2001; Malicki et al. 1996; Paquet et al. 1993; Roth et al. 1992; Skierucha 2000; Topp et al. 1980) accounting for the influence of SOM, SOM itself is not treated as a necessary input in Bircher 2016 to derive the dielectric constants of organic soils. Mironov 2019, however, incorporates the dielectric impacts of SOM and soil bulk density while omitting the clay fraction. In contrast, Park 2019 and Park 2021 consider both mineralogy and SOM. Though comprehensive, the confidence in representing the dielectric interactions among various soil properties and the quality of those global-scale soil databases greatly limit the practical uses of Park models. For example, SOM as the most critical index to classify mineral and organic soils was estimated by multiplying SOC

content with a fixed factor of 1.724 (Mironov et al. 2019; Pribyl 2010) However, the conversion factor between SOC and SOM is unlikely a global constant while Pribyl (2010) pointed out that this conversion factor would vary from 1.4 to 2.5 across different geographical regions.

Additionally, mineral-soil-based dielectric models are usually based on the assumption that the soil is composed of sand, silt, and clay, and thus the summation of their fractions is 100%. However, this assumption is likely inappropriate over organic-rich soils where SOM has a grate gravimetric contribution. Here, the texture fractions extracted from the SoilGrids 250m were normalized. As a result, the summation of minerals and SOM currently exceeds 100%, while a further re-normalization is difficult to proceed with, as the SOM contents (sometimes over 100%) were empirically estimated. Despite these issues, at this time, these data sets might be the most practical sources to support running those dielectric models over a wide spatial coverage. Therefore, a soil property data set that can accurately describe the gravimetric relationship among sand, silt, clay, and SOM is pressingly needed.

3.4.4.3. Limitations of In-Situ Benchmarks

Besides the limits of the model applicable range and the quality of input data sets of soil properties, the other critical factor that directly affects the assessment results is the quality of the benchmarks, i.e., in-situ SM measurements. As mentioned, breaks, missing values, and jumps were commonly found during the examination of in-situ SM time series. Furthermore, many calibration functions used to deduce in-situ SM values are designed for mineral soils only due to the unavailability of organic-soil-based calibration functions over those regions. As a result, in-situ SM values might have an underestimation issue.

Due to the limited availability of in-situ measurements over Alaska, only one ground station was selected as the regional benchmark for each validation pixel. However, the estimated SMAP retrieval performance over these areas was likely degraded given the unmatched spatial representatives and measuring depths between the passive microwave SM derivations and ground measurements. Additionally, inconsistent SM variations from the radiometer snapshot and the ground sensors may have arisen during the transition period between two years (e.g., from the end of August 2015 to the beginning of June 2016), adversely affecting the validation metrics. In spite of these factors, this study presents an evaluation that maximizes the use of existing data sets and can serve as a valuable reference for further investigations as more data become available.

3.4.4.4. Characteristics of Park Models

Compared to other conventional semi-empirical dielectric models (Dobson et al. 1985; Mironov et al. 2012; Mironov et al. 2019; Park et al. 2017; Park et al. 2019; Wang and Schmugge 1980), Park models describe the fractions of bound water and free water differently (Park et al. 2017; Park et al. 2021; Park et al. 2019). First, Park models use the wilting point as the beginning point where free water starts to occur whereas other models set that value using an independent term named maximum bound water fraction. When the volumetric SM is between the maximum bound water fraction and porosity, most dielectric models fix the bound water content and the dielectric contribution of bound water. However, in the same SM range, Park models assume that the content of bound water and free water alters with the volumetric SM. Specifically, SM is treated as a weighted summation of the bound water and free water, where the sum of the weights of bound water (w_b) and free water (w_f) is constrained as one. It is assumed that w_b is

one when SM is equal to wilting point. On the contrary, w_b declines to zero when SM reaches porosity.

According to **Figure 3.3**e and f, there are a few rapid drops in the curves of Park 2019 and Park 2021 when SOM exceeds 60%. Such scenarios could be explained by the wilting-point and porosity calculation equations used in Park 2019 and Park 2021. As shown in **Figure S3**, the porosity equation of Park 2019 could lead to a porosity greater than 1 m^3/m^3 when SOM ranges from 30% to 35%. Meanwhile, in Park 2019, the derived wilting point could surpass the porosity when SOM is over 60%. Although the above issues have been substantially mitigated for Park 2021 with valid magnitudes of its derived porosity and wilting point, an evident bending near the wilting point could still be observed in its simulated T_B curves at highly organic soils. Therefore, caution should be paid when applying Park 2019 and Park 2021 over organic-rich soils.

3.4.4.5. Selection of A Globally Optimal Combination of Dielectric Models

In general, Mironov 2019 can be concluded as the prime dielectric model for use in the SMAP SCA-V algorithm over organic-rich soils. Such a determination was not only yielded from the validation results, but also incorporated the input parameters and configurations of various models. Specifically, Mironov 2019 requires fewer input parameters compared to Park 2019 and Park 2021, making it less susceptible to the uncertainties introduced by different soil property data sources, while accounting for the SOM effects. Additionally, Mironov 2019 was developed based on a physically refractive mixing dielectric model, where the parameters were calibrated and validated across several soil samples, with a SOM ranging from 35% to 80%. In contrast, Bircher 2016 was derived from straightforward regression analyses between two measured variables, while Park 2019 and Park 2021 lack effective calibration. Furthermore, Mironov 2019 consistently demonstrated a slight edge over the other models, in terms of the averaged ubRMSE and R. This accuracy advantage of Mironov 2019 would likely extend to other regions with organic-rich soils, given similar climatic conditions and vegetation types with Alaska.

While the operational SMAP retrieval algorithms apply a single dielectric model globally, finding a universal dielectric model that outperforms the other models across all possible conditions seems overambitious. As described above, mineral-soil-based dielectric models do not include the SOM effect on soil dielectric constants, whereas organic-soil-based models often ignore the influence of soil texture. Although Park 2019 and Park 2021 consider both soil texture and SOM, they are prone to higher errors, due to a few improper formulations and excessive uncertainties introduced by various input data sources. Hence, based on the previous studies and the results obtained here, the separate use of Mironov 2009 and Mironov 2019 in the SMAP SCA-V algorithm over mineral and organic soils is proposed. The selection of utilizing Mironov 2009 is somewhat arbitrary, as Mironov 2009 has not been comprehensively assessed against Mironov 2013 and Park 2017 over mineral soils. The applicability of Mironov 2009 has been extensively validated, and the use of Mironov 2009 will not further degrade the current retrieval quality.

The simultaneous use of Mironov 2009 and Mironov 2019 requires a sophisticated SOM threshold that can demarcate mineral and organic soils. However, there is presently no rigorous set of rules for this threshold. Mironov et al. (2019) states that soil can be categorized into organic soil if the SOM is more than 20%, whereas Broll et al. (2006) and Zanella et al. (2011) declare that organic soil should contain a SOM of at least 30%. According to the results of the synthetic experiments, a SOM of 15% might be an optimal threshold for distinguishing soil types, as the T_B curves of different models are closely clustered and the divergence between mineral- and organic-soil-based models seems to start after a SOM exceeding 15%. Such a threshold conforms to Huang et al. (2008) who classifies soils into organic soil or highly organic soil when the SOM is more than 15%.

The utilization of an optimal organic-soil-based dielectric model (i.e., Mironov 2019 here) is anticipated to improve the overall precision of the SMAP SM retrievals over organic soils. Since SM is a crucial factor in determining carbon fluxes in boreal regions, having precise knowledge of SM variations can effectively monitor the health of local ecosystems and predict the trends in carbon storage. In the current context of global warming, the snow extent has rapidly dropped in the Northern Hemisphere. Consequently, more snow-covered regions become bare soils, and the period of thawing seasons tends to last longer. Hence, decreasing SM retrieval uncertainties over

these high-SOM areas would greatly aid in tracking the potential significant hydrologic shifts triggered by climate change and permafrost thawing.

Meanwhile, the deficiencies in the quality of soil property products and in-situ data sets in the Northern environment have been identified. For instance, the universal conversion formula between SOC and SOM is still rudimentary, occasionally leading to an estimation over 100%. As such, the limitations discovered in this study offer a strong motivation and direction for developing soil property data sets with better applicability. Additionally, the necessity for accurate SM in high-latitude areas highlights the need for more ground stations and dense SM observation networks over the circumpolar zone.

3.4.4.6. Future Work

Here, the determination of the SOM threshold at 15%, based solely on synthetic experiments, likely caused spatial inconsistencies at the boundary of the mineral and organic soils. Hence, location/time-dependent SOM thresholds may be necessary to produce smooth SM maps in high-latitude regions. An alternative approach would be the mixed use of mineral- and organic-soil-based models over each pixel, provided that an accurate relative proportion of SOM and clay is available in advance.

Although this study evaluated various dielectric models under the SMAP SCA-V algorithm, their use in other radiative transfer model-based algorithms and with observations from different polarizations, angles, and frequencies remains to be investigated. Of particular interest is the dual-channel algorithm (DCA), the current SMAP baseline algorithm, which exhibited moderate edges over agricultural sites. The objective of the DCA algorithm is to achieve the optimal vegetation optical depth (VOD) and SM simultaneously, by minimizing the aggregated differences between the simulated and observed brightness temperatures at both horizontal and vertical polarizations. Thus, the alternation of the dielectric model could indirectly affect the derived vegetation water content. In addition to passive microwave remote sensing, the dielectric mixing model is also critical for other fields, such as SMAP L4 and the European Centre for Medium-Range Weather Forecasts (ECMWF) Community Microwave Emission Model (CHEM). Radar sensors also require a dielectric model to simulate the backscatter coefficients. However, there is currently no clear consensus on the best dielectric model for these platforms, making further investigations necessary and valuable.

3.5. Conclusions

In this study, the skills of nine dielectric models over organic soil in Alaska have been evaluated and compared in the context of the SMAP SCA-V algorithm. Four out of nine models carefully account for the SOM effect on the complex dielectric constant of the soil-water mixtures while the remaining models were designed for use in mineral soils. The dielectric responses (expressed in a form of T_B) of those models to the increasing SOM were comprehensively investigated through artificially controlling input values. At a given SM over 0.1 m^3/m^3 and SOM higher than 15%, the simulated T_B values from organic-soil-based dielectric models are greater than those estimated from mineral-soil-based dielectric models. In other words, relative to mineral-soil-based dielectric models, organic-soil-based models are inclined to obtain higher SM estimates from the identical observed radiations. Furthermore, a SOM threshold of 15% was suggested for the separate use of mineral- and organic-soil-based dielectric models in the retrieval algorithm as the divergence of T_B curves of mineral- and organic-soil models was observed when SOM exceeds 15%.

The predictive power of each dielectric model is represented by several statistic metrics computed by comparing its SM retrievals with in-situ measurements. Compared to satellite products reflecting SM variations over a large spatial extent, in-situ point-based SM measurements exhibited more temporal variability. Additionally, even over the same place, the annual correlations between satellite-based SM retrievals and in-situ data would fluctuate a lot. Consistent with the results from synthetic experiments, organic- and mineral-soil-based models tended to induce wet and dry biases. In an integrated evaluation, Mironov 2019 presented a

slightly but consistently better performance over other dielectric models, which showed a mean ubRMSE of 0.0507 m^3/m^3 and a mean R of 0.465.

Furthermore, an inter-comparison between SM retrievals within a one-week time interval from mineral- and organic-soil-based dielectric models was conducted at a global scale. Such a comparison would be useful to capture clues about the performance of organic-soil-based models over mineral soils. Mironov 2009 and Mironov 2019 were elected as the representatives of mineral- and organic-soil-based models, respectively. As a result, SM estimates from Mironov 2019 were at least 0.05 m^3/m^3 higher than those from Mironov 2009. When SM is below 0.1 m^3/m^3, SM retrievals from Mironov 2019 were occasionally smaller than SM retrievals from Mironov 2009 in mineral soils.

It should be noted that the performance of each dielectric model heavily depends on its designed application range, the quality of input data sets, as well as the accuracy of in-situ benchmarks. Different assessment results might be obtained with the update of dielectric models, in-situ measurements, and soil parameters. As such, a routine evaluation study that incorporates all the potential dielectric models and the most recent soil auxiliary data sets is recommended. In an integrated consideration of model inputs, the model physical foundation, and the practical accuracy, the separate use of Mironov 2009 and Mironov 2019 in the SMAP SCA-V algorithm for mineral soils (SOM < 15%) and organic soils (SOM $\geqslant$ 15%) would be the optimal option at this time. Considering the SOM magnitudes at the 36 km scale, developing a sophisticated dielectric model accounting for variable SOM from 10% to 30% is expected for passive microwave remote sensing of SM.

Chapter 4. Temporal Gap-Filling of 12-Hourly SMAP Soil Moisture over the CONUS Using Water Balance Budgeting

R. Zhang, S. Kim, H. Kim, B. Fang, A. Sharma, and V. Lakshmi (2023). Temporal Gap-filling of the SMAP-based 12-hourly Soil Moisture Product Over CONUS using Water Balance Budgeting. *Water Resource Research*, major revision.

Abstract

The presence of temporal gaps is an inevitable concern in using satellite-based soil moisture (SM) products. In this study, an entirely observation-based method has been developed to derive volumetric soil water content for filling the existing gaps in Soil Moisture Active Passive (SMAP) retrievals. Through a water balance equation, each 12-hour variation of water amount in the topsoil layer is ascertained based on the observed precipitation using the Global Precipitation Measurement Mission (inflow) and the hydrologic loss function (outflow) built on SMAP dry-downs. A temporally seamless SM product composed of SMAP dry-downs and the precipitation-driven moisture approximations (PQD) was generated as a secondary outcome in ascertaining the optimal parameters required for water balance budgeting. The PQD dataset preserves the features of the original SMAP SM dynamics given the median Pearson correlation (R) at 0.69 and the unbiased root-mean-square error (ubRMSE) at 0.05 m^3/m^3. Next, all the obtained parameters and available SMAP observations were used to produce the SMAP-based 12-hourly SM product (PQF) over the conterminous United States. By validating against in-situ measurements, the PQF not only exhibits good performance given a median R of 0.63 but it also captures most of the SM peaks induced by heavy rainfall events. Additionally, the validity of the proposed scheme has been further verified by the comparable performance of the exclusive filled-on SM estimates.

The availability of a continuous SM dataset (i.e., PQF) and its paired hydrologic losses could advance the quantification connections among the hydrologic components and benefit the understanding of land-surface hydrology.

4.1. Introduction

Soil moisture (SM) is a variable of great importance in its capability of influencing land-atmosphere interactions and its critical role in the hydrologic cycle (Koster et al. 2004; Koster et al. 2010; Petropoulos 2013; Seneviratne et al. 2010). Satellite-based passive microwave remote sensing has been identified as a reliable tool to monitor the temporal variations of surface SM at global scales, striking a balance between measurement accuracy, geographical coverage, and cost-effectiveness. In practice, SM observations from various spaceborne microwave sensors have been widely implemented to improve understanding of the process of Earth's systems (e.g., climate variability and drought detection), and to facilitate anthropogenic development strategies (e.g., water resource management and agricultural monitoring) (Fang et al. 2021a; Findell et al. 2011; Ge et al. 2011; Koster et al. 2010; Miralles et al. 2014; Taylor et al. 2012). However, the presence of spatiotemporal gaps in satellite-based SM observations is a limiting factor in their applications. Such a deficiency is primarily attributed to the intrinsic limitation of sun-synchronous orbits, instrumental configurations, and the applicable ranges of retrieval algorithms (Wang et al. 2012).

Despite an observation available every 1 to 3 days, and a gradual evolution of SM over a large spatial scale, satellite-based SM datasets are often unable to provide a complete picture of hydrologic process and likely ignore critical land responses to short-term weather extremes. A typical example is depicted in **Figure 4.1** where SM from the Soil Moisture Active Passive

(SMAP) (Entekhabi et al. 2014) misses a number of positive increments in wetness generated by the precipitation relative to in-situ SM records. Previous studies showed that SM sampled at such a coarse temporal frequency may bias the modeled water and energy fluxes which follow as well as could underestimate its antecedent precipitation volume (Brocca et al. 2019; Crow and Wood 2002). Therefore, advanced interpretation of SM dynamics and the increased flash hydroclimatic extremes (e.g., flash drought and flood events) require a continuous SM dataset at a sub-daily time frame.

Various methods have been developed to derive frequently updated, gapless SM products by temporally and/or spatially filling the voids contained in the original satellite observations. Such approaches can be broadly assigned into two categories. The first type of gap-filling study estimates the missed points by analyzing the varying patterns from the available measurements, such as the simple linear interpolation with a short time window (Brocca et al. 2019). Advanced statistical modeling techniques, such as the three-dimensional optimization underlying the discrete cosine transforms, ordinary kriging, and the convolutional neural network, are often adopted to achieve geophysical records of spatially complete and temporally seamless (ElSaadani et al. 2021; Kim et al. 2016; Pham et al. 2019; Wang et al. 2012; Zhang et al. 2021a). However, gap predictions relying on the spatial or temporal modes from observed SM alone could inherit the errors of the original products and incapable of handling unusual meteorological events (Zhang et al. 2021a). The second route to increase SM availability is to combine soil moisture information from several different satellite sensors, such as the European Space Agency Climate Change Initiative (ESA CCI) SM dataset and the SMOSSMAP-IB product (Dorigo et al. 2017; Li et al. 2022). Additionally, the low-earth orbit satellite constellation mission known as Cyclone Global Navigation Satellite System (CYGNSS), which observes L-band GNSS signals

of opportunity, has been utilized to retrieve SM with higher temporal repeats (i.e., sub-daily). However, CYGNSS-based SM retrievals showed unsatisfactory precision/accuracy compared to passive microwave satellite systems, and they were limited to their spatial coverages (38º N to 38º S) (Kim and Lakshmi 2018; Ruf et al. 2018).

Precipitation tends to induce an increase in surface SM at a pace greatly faster than the dry-down rate. Hence the incorporation of varying synchronous precipitation trends could aid in filling SM gaps. Given those, the main objective of this study is to fill SMAP's temporal gaps over the conterminous United States (CONUS) using the temporally high-resolution precipitation product, Global Precipitation Measurement Mission (GPM), from 2015 to 2021. Through water balance budgeting, the existing SM gaps were filled by the estimations converted from their corresponding 12-hour GPM precipitation volumes. This conversion was accomplished with the scheme initially proposed by Akbar et al. (2018) where the hydrologic terms other than precipitation and SM were treated as the aggregated loss and were quantified via SMAP dry downs. Here, we improved that scheme by adding parametric-based loss modeling and advanced ancillary datasets for better achieving our research goal rather than deriving the regionally hydrologic length scales that were found in Akbar et al. (2018). Through this, these precipitation-driven SM data preserves the magnitudes of the original SMAP retrievals. Compared to previous filling approaches, the proposed scheme is remarkedly simplified for utilization and could provide integrated hydrological loss quantifications in addition to high-quality SM estimates. Meanwhile, this newly yielded continuous dataset is envisioned to track the footprints of precipitation events and manifest as peaks in SM. The feature of totally observation-based simulation would be an exceptional benefit.

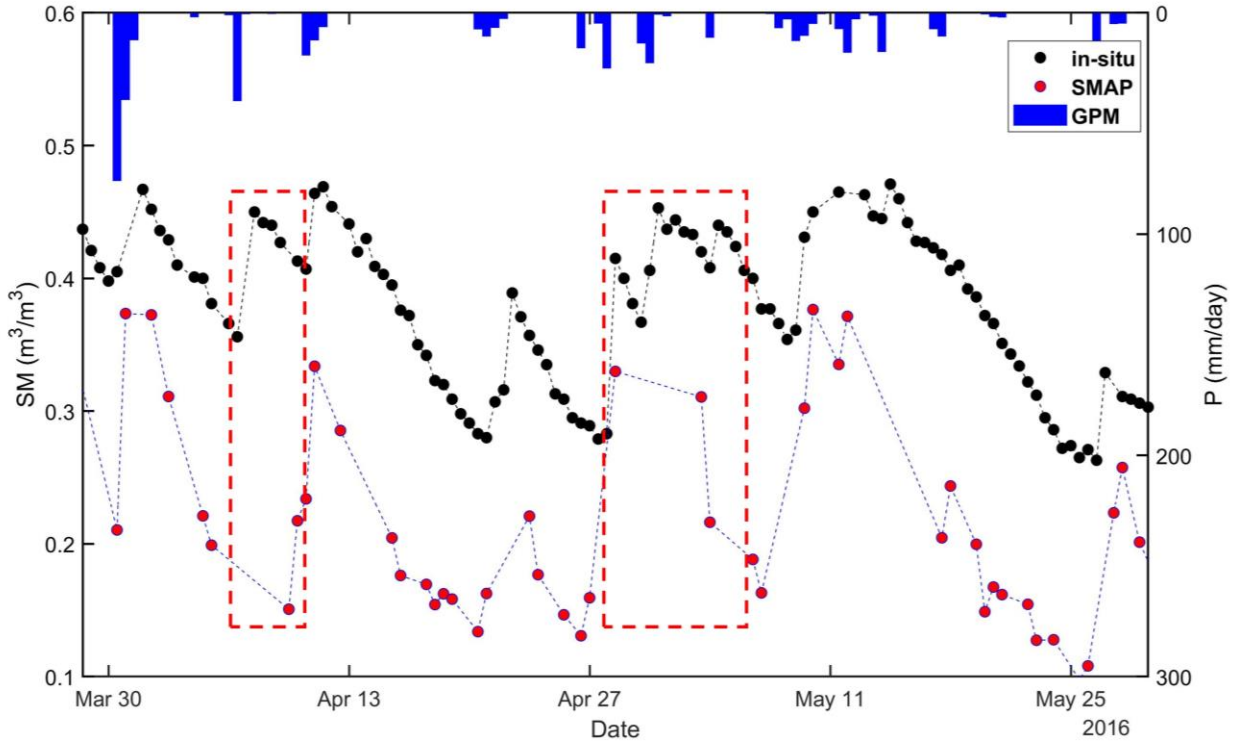


Figure 4.1. An example of the soil moisture time series from SMAP observations and in-situ measurements from the station at [40.05º N, 88.37ºW]. The intervals highlighted by the red boxes reflect that soil moisture responses to rainfall events on April 6 and April 30, 2016, have been completely missed and monotonically decreased. Such scenarios frequently occur due to the temporally sparse sampling of SMAP observations and simplify a number of up-and-down variations as dry-down trends. The omission of these details will misguide researchers in the process interpretation and result in underestimations of hydrologic dynamics.

4.2. Data

As summarized in **Table 4.1**, various datasets covering the period of six years (April 1, 2015 – March 31, 2021) have been adopted in this study. These include 1) the National Aeronautics and Space Administration (NASA) SMAP SM product, 2) the reanalysis SM dataset of the land component of the 5th generation of European Re-Analysis (ERA5-Land) developed by the European Centre for Medium-Range Weather Forecast (ECMWF) (Muñoz-Sabater et al. 2021a), 3) in-situ SM measurements from 1084 stations of the International Soil Moisture Network (ISMN) (Dorigo et al. 2021; Dorigo et al. 2013), 4) in-situ SM measurements from 40 stations of

the Texas Soil Observation Network (TxSON) (Caldwell et al. 2019), and 5) half-hourly precipitation estimates from the Integrated Multi-SatellitE Retrievals for Global Precipitation Measurement (GPM IMERG) (Hou et al. 2014).

Given the objective of building a 12-hourly continuous SM dataset at a UTC timescale, the local-time-based SMAP retrievals were interpolated into the closest time slots of UTC 00:00 and 12:00. This adjustment process was constructed on a crucial hypothesis of invariable SM within a (+/-) 3-hour interval. Accordingly, the hourly available ERA5-Land and the in-situ SM data synchronized with the UTC 00:00 and 12:00 were extracted and retained for validation. After temporal processing, the gridded GPM IMERG and ERA5-Land products were then resampled into the Equal-Area Scalable Earth (EASE) 9-km scale to be compatible with the SMAP spatial resolution.

Table 4.1. Summary of datasets used in this study.

Variable	Product Name	Spatial Resolution	Unit	Reference
Volumetric Soil Moisture	SMAP L3 enhanced soil moisture product (version 5)	9 km	m^3/m^3	O'Neill et al. (2021c)
	ISMN	Point	m^3/m^3	Dorigo et al. (2021)
	TxSON	Point	m^3/m^3	Caldwell et al. (2019)
	ERA5-Land 0-7cm soil moisture	0.1º	m^3/m^3	Muñoz-Sabater et al. (2021a)
Precipitation	GPM IMERG Final Precipitation Level-3 Half-hourly Product (version 06B)	0.1º	mm/hour	Huffman et al. (2019)

4.2.1. SMAP Soil Moisture

The SMAP mission was launched on January 31, 2015, by NASA for quantifying the representative water content at the top of a 5 cm soil column and detecting freeze/thaw states at a quasi-global scale (Entekhabi et al. 2014). The SMAP sensor crosses the equator constantly at

around 6 a.m. and 6 p.m. (local solar time) and monitors SM variations with a revisit frequency of 1 to 3 days (O'Neill et al. 2021c). In order to satisfy the research requirements of hydrometeorology and hydroclimatology, SMAP originally intended to incorporate the attributes of active and passive microwave sensors to provide high-resolution SM retrievals. However, the malfunction of the SMAP radar in July 2015 hampered the initial goals. Alternatively, the Backus-Gilbert optimal interpolation technique is adopted on the oversampled measurements of the SMAP radiometer to derive an enhanced SM product posted at the 9-km EASE grids (O'Neill et al. 2021c).

In this study, the SMAP Enhanced Level-3 Radiometer Global Daily 9-km EASE-Grid Soil Moisture (version 5) product (hereinafter referred to as SMAP) has been selected. The SMAP Level-3 product is a daily composite of the half-orbit SMAP Level-2 products where the SM retrieval algorithm is conducted. An integrated consideration of both geographical coverage and the quality of the filled-in SM estimations, a series of filtering procedures have been adopted before being temporally interpolated into the closest UTC 00:00 and 12:00. Specifically, the regions of vegetation water content (VWC) below 7 kg/m^2 are retained to include more areas of eastern CONUS (Akbar et al. 2018). It should be noted that this VWC of 7 kg/m^2 is less restricted relative to the recommended threshold of 5 kg/m^2 (O'Neill et al. 2021a). In contrast, a more rigid water fraction threshold of lower than 1% of water bodies within each 9-km pixel has been applied. Moreover, the pixel-wise soil porosity values are computed using the bulk densities under the SMAP ancillary dataset (Das and O'Neill 2020) as they are required by building the loss functions in the later part (Section 4.3.2).

4.2.2. In-Situ Soil Moisture Measurements

In-situ measurements are often acknowledged as the most reliable source and widely performed as benchmarks to assess remotely sensed SM retrievals and model estimations. The ISMN (Dorigo et al. 2021; Dorigo et al. 2013), as a centralized data platform, regularly compiles and reconciles ground SM observations from different networks all over the world. A series of filtering procedures on those ground-based SM records have been carried out to guarantee validity of the assessment. The TxSON measurements were firstly harmonized into UTC, aligning with the ISMN data. Subsequently, those observations symbolized as 'good' quality and with the measuring depth shallower than 10 cm were retained. Meanwhile, stations with effective SM samples of less than 30 were excluded.

Although the evaluation results based on the sparse networks tend to be slightly inferior to those obtained from the core validation sites, they are still of high value due to the wide geographical coverage (Chan et al. 2018; Zhang et al. 2019). In combination with the probable unevenness introduced by the simultaneous inclusion of assessment metrics from pixels with different amounts of stations, for each pixel, only one station candidate would be preserved and presumed as the areal delegate. Following the steps adopted by Zhang et al. (2021b), those representative stations have been decided by comparing the averaged correlations among the SMAP, ERA5-Land, and in-situ measurements. Specifically, for each station (STNx) in an identical grid, the Pearson Correlation (R) values between any two of the above datasets (i.e., R [SMAP, ERA5-Land], R [SMAP, STNx], and R [ERA5-Land, STNx]) have been separately computed. The station with the highest mean R value is considered to be the pixel representative.

4.2.3. GPM IMERG Precipitation

The half-hourly GPM IMERG precipitation Level-3 half-hourly product (Version 6B) has been applied to provide the water inputs to the upper soil system. Its densely temporal availability was attained via the simultaneous use of the constellation from multiple passive microwave satellite-based sensors as well as geosynchronous infrared satellites (Huffman et al. 2019). Here, the 30-minute Level 3 GPM IMERG Final-run (hereinafter referred to as GPM) precipitation volumes were aggregated for the time intervals from UTC 00:00 to 12:00 and UTC 12:00 to 00:00 prior to spatially re-gridding.

4.3. Methods

4.3.1. Estimation of the Integrated Hydrologic Loss

The hydrologic process occurring on the topsoil layer has been simplified into three components: precipitation (input), hydrologic loss (output), and SM (change in storage) (Eq 1). Since the amounts of water inputs are fixed, accurate quantification of hydrologic divergence (Q) is critical to ensure the quality of predicted SM. Akbar et al. (2018) showed that Q can be represented by a product of the hydrologic length scale (ΔZ) and the volumetric loss (L) (Eq 2). Further, L could be roughly estimated via the SMAP dry-down SM.

$$\Delta Z \cdot \frac{SM_{t+\Delta t} - SM_t}{\Delta t} = P(t \sim t + \Delta t) - \Delta Z \cdot L(t \sim t + \Delta t) \quad (1)$$

where SM_t and $SM_{t+\Delta t}$ (m^3/m^3) represent volumetric soil moisture at time points$t$ and $t + \Delta t$, respectively. Δt (day) is the time interval. ΔZ (mm) denotes the hydrologic depth within which soil moisture data have similar dynamics. P and Q are the precipitation (mm/day) and loss rates (mm/day) during the time interval.

$$Q = \Delta Z \cdot L \quad (2)$$

where L is the water loss in the volumetric unit ($m^3/m^3/day$).

The SMAP observed dry downs can be defined as the monotonically decremental SM segments. Each dry-down segment is composed of at least three consecutive SMAP observations having zero inter-measurement precipitation (Akbar et al. 2018). In other words, the decreases in these SM measurements are completely equivalent to the hydrologic losses. Hence, the SMAP dry-down segments are first identified (**Figure 4.2**a) and collected altogether (**Figure 4.2**b). Then, the daily rate of L during each dry-down period is estimated (Eq 3). However, it should be noted that the dry-down segments with fluctuation amplitudes of magnitudes smaller than 1% of the SMAP SM range were excluded. This is because such minor variations are likely caused by observation noises and/or retrieval uncertainties (Akbar et al. 2018; McColl et al. 2017).

$$L_{obs} = -\frac{SMAP_{dd}(t + \Delta t_{obs}) - SMAP_{dd}(t)}{\Delta t_{obs}} \quad (3)$$

where the subscripts obs and dd indicate the observation-based derivations, and the SMAP dry-down soil moisture measurements.

However, the above estimations are only applicable for the qualified SMAP dry-down periods, and they are unable to reflect the losses over the remaining spans at a routine frequency. A simple solution is to assume that the integrated loss over a time interval is a function of the SM value(s) at the beginning and/or the end of this period. Such an approach has been proven feasible and widely adopted (Akbar et al. 2018; Brocca et al. 2019; Koster et al. 2018). Here, we presumed that the loss rate (L_{est}) was predominately governed by the initial SM (SM_t) of a given time slot and that this relation can be sustained for 12 hours (Eq 4). The quantitative relationship between

the loss and the initial SM started by regressing the observed moisture losses to SM at the beginning of the corresponding dry-down limbs. Here, a dry-down limb is defined as a subpart of a dry-down segment, and each limb contains two successively decremental SM observations.

$$L_{est}(t{\sim}t+0.5) = f(SM_t) \quad (4)$$

where $L_{est}(t{\sim}t+0.5)$ represent the estimated volumetric water loss between t and 0.5 days after t. $f()$ displays the regressed relationship between the loss and initial soil moisture.

Specifically, two regression methods, locally weighted linear smoother (LOWESS) with a span of 65% (Akbar et al. 2018) and the quantile regression approach with a performance-driven percentile (β) have been adopted and compared. Quantile regression can be described by Eq 5 and models the conditional quantiles of a dependent variable (y) for an independent variable (x). The objective of quantile regression is to minimize the summation of two aggregated absolute error terms separately multiplied by β for underprediction and $(1-\beta)$ for overprediction (Eq 6), different from linear regression that minimizes the sum of squared errors (Magan et al. 2020; Wasko and Sharma 2014). The specific value of β is determined through an optimization procedure described in Section 3.2 (**Figure 4.2**c). In general, the loss amounts monotonically elevate with the increase of SM. Such a regressed link, however, was derived using the dry-down SM alone and it might be inappropriate to apply when the moisture content falls beyond that range. As such, this regression relationship (Eq 4) is set to be useful only when SM value is between the minimum (p_1) and maximum (p_2) SMAP dry-downs, and the above range is labeled as Segment B of the reconstructed loss function (**Figure 4.2**c).

$$y_i = a_0^{(\beta)} + a_1^{(\beta)} x_i + \epsilon_i^{(\beta)} \quad (5)$$

where x_i and y_i are data elements for the independent variable x, and dependent variable y; β is the conditional quantile ranging from 0 to 1; and $a_0^{(\beta)}$, $a_0^{(\beta)}$, and $\epsilon_i^{(\beta)}$ represent the parameters and error terms (zero expectation) under quantile β.

$$F\left(a_0^{(\beta)}, a_1^{(\beta)}\right) = \beta \sum_{y_i \geq a_0^{(\beta)} + a_1^{(\beta)} x_i} \left| y_i - a_0^{(\beta)} - a_1^{(\beta)} x_i \right| + (1-\beta) \sum_{y_i < a_0^{(\beta)} + a_1^{(\beta)} x_i} \left| y_i - a_0^{(\beta)} - a_1^{(\beta)} x_i \right| \quad (6)$$

where $F()$ is the cost function needed to be minimized.

In the cases where the initial SM values are outside the limits of Segment B, a straightforward linear extrapolation technique has been applied for estimating the loss amounts. At the dry end (Segment A), the possible minimum SM and its corresponding loss are arbitrarily assumed as 0.02 m^3/m^3 and 0 (m^3/m^3/day) following Akbar et al. (2018). The loss function can then be determined through the line connecting $[0.02, 0]$ and $[p_1, L_{est}(p_1)]$ when SM is under $[0.02, p_1]$ (Eq 7). Similarly, the rate of moisture loss can be estimated through the linear fitting between $[p_2,\ L_{est}(p_2)]$ and $[porosity\ (\Phi), L_{est}(p_2) + \alpha \cdot (\Phi - p_2)]$ over Segment C $[p_2, \Phi]$, and α is an unknown slope parameter (Eq 7). Once the initial SM values reach saturation (Segment D), the excessive amounts of water possibly owing to runoff and spillage are partitioned into the loss portion for the next 12-hour interval. Moreover, they are quantified through the water balance equation on the basis of the hypothesis that SM always returns to the porosity 12 hours after oversaturation (Eq 7). Generally, the loss functions can be easily divided into three pieces according to the slopes (**Figure 4.2**c). Compared to the intermediate SM, the losses are more sensitive to the magnitudes of SM at the dry and wet ends (Koster et al. 2018; Salvucci 2001).

Since the skills of Segments A and C highly depend on Segment B, the number of dry-downs used to construct Segment B is critical. Considering this, for any individual 9-km pixel with dry-down segments of fewer than 117 (i.e., the 25th percentile of the SMAP dry-down numbers of all the 9-km grids included here), dry-down segments from the surrounding 4×4 9-km pixels were collected for building the Segment B. If the number of aggregated dry-down segments from 16 grids was still lower than that threshold, the counterpart would be masked out.

$$L_{est}(t\sim t+0.5)=\begin{cases}(SM_t-0.02)\cdot\frac{L_{est}(p_1)}{(p_1-0.02)} & 0.02<SM_t<p_1 \quad \text{Segment A}\\ f(SM_t) & p_1\le SM_t\le p_2 \quad \text{Segment B}\\ L_{est}(p_2)+\alpha\cdot(SM_t-p_2) & p_2<SM_t\le \Phi \quad \text{Segment C}\\ \frac{P(t\sim t+0.5)}{\Delta Z}-\frac{(\Phi-SM_t)}{\Delta t} & SM_t>\Phi \quad \text{Segment D}\end{cases} \tag{7}$$

where p_1 and p_2 are the maximum and minimum SMAP dry-down soil moisture, and Φ is the porosity that is initially estimated via $1-\frac{BD}{2.65}$. BD is the soil bulk density provided by the SMAP ancillary dataset (Das and O'Neill 2020).

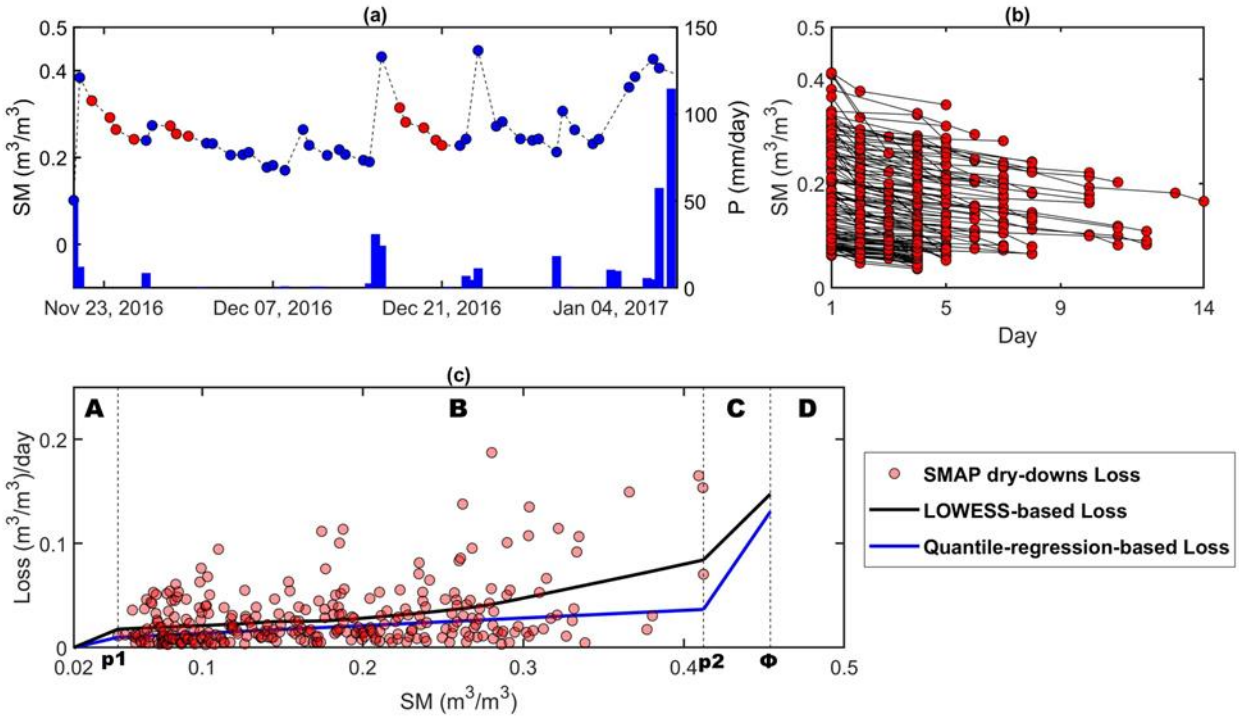


Figure 4.2. (a) An example of the SMAP soil moisture and the GPM precipitation time series for a pixel centered on [36.25ºN,120.58ºW] from November 1, 2016, to January 1, 2017. The red dots represent the dry-down soil moisture; (b) All the SMAP soil moisture dry-downs are stacked together after preprocessing procedures; (c) The quantitative relationship between the dry-down soil moisture and its paired loss. Regression analyses, including the locally weighted linear smoother (LOWESS) with a constant span of 65% (black line) and the quantile regression with a to-be-determined percentile (blue line), have been performed within the SMAP dry-downs range (Segment B). Lines in Segments A and C are linearly extended along the line of Segment B. Again, p_1 and p_2 denote the minimum and maximum SMAP dry-down soil moisture while Φ is the soil porosity.

4.3.2. Forward Simulation of Rainfall-Driven Soil Moisture

Given that L has been described as a function of SM_t, $SM_{t+0.5}$ can be predicted as long as P and SM_t are available (Eq 8). However, there are still at least two unknown parameters (ΔZ and α) that could affect the temporal evolution of the water balance cycle. A third parameter, β, will also be required if the quantile regression is used to build Segment B. The values of these pixel-wise parameters were determined by performing an optimization procedure with an objective function to minimize the root-mean-square error (RMSE) between the rainfall-driven SM simulations and the synchronous SMAP observations.

$$SM_{t+0.5} = SM_t + 0.5 \cdot \left[\frac{P(t \sim t+0.5)}{\Delta Z} - L_{est}(t \sim t+0.5)\right] \quad (8)$$

Using Eq 8, SM can be continuously simulated by inserting one initiated value for SM at the beginning of the study period. Simulated SM products in this way are expressed as PLO (P: Precipitation-driven simulation + L: 'LOWESS' regression + O: One initiated soil moisture), and PQO (P: Precipitation-driven simulation + Q: Quantile regression + O: One initiated soil moisture).

However, these precipitation-reconstructed estimates are prone to suffer from the errors accumulated for a six-year simulation interval, resulting in a large deviation between the predicted and observed SM in the late or particular stage of simulation. These errors could be sourced from the occasional mismatching between rainfall and SM datasets as well as from the inappropriate derivations of hydrologic losses. In light of this, the SMAP dry downs were adopted for the simulation process. Specifically, each dry-down SM was directly placed into the simulated time series to piecewise initiate the precipitation-driven process until the next SMAP dry-down. The intent of including those SMAP dry downs is to avoid the unreasonable magnitude differences between the simulated and observed SM values. Therefore, the SMAP dry-downs could also be labeled as correction points or as lifting-up measures regarding the rapidly dropped SM of PLO. The number of SMAP dry-down SM is around 12% of the total SMAP observations (4% of 12-hour continuous SM simulations over the six-year period). Thus, the precipitation-driven SM products including the original SMAP dry-downs are entitled as PLD (PL + D: Dry-down soil moisture) and PQD (PQ + D: Dry-down soil moisture), respectively.

Finally, PQF (PQ + F: gap-Filled) was generated to supplement the missed SM at a 12-hourly scale of the combined descending and ascending SMAP L3 product. Similar to PQD, the simulated SM was immediately displaced by the simultaneous SMAP observation. As the optimization objective is to minimize the RMSE between the simulated SM data and their temporally paired SMAP observations, the benchmark dataset and the optimization process will become ineffective when all the SMAP data are involved in the simulation. Hence, PQF was derived using the optimal parameters obtained from the PQD, and a rerun of Eq 8. PQF is the ultimate product conformed to the major purpose here that solely fills the SMAP temporal gaps. A summary of different simulated SM products is described in **Table S3** in the supporting information (hereafter Figures and Tables in the supporting information use the prefix "S").

The optimal methodology flow was determined based on the simulation results from the representative 9-km pixel (**Figure S4**). In order to cover a wide variety of land surface conditions, the CONUS matrix (285×644) composed of 9-km grids was first divided into 15×28 coarse-scale blocks. Each block encompassed 19×23 9-km pixels, and then a pixel was randomly selected from each block. Before the random procedure, the pixels with dry-down limbs of fewer than 50 and the blocks of fewer than 10 effective pixels were excluded. As a result, 260 pixels out of 420 blocks were determined and extracted as the representative points.

4.3.3. Detection of Rainfall-Induced Soil Moisture Peaks

At a large spatial scale, the vertical variations of hydrologic components are often assumed to be dominant relative to lateral hydrologic exchanges (Akbar et al. 2018). Hence, an increase in surface SM could be attributed to the precipitation input alone, and in turn, the occurrence of each SM peak corresponds to a rainfall event. Again, identifying those peaks and their causes is important for characterizing SM dynamics and runoff predictions. In order to detect the rainfall-

induced peaks, the precipitation intervals were first determined. A precipitation threshold of 0.5 mm/day was applied to exclude spurious events. Multiple precipitation events in consecutive 12-hour time intervals were merged and treated as one precipitation interval. Then, the preliminary SM peak intervals were established by adding a 12-hour lagging slot to each precipitation interval (the interval has the same width as the precipitation one but moves forward one more time step). A rainfall-induced SM peak interval can be ultimately determined if the detected peaked SM value(s) falls into the preliminary slots. It should be noted that SM peaks are not always detected after rainfall events. From the same precipitation dataset, the peak occurrences from different SM data sources were collected and compared to assess their abilities in capturing rainfall-induced SM peaks.

4.3.4. Statistical Metrics

Statistical metrics are required to reflect the quality and accuracy of different datasets. Here, the quantitative metrics, i.e., unbiased root-mean-square error (ubRMSE) and R, were adopted to describe the discrepancies in magnitudes and temporal correlations between the precipitation-based SM and the benchmark datasets. Specifically, the SMAP observations and in-situ measurements were selected as the reference to separately validate PQD and PQF.

For purposes of capturing rainfall-induced SM peaks, three categorical scores, i.e., the probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) have been computed. POD refers to the fraction of correctly identified rainfall-induced soil moisture peaks while FAR is the fraction of identified peaks that are not observed by in-situ measurements. CSI provides an integrated performance score. The formulas of all the statistical metrics are shown in **Table 4.2**.

Table 4.2. Statistical metrics used for product assessment.

Statistical Metrics	Equation	Minimum	Maximum
ubRMSE	$\sqrt{E[(\theta_{sim}-\theta_{ref})^2]-E[(\theta_{sim}-\theta_{ref})]^2}$	0 (optimal)	
R	$\frac{E[(\theta_{sim}-E[\theta_{sim}])(\theta_{ref}-E[\theta_{ref}])]}{\sigma_{sim}\sigma_{ref}}$	-1 (optimal)	1 (optimal)
POD	$\frac{H}{H+M}$	0	1 (optimal)
FAR	$\frac{F}{H+F}$	0 (optimal)	1
CSI	$\frac{H}{H+F+M}$	0	1 (optimal)

where $E[]$ represents the expectation value; θ_{sim} and θ_{ref} denote the soil moisture from simulation product and reference dataset; σ_{sim} and σ_{ref} refer to the standard deviations of simulated and referenced soil moisture data; H is the number of rainfall-induced soil moisture peaks simultaneously detected by a simulated dataset and in-situ measurements; F is the number of rainfall-induced soil moisture peaks detected by a simulated dataset but not observed by in-situ measurements; and M is the number of rainfall-induced soil moisture peaks observed by in-situ measurements but not detected by the simulated product.

4.4. Results and Discussion

4.4.1. Determination of Optimal Simulation Flow

The performance of the precipitation-driven SM dataset primarily depends on the accuracy degree of loss estimations. A preliminary examination of the PLO generally exhibited an over-rapidly drying down rate compared to the SMAP observations. Such speedy declines could be attributed to the loss overestimations. As shown in **Figure 4.2**c, hydrologic losses of similar-magnitude SM values scatter across a wide range. Hence, the use of a constant smoother (i.e., 65%) seems improper, especially for SM values near p_2 with a limited number of corresponding losses. As such, we introduced β for quantile regression. Despite being more or less arbitrary, the performance-driven parameter β is expected to return more suitable loss estimations for SM predictions. Additionally, the lifting-up measure (as mentioned in Section 4.3.2) was also adopted to prevent inordinate deviations of precipitation-reconstructed SM estimates away from

the SMAP retrievals. Therefore, an investigation of the efficiency and effectiveness of different methodology flows (i.e., PLO, PLD, and PQD) was conducted.

The common parameters (ΔZ and α) (**Figure 4.3**a and b) and performance metrics (ubRMSE and R) (**Figure 4.3**c and d) of precipitation-driven simulations from three different methodology flows (i.e., PLO, PLD, and PQD) are illustrated below. Simulations are only performed over the representative pixels (**Figure S4**), and the metrics were obtained by comparing them against the SMAP observations. The PQD shows a pronounced superiority over the PLO and PLD (**Figure 4.3**c and d). Such an improvement demonstrated that the introduction of variable β availed more accurate loss estimations. Additionally, the incorporation of lifting-up measures seemed useful to insignificantly upgrade the overall performance of the simulated SM datasets. Therefore, PQD was decided as the operated methodology flow for the precipitation-reconstructed SM simulations and parameters' derivations for the rest portion. The inclusion of quantile regression and correction points was used to produce better 12-hourly SM with a near-SMAP accuracy.

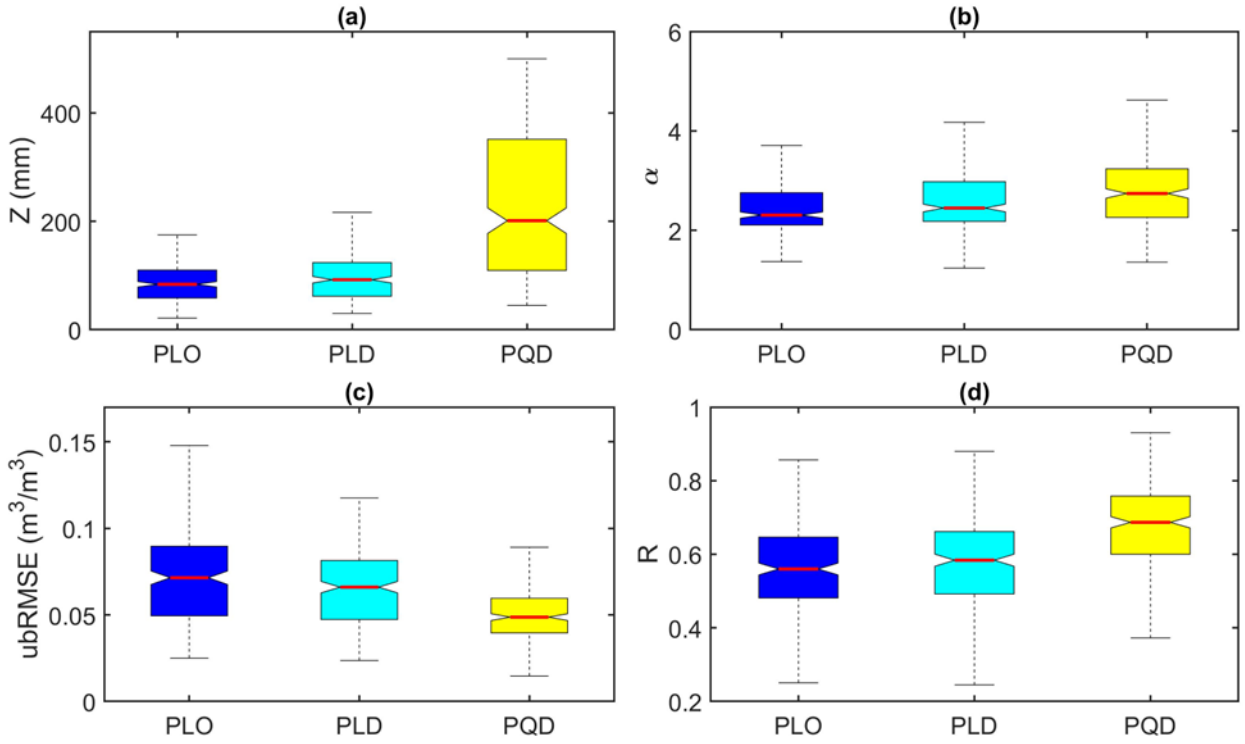


Figure 4.3. Boxplots of the optimal parameters (a) ΔZ (mm) and (b) α, and performance metrics of (c) ubRMSE (m^3/m^3) and (d) R calculated by separately comparing the PLO, PLD, and PQD against the SMAP retrievals.

4.4.2. Comparison between SMAP and PQD Soil Moisture

The spatial distributions of R between the precipitation-reconstructed SM derivations (PQD) and the SMAP retrievals are illustrated in **Figure 4.4**a, c, and d. The median R of 0.69 indicates that temporal variations described by the PQD agree well with the SMAP observations across the entire study period. However, there is a notable degradation in the performance of SM simulations over cold seasons (November to April) relative to warm seasons (May to October) (**Figure 4.5**b). In the high-latitude areas, R values quickly drop from 0.7 to 0.5 (**Figure 4.4**c and d). Such a large seasonal difference could be attributed to the low availability of the SMAP benchmarks during winter as a result of long frozen periods and/or frequent snowfall events. The sparse SMAP sampling during cold seasons likely led to the optimal parameters (i.e., ΔZ, α, and β) being more beneficial for simulating SM in warm seasons. It was corroborated by the close

magnitudes of the median ΔZ (139 mm) and the median ΔZ (135 mm) obtained by Akbar et al. (2018) that only used the SM from May to September. On the contrary, an improvement of R values in the western CONUS during cold seasons can be observed (**Figure 4.4**d), probably due to the seasonal rainfall patterns over these domains (i.e., rare rainfall during summer). In view of R, the simulation scheme proposed here is more feasible for periods with sufficient SM observations and with adequate rainy days.

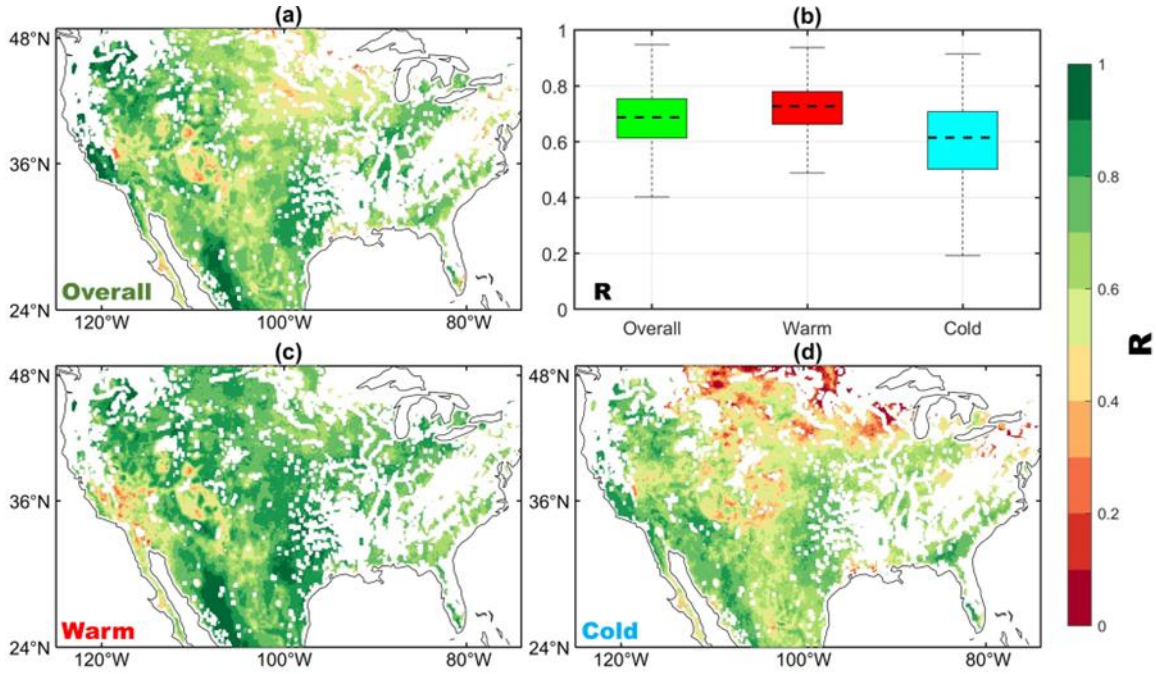


Figure 4.4. Spatial distribution of R between the SIMU-PQD simulations and the SMAP observations over the CONUS, where (a), (c) and (d) represent the metrics obtained using the soil moisture data of the entire study period, warm seasons (May to October), and cold season (November to April).

The ubRMSE describes the discrepancies in absolute magnitude between the PQD and the SMAP retrievals. Overall, there is an east-west gradient in the distribution of the ubRMSE map (**Figure 4.5**a, c, and d). The relatively higher deviations of the PQD in eastern CONUS could be partly attributed to the generally larger SM values due to more precipitation volumes. Additionally, the SMAP SM data in the eastern sides were mostly retrieved under the VWC of more than 5 kg/m^2 which is a threshold commonly used to screen the low-quality SM estimations

from densely vegetated areas out (O'Neill et al. 2021a). Therefore, the loss estimations from the dry downs of inferior SMAP SM data may not quantitatively function as well as the regions with VWC below 5 kg/m^2. Overall, an integrated consideration of the ubRMSE (0.05 m^3/m^3) and R (0.69) could indicate the comparable performance of the PQD relative to the SMAP retrievals as well as the validity of the derived parameters. Those parameters are necessary for cycling the water balance budgeting (Eq 7 and Eq 8).

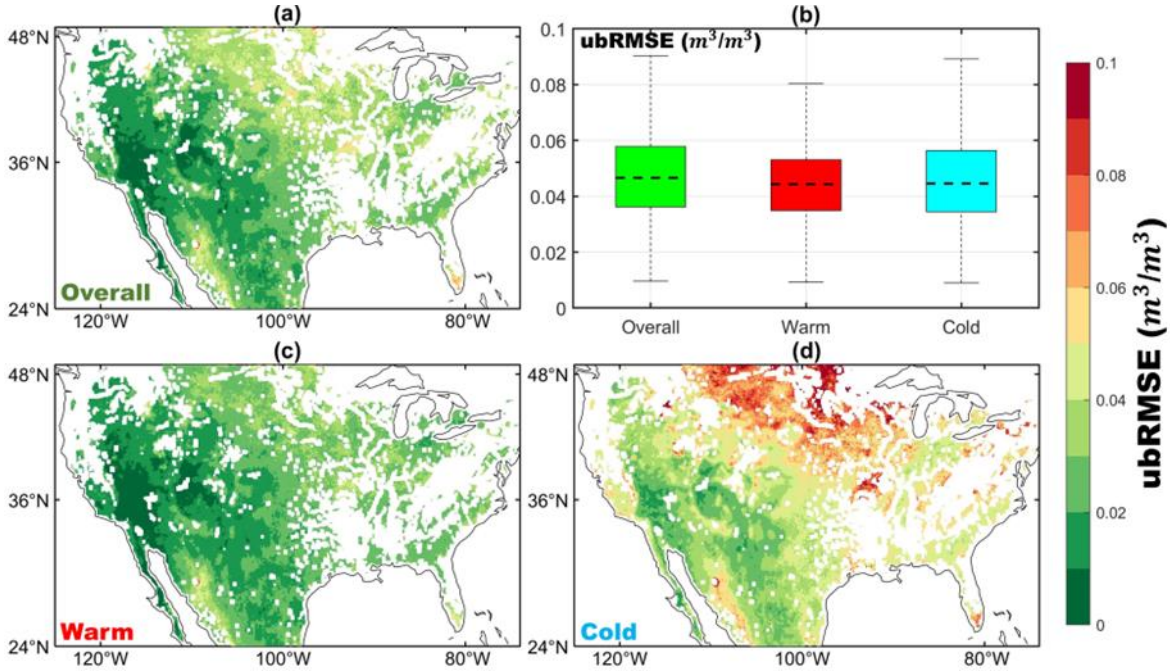


Figure 4.5. Spatial distribution of ubRMSE (m^3/m^3) between the SIMU-PQD simulations and the SMAP observations over the CONUS, where (a), (c) and (d) represent the metrics obtained using the soil moisture data of the entire study period, warm seasons (May to October), and cold season (November to April).

4.4.3. Validation of Gap-filled Soil Moisture Products

The effectiveness of the optimized parameters (i.e., ΔZ, α, and β) was proven by the near-SMAP accuracy of the PQD. To proceed, the gaps of the SMAP combined descending and ascending product were supplemented by re-running the simulations using the obtained parameters, precipitation inputs, and all the available SMAP observations (Eq 8). Again, this temporal gap-filled SM dataset is abbreviated as PQF.

Time series examples of the PQD and PQF from three contrasting locations over the CONUS are displayed in **Figure 4.6**. Estimations from those two sources (i.e., PQD and PQF) are overall consistent and exhibit incremental responses to precipitation events. However, the PQF presents more temporal variability during dry periods. During the long-term drying periods of zero or extremely low precipitation, the PQD SM shows monotonically decremental trends given the use of Eq 8 (**Figure 4.7**). In contrast, PQF data present frequent fluctuations forcedly modulated by the inserted SMAP observations (**Figure 4.7**). Such a difference disclosed the limitations of this parsimonious scheme without adequately considering the physical variations, such as dew in the morning, into account.

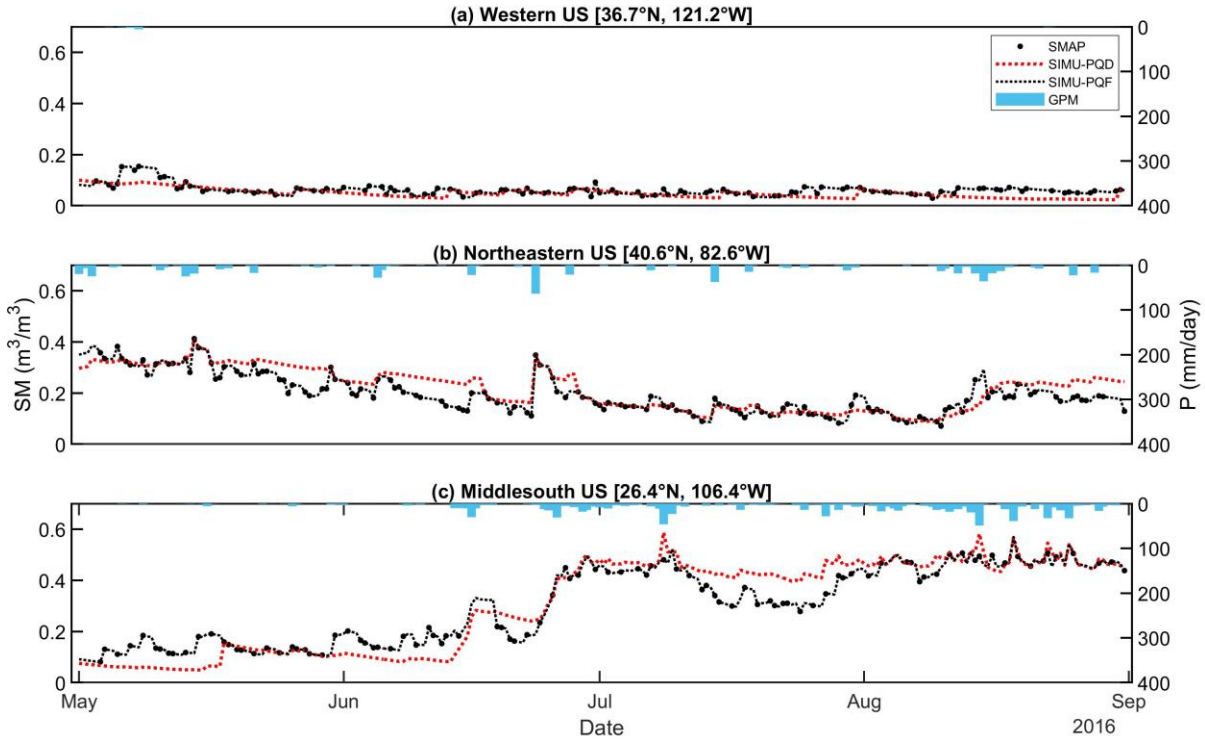


Figure 4.6. Example time series of the PQD and PQF soil moisture simulations.

The performances of the PLO, PQD, and PQF were further evaluated by comparing them against in-situ measured SM over 526 CONUS pixels (**Figure S6**). Additionally, assessment metrics of another two SMAP-based continuous SM datasets yielded using the three-dimensional discrete

cosine transform-penalized least square regression (DCT) (Wang et al. 2012), and the simple linear interpolation (LIP), were also calculated. The ability of the DCT algorithm to completely overlap its reconstructed and the original data was highlighted by Wang et al. (2012) . Here, rather than using the entire three-dimensional SMAP array within the study domain, only observations from the neighboring grids surrounding the targeted pixel (as a center of a 9×9 matrix) have been included for conducting a local DCT gap filling.

The PQF, DCT, and LIP display comparable ubRMSE and R medians (**Figure 4.7**a and b) as they are largely framed by identical SMAP observations. As discussed above, the parsimonious scheme smoothed SM variations of the PQD relative to PQF, likely leading to a minor degraded performance of PQD. Given this, a sufficient number of satellite-based SM observations are necessary to mitigate the errors accumulated during simulations. Consistent with the results of Section 4.4.1, the distinct inferiority of the PLO is more attributed to the improper loss estimations. Compared to the metric scores of the SMAP observations, the gap-filled portions show comparable performances (**Table S4**). Although the ubRMSE values of SMAP observations slightly outperform the filled-in SM (when more digits are examined), the R values of filled-in SM of the DCT and LIP products are unexpectedly higher than those of the SMAP data (**Table S4**). Such small discrepancies between the performances of the SMAP retrievals and those exclusive filled data are adequate to demonstrate the validity of the proposed scheme and the derived ancillary parameters. It should be noted that the average number of filled-in SM data was around two times relative to that of the SMAP available samples.

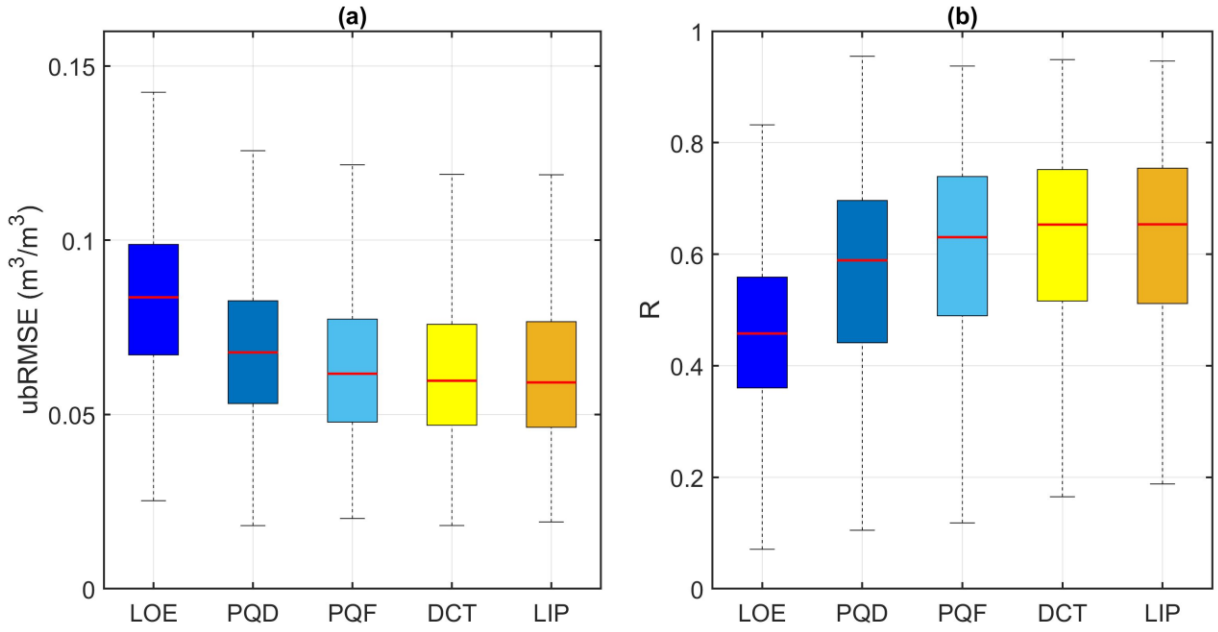


Figure 4.7. Boxplots of the quantitative performance metrics of (a) ubRMSE (m^3/m^3) and (b) R estimated by comparing five SMAP-based gap-filled products through five different ways against in-situ measurements.

In terms of land response to incoming precipitation, the consistency of rainfall-induced SM peaks between each gap-filled dataset and in-situ measurements was separately analyzed. Those peaks caused by heavy rainfall were majorly investigated given their importance. The POD refers to the ratios of peaks successfully detected by the simulation products to all the peaks caught by in-situ measurements, and the optimum POD is 1. **Figure 4.8**a shows that the POD medians of the PLO and PQD are near 1, suggesting that those two products could capture almost all heavy rainfall-induced peaks. The FAR (optimal FAR is 0) represents the fractions of peaks identified by the simulation dataset but not observed by in-situ measurements, and it is described in **Figure 4.8**b. In contrast, it reveals that the PLO, PQD, and PQF hold excessive peaks. Such frequent peak occurrences can be attributed to the forward simulation procedure. Generally, a precipitation event is bound to produce one immediate SM peak at the end of the precipitation event (Eq 8). In nature, however, the peak responses of SM to rainfall are not limited by a strict 12-hour interval and they might occur within the 12-hour interval and shortly

return to normal. Furthermore, the CSI (optimal CSI is 1) reflects an integrated ability of the simulation product to capture rainfall-induced SM peaks. The PLO, PQD, and PQF have better CSI values than the DCT and LIP (**Figure 4.8**c).

Although the PQF, DCT, and LIP exhibited similar values of ubRMSE and R, the gradient CSI values indicate the latter two gap-filled products missed some critical SM increases. The POD and CSI of the PQF improved by around 35% and 32% relative to LIP whereas the FAR declined by 51% given the extremely small magnitude. The DCT could partially compensate for the temporal missing by incorporating spatially surrounding observations but cannot compete with the PQF in capturing SM peaks induced by heavy rainfall. When considering the categoric metrics of capturing SM peaks induced by all the rainfall, the advantages of the PQF over the DCT and LIP can be still observed (**Figure S7**) in a similar manner shown in **Figure 4.8**a and c. Therefore, an integrated consideration of both accuracy and peak-capturing capacity could conclude that the PQF provides continuous SM observations of the best quality.

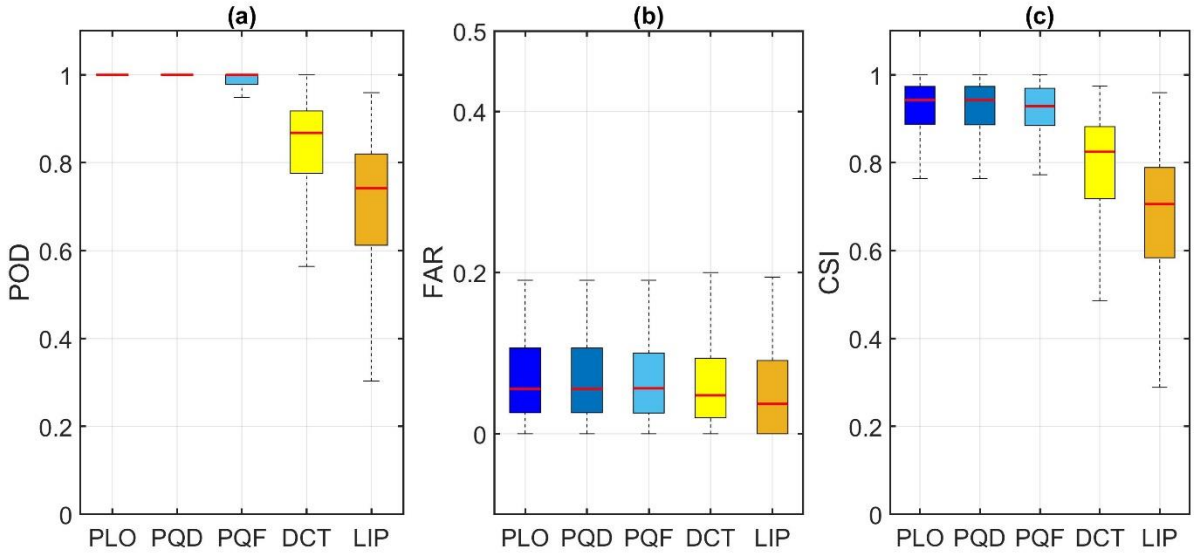


Figure 4.8. Boxplots of the categorical performance metrics of (a) POD, (b) FAR and (c) CSI for five different SMAP-based gap-filled soil moisture datasets in capturing soil moisture peaks caused by the

heavy rainfall events (exceeding 95% locally non-zero 12-hour precipitation volumes). Those soil moisture peaks observed by in-situ soil moisture measurements are used as the benchmarks.

4.4.4. Caveats and Future Works

In addition to supplying reasonable estimations for the missing data in SMAP, the proposed scheme could also provide quantification for other hydrological components of the simplified water balance equation. Yet, the conventional gap-filling methods are usually unable to afford such information. Those ancillary estimations could be critical for many hydrologic-related studies. The hydrologic length scale (ΔZ) could be an important factor in learning SM dynamics. ΔZ is defined as the depth within which SM variations are temporally analogous (Akbar et al. 2018). In other words, surface SM has a strong correlation with the subsurface soil moisture above the depth of ΔZ. **Figure S5** depicts the spatial distribution of ΔZ across the CONUS, showing consistent east-west gradients and a similar median ΔZ relative to Akbar et al. (2018). Since ours ΔZ was estimated by introducing the optimization parameter β, it can be seen as more in line with the natural process than the value of Akbar estimated by applying an arbitrary LOWESS parameter. The stored precipitation fraction (McColl et al. 2017), an index of SM memory, was computed using the SMAP observations and a constant ΔZ of 50 mm. The retained rainfall proportions will increase, and its spatial patterns will also alter if the map of the diverse ΔZ with values generally deeper than 50 mm is incorporated. In addition to ΔZ, the availability of loss estimates could also be useful for other research, such as the Soil Moisture to Rain algorithm (SM2RAIN) (Brocca et al. 2014; Brocca et al. 2019). The completely observation-driven processes and the simplicity are the additional advantages of this approach.

Despite those, the limitations of the precipitation-driven SM estimation cannot be fully ignored. The varying patterns of the PQD and PQF are more abrupt than the observations, which could be

partly ascribed to omitting information from the diurnal cycle of SM, such as the nighttime re-moistening (O'Neill et al. 2021a). Additionally, SM simulations tend to have a wider magnitude range than observations, and those extremely large and occasionally negative SM estimations are likely due to the inappropriate parameters and loss underestimations under extreme precipitation scenarios. Furthermore, a greater uncertainty within the simulations that fill the gap interval of a longer duration is expected. Hence, a flag file has also been generated to mark those filled-in SM for long-term gaps and with impractical values. Yet, it should be noted that we only focus on filling the temporal gaps left in the existing SMAP dataset. Extension from the available data-covered regions to other spatial gaps is out of scope. Since the spatial gaps of satellite-based SM products are often caused by filtering procedures and land-water masks, arbitrary supplements via surrounding information may be inappropriate.

In this study, an important assumption that soil water content hardly varies within a 3-hour time interval has been adopted. The selection of a 3-hour duration is based on a preliminary examination of the time differences between the actual SMAP overpass moments, and the UTC 00:00 and 12:00. This assumption made it possible to express SMAP data in the fixed UTC timeframe before extracting dry downs. However, such a 3-hour deviation could be a concern to cause the mismatching between rainfall and SM responses.

The combined consideration of the novel features and deficiencies of this study spurs and outspreads topics for future work. A fine-spatial-scale precipitation dataset is required by many hydrologic analyses at catchment scales, but it is often restricted by the availability (Lobligeois et al. 2014; Sadeghi et al. 2021; Sharifi et al. 2019). Given the development of SM downscaling frames and the SM2RAIN algorithm in recent years, the generation of high spatial-resolution rainfall products using 1 km SM data become feasible (Fang et al. 2018; Fang et al. 2020; Fang

et al. 2022; Fang et al. 2021c; He et al. 2022; Peng et al. 2021). However, the 1-km SM2RAIN-based precipitation may also suffer from underestimation and heavily depends on the availability of high-resolution SM products. The PLO and PQO estimations can be roughly seen as the quantified water indexes converted from the precipitation (rainfall → soil water content). The SM spatial downscaling scheme, for example, the thermal-inertia-theory-based algorithm proposed by Fang et al. (2022), could be conducted to re-distribute those partitioned rainfall water volumes on land (at 9-km) using the local (1km) vegetation and land surface temperature. Through SM2RAIN, the above obtained 1-km rainfall water indexes on land could be transformed back to 1-km rainfall (soil water content → rainfall). Such a full cycle flow (rainfall → soil water content → rainfall) is expected to alleviate the underestimation concern by running the half-cycle-based SM2RAIN (i.e., soil moisture → rainfall) for spatial disaggregation of the coarse-scale rainfall datasets. Meanwhile, the non-gapped 12-hour 9-km SM simulations increase the available samples in the downscaled datasets.

Additionally, the other promising direction would be a further improvement of the re-constructed SM simulations by incorporating machine learning and data assimilation techniques, as well as adding considerations related to the diurnal cycle of SM. These measures are anticipated to return more accurate SM values principally because of better loss estimations. The combined use of the improved loss model and the long-term precipitation dataset, such as Multi-Source Weighted-Ensemble Precipitation (MSWEP) (Beck et al. 2019b), makes it possible to yield a consistent SM dataset spanning 40 years. Moreover, the regionally available high-resolution precipitation observations (e.g., 4 km) (Beck et al. 2019a; Lin and Mitchell 2005; Sadeghi et al. 2021) could also be used to derive soil water content at the same resolution if assuming a homogeneous loss at the large spatial scale (e.g., 9-km). In practice, those 30-minute

precipitation datasets have the potential to infer the moisture content at the sub-hourly step. It is also feasible to produce the global-scale PQO, PQD, and PQF if desiring a local-time-based dataset.

4.5. Summary and Conclusions

In this study, a 12-hourly continuous SM product (PQF) from 2015 to 2021 was generated. This newly yielded dataset is composed of SMAP retrievals and filled-in SM estimates derived through water balance budgeting. Specifically, the existing gaps of the combined descending and ascending SMAP dataset were supplemented by SM predictions using the precipitation and hydrologic loss in their antecedent 12-hour slots. The whole simulation procedure was purely observation-based using the satellite-based precipitation product (i.e., GPM) and the loss approximations established on the SMAP dry-downs. Compared to the LOWESS with a fixed smoother, the quantile regression with a to-be-determined parameter was introduced to bridge the relationship between the losses and initial SM from the SMAP observations given the scattered loss ranges at similar wetness. The magnitudes of all the parameters for driving the water balance cycle were then determined by minimizing the RMSE between the precipitation-derived SM simulations and the synchronous SMAP observations. In addition to the PQF, the precipitation-reconstructed SM product (PQD) that combined the SMAP dry-down SM and precipitation-based filled-in SM, was also developed.

The superiority of the parametric regression in building Segment B has been highlighted by a better performance of the PQD relative to the PLD over the CONUS representative pixels. Additionally, the inclusion of the SMAP dry-downs in gap-filling would be somewhat helpful to mitigate the overly rapid drying rates observed in PLO. Using the certified methodology flow,

the PQD was produced over the CONUS. Validation results showed that the PQD overall resembled the SMAP retrievals despite an overall performance degradation during cold seasons. Moreover, the median R of 0.69 and the median ubRMSE of 0.05 m^3/m^3 proved the validity of the derived parameters. It should be noted that all the above assessments were conducted by comparing the simulations with the SMAP product. Subsequently, the PQF was generated by rerunning the water balance equation with the obtained parameters. In-situ SM measurements were selected as the benchmarks to evaluate the PQF and the SMAP-based seamless SM datasets through the other filling approaches (i.e., DCT and LIP). The assessment results revealed that the PQF not only displayed good accuracy (ubRMSE: 0.06 m^3/m^3; R: 0.63) comparable with that of the DCT and LIP but also exhibited a greatly stronger ability to reflect positive responses to heavy rainfall events. The similar performance of exclusively filled-in SM to the original SMAP retrievals convinced the effectiveness of the proposed scheme. The superior integrated performance of the PQF provides an alternative option for researchers interested in studying sub-daily SM variations. Besides, the availability of the hydrologic parameters and the quantified loss estimations could directly benefit other related studies and serve as the foundations for further investigations.

Chapter 5. Identifying Relative Strengths of SMAP, SMOS-IC, and ASCAT to Capture Temporal Variability

R. Zhang, S. Kim, A. Sharma, and V. Lakshmi (2021). Identifying Relative Strengths of SMAP, SMOS-IC, and ASCAT to Capture Temporal Variability. *Remote Sensing of Environment*. 2021, 252, 112126. https://doi.org/10.1016/j.rse.2020.112126

Abstract

This study evaluates the relative strengths of three remotely sensed soil moisture (SM) products to capture temporal variability at a global scale, the products being the Soil Moisture Active Passive (SMAP), Soil Moisture Ocean Salinity INRA-CESBIO (SMOS-IC) and Advanced Scatterometer (ASCAT). For this, the conventional reference-based Pearson correlation (R) and a statistical technique called Triple Collocation (TC)-based R are calculated. In addition, two alternatives for linear combination of the three data sources for maximizing R against the truth are evaluated, the first using a reference product (i.e., assumed truth) and the second based on TC where three data sources are combined without the need for an underlying reference or assumed truth. The estimated optimal combination weights represent quantitative contributions of the three products in forming the new combined product having the maximized R. Two reanalysis products: the European Centre for Medium-Range Forecast (ECMWF) Interim product (ERA-Interim) and the Modern-Era Retrospective Analysis for Research and Application Land version 2 reanalysis product (MERRA2), are used as the references as well as data alternatives to calculate the conventional reference-based R and the TC-based R combinations. Both types of R, and their derived optimal weights are then compared globally and analyzed under various climate, land cover, and vegetation conditions. Despite the differences between the conventional

R and the TC-based R, both metrics displayed consistent spatial distributions and can reflect the temporal variations of each studied dataset without considerable impact from adopted references. All products had difficulty in retrieving SM over arid and polar regions while exhibiting good performance in areas such as South America and Australia. While ASCAT presented higher R values over tropical, savannas, and the vegetation water content interval of 2–5kg/m^2, SMAP and SMOS-IC displayed overall comparable and continually high temporal performances across almost all conditions. In the case of the derived optimal weights, a global complementarity of the areas was observed where each satellite-based observation product showed its respective advantage in capturing SM variations in different geographic areas.

Keywords: Soil Moisture, SMAP, SMOS-IC, ASCAT, correlation coefficients, optimal weights

5.1. Introduction

Soil moisture (SM) is an important factor modulating both terrestrial-atmosphere interactions and hydrological circulation (Kerr et al. 2016; Srivastava et al. 2013; Van den Hurk et al. 2002). Given that SM provides useful information for many real-world applications, including flood and weather forecasting, there is an added significance of accurately estimating SM in both space and time (Entekhabi et al. 1999; Entekhabi et al. 1996).

Remote sensing has been considered as a promising technique for providing regional SM information at a global scale (De Jeu and Owe 2003; Kim et al. 2019; Petropoulos et al. 2015). Microwave remote sensors can measure the soil dielectric constant which is strongly correlated to the water content within the topsoil layer (Kerr et al. 2016; Pellarin et al. 2003). In comparison to optical/infrared radiation and analogous remote sensing alternatives (Kerr 2007; Petropoulos et al. 2015), microwave remote sensing can operate under all-weather conditions and during nighttime (Srivastava et al. 2015). Additionally, microwave systems are less sensitive to surface roughness and can penetrate vegetation canopy (Ulaby et al. 1986b; Wigneron et al. 1998). Due to their greater temporal resolution and lower influence by roughness disturbances than active sensors, passive microwave radiometers have been widely used to estimate SM since late the 1970's and form the focus of our investigation (Al-Yaari et al. 2014; Bertoldi et al. 2014; Kim et al. 2019).

Understanding error characteristics and limitations of SM retrievals is an important step to improve retrieval algorithms. Given that, validation studies that evaluate the accuracy of the remotely sensed SM products have been widespread and useful (Al-Yaari et al. 2019; Ma et al. 2019; Zhang et al. 2019). Errors in satellite SM data can be presented by statistical metrics (e.g.,

bias and root mean squared error). These errors generally result from uncertainties in the retrieval process and/or limitations of the satellite instruments, with uncertainties inevitably inducing biases and/or unreasonable dynamic ranges of the retrievals. Moreover, it is often hard to exactly evaluate satellite products due to the absence of references at spatial scale corresponding to those of the satellite data (Clewley et al. 2017; Crow et al. 2012). Consequently, the temporal Pearson correlation coefficient (R) has been considered as a comparatively reliable metric in the performance comparisons of spatially mismatched SM products (e.g., in-situ and remotely sensed SM retrievals). This is because it is less sensitive to the bias or amplitude of variations (Entekhabi et al. 2010b), and the temporal dynamics of SM is relatively stable over a large area compared to absolute values that are affected by the high variability of local SM (Brocca et al. 2009a). In addition, in many applications, the temporal variability in geophysical variables has been considered as an important statistic (Guillod et al. 2015; Gupta et al. 2006; Koster and Suarez 2001) and scaling has been subsequently applied to adjust the range of variables before use in applications (Brocca et al. 2009b; Reichle et al. 2004; Yilmaz and Crow 2013).

In view of this, this study uses Pearson correlation (R) to assess the temporal performances of SM retrievals derived from various satellite instruments. Two types of R including the conventional R calculated based on the arbitrary references and the R values derived using the triple collocation (TC) approach were used and compared. For the conventional R, three different reference products have been selected, which are the in-situ measurements, and two reanalysis products: European Centre for Medium-Range Weather Forecast (ECMWF) ERA-Interim SM product and Modern-Era Retrospective Analysis for Research and Application Land version 2 reanalysis product (MERRA2).

Together with estimating R, we also evaluate the relevance of one data source compared to another using an optimal weight (w), to identify relative contributions of individual SM products to capture the temporal variability. Here, the optimal weight is defined as the weight (within its constrained range) of the parent product being linearly merged to maximize R between the combined product and an objective reference dataset (Hagan et al. 2020; Kim et al. 2018; Kim et al. 2015). In the linear combination scheme, the optimal weight can be interpreted as the quantitative contribution of each parent product.

Different from the R that indicates the absolute performance of the product with regard to the reference datasets, the optimal weight can reflect the dependent performances among the parent products relative to each other. Additionally, the optimal weight is expected to contain the unique strength of each parent product. This has been demonstrated by the improved performances of the combined product compared to individual parent product (Kim et al. 2015). In other words, when two parent products contain much overlapped information, the contribution from one product tends to largely reduce. In light of this, it is critical to investigate the relationship between the R and the optimal weights for further enhancing the quality of the combined products.

Using the two metrics, R and optimal weight, this study compared three high-quality SM products at a global scale: Soil Moisture Active Passive Level 3 (SMAP); Soil Moisture Ocean Salinity INRA-CESBIO (SMOS-IC); Advanced Scatterometer (ASCAT). Here, the study period is from April 2015 to March 2018, nearly double of recent related studies, providing more confidence in the conclusions made (Al-Yaari et al. 2019; Ma et al. 2019).

5.2. Methodology

5.2.1. Triple Collocation-Based Data-Truth Correlation

Together with using the conventional R calculated based on the specified reference products, this study applied Triple Collocation (TC) to estimate the data-truth R. TC is an inter-comparison between three independent products which allows approximate estimation of the error variance of each product along with the correlation between each product and the unknown truth. The approximation in TC becomes possible with four assumptions: 1) linearity between the truth and the errors; 2) stationarity for both the truth and the error; 3) zero error cross-correlation; 4) error-truth orthogonality. Detailed equations and their derivations are available in Stoffelen (1998); McColl et al. (2014); Gruber et al. (2016). In this study, the data-truth correlations are estimated using Equation **Error! Reference source not found.**.

$$R_{X,T} = \sqrt{\frac{Cov(XY)Cov(XZ)}{Cov(XX)Cov(YZ)}} \quad (1)$$

where $R_{X,T}$ represents the correlation between the product X and the TC-derived data truth (T). $X, Y,$ and Z represent the different datasets within the triplet, and Cov means the covariance of any two studied products.

Following a typical combination of datasets used in TC approach (i.e. active, passive and modeling products) (Chen et al. 2018; Kim et al. 2018), we adopted two data triplets to estimate the TC-derived data-truth R of the satellite-based products, which are 1) SMAP, ASCAT and a reference product (i.e. in-situ observation, ERA-Interim and MERRA2 in this study), and 2) SMOS-IC, ASCAT and a reference product. For consistency in estimating the TC-based R, those

for SMAP and ASCAT were extracted from the first triplet and that for SMOS-IC was estimated from the second triplet.

5.2.2. Estimating Optimal Weights for Maximizing R

In this study, the relative contribution in capturing temporal SM variations was represented and compared using optimal weights in a linear combination scheme. The linear combination can effectively incorporate the advantages of each parent product and potentially generate a new product with better performance than any individual dataset being merged (Kim et al. 2015, 2016).

The SM dataset (n data length × p datasets) are linearly combined by assigning a weight vector $\boldsymbol{w}$ (p×1) to the dataset to generate a combined product $\mathbf{SM}_c$ (n×1) using Equation (2).

$$\mathbf{SM}_{\mathrm{c}} = \mathbf{SM} \times \boldsymbol{w} \tag{2}$$

The optimal weight ($\boldsymbol{w}^*$) can be obtained by solving a constrained optimization problem of maximizing the correlation between $\mathbf{SM}_{\mathrm{c}}$ and the reference, i.e., $\mathrm{R_{cr}}$, as

$$\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}} \mathrm{R_{cr}}(\boldsymbol{w}) \quad \text{subject to } 0 \le \mathrm{w}_i \le 1, \sum_{i=1}^{\mathrm{p}} w_i = 1 \tag{3}$$

The weight is able to effectively reflect the sensitivity of the combined product to the corresponding parent product, or the contribution an individual dataset makes in forming the combination.

In case of merging two products ($\mathbf{SM}_1$ and $\mathbf{SM}_2$) for which standard deviations are the same (i.e. scaled to each other or a reference), the optimal weights (w_1 and w_2) are affected by R between the two products (R_{12}) and the data-reference R ($\mathrm{R}_{1\mathrm{R}}$ and $\mathrm{R}_{2\mathrm{R}}$) (Kim et al. 2015).

$$w_1 = \frac{\mathrm{R}_{1\mathrm{R}} - \mathrm{R}_{12}\mathrm{R}_{2\mathrm{R}}}{(\mathrm{R}_{2\mathrm{R}} - \mathrm{R}_{12}\mathrm{R}_{1\mathrm{R}}) + (\mathrm{R}_{1\mathrm{R}} - \mathrm{R}_{12}\mathrm{R}_{2\mathrm{R}})} \qquad (4)$$

$$w_2 = 1 - w_1$$

In this study, when considering three products being merged for maximizing R, Equation (5) was extended as per Kim et al. (2015).

$$w_1 = \frac{AS_2S_3}{CS_1S_2 + BS_1S_3 + AS_2S_3}$$

$$w_2 = \frac{BS_1S_3}{CS_1S_2 + BS_1S_3 + AS_2S_3} \qquad (5)$$

$$w_3 = \frac{CS_1S_2}{CS_1S_2 + BS_1S_3 + AS_2S_3}$$

where,

$$\mathrm{A} = \mathrm{R}_{12}\mathrm{R}_{2\mathrm{R}} - \mathrm{R}_{1\mathrm{R}} + \mathrm{R}_{13}\mathrm{R}_{3\mathrm{R}} + \mathrm{R}_{23}^2\mathrm{R}_{1\mathrm{R}} - \mathrm{R}_{12}\mathrm{R}_{23}\mathrm{R}_{3\mathrm{R}} - \mathrm{R}_{13}\mathrm{R}_{23}\mathrm{R}_{2\mathrm{R}}$$

$$\mathrm{B} = \mathrm{R}_{12}\mathrm{R}_{1\mathrm{R}} - \mathrm{R}_{2\mathrm{R}} + \mathrm{R}_{23}\mathrm{R}_{3\mathrm{R}} + \mathrm{R}_{13}^2\mathrm{R}_{2\mathrm{R}} - \mathrm{R}_{12}\mathrm{R}_{13}\mathrm{R}_{3\mathrm{R}} - \mathrm{R}_{13}\mathrm{R}_{23}\mathrm{R}_{1\mathrm{R}}$$

$$\mathrm{C} = \mathrm{R}_{13}\mathrm{R}_{1\mathrm{R}} - \mathrm{R}_{3\mathrm{R}} + \mathrm{R}_{23}\mathrm{R}_{2\mathrm{R}} + \mathrm{R}_{12}^2\mathrm{R}_{3\mathrm{R}} - \mathrm{R}_{12}\mathrm{R}_{13}\mathrm{R}_{2\mathrm{R}} - \mathrm{R}_{12}\mathrm{R}_{23}\mathrm{R}_{1\mathrm{R}}$$

where the subscripts, 1, 2, 3 and R, represent the parent products being merged and the reference, and S_i represents the standard deviation of a given candidate.

In order to directly evaluate the product's contribution to capture temporal variability by removing effects from differences in standard deviations, all parent products were rescaled

against a reference using Equation (6) before calculating the optimal weights. Note that the R values among the newly yielded datasets and references would not be affected by the rescaling process.

$$\mathbf{SM}_{\text{rescaled}} = (\mathbf{SM}_{\text{raw}} - \overline{\mathbf{SM}_{\text{raw}}}) \times \frac{\text{std}(\mathbf{SM}_{\text{ref}})}{\text{std}(\mathbf{SM}_{\text{raw}})} + \overline{\mathbf{SM}_{\text{ref}}} \tag{6}$$

where $\mathbf{SM}_{\text{rescaled}}$, $\mathbf{SM}_{\text{raw}}$ and $\mathbf{SM}_{\text{ref}}$ represent the SM values of rescaled product, raw product, and reference, respectively. $\overline{\mathbf{SM}}$ and $\text{std}(\mathbf{SM})$ are the average and standard deviation values of SM data.

As mentioned, the TC-based data-truth R has the ability to represent the correlation between the product and the SM truth. When the SM truth is identified as the objective dataset for the linear combination, the relative contribution of each parent product can be reflected by the optimal weights derived from the TC-based data-truth R. To distinguish the optimal weights obtained by adopting the arbitrary reference products and the SM truth as the targets of the combination products, the optimal weights computed from the conventional R and the TC-based R are referred to be as w_{con} and w_{TC}, respectively. While the TC assumptions are satisfied and the parent products have been processed by the rescaling process, Equation (5) can be further simplified as:

$$\begin{aligned} w_{TC1} &= \frac{R_{TC1}(1 - R_{TC2}^2)(1 - R_{TC3}^2)}{S} \\ w_{TC2} &= \frac{R_{TC2}(1 - R_{TC1}^2)(1 - R_{TC3}^2)}{S} \\ w_{TC3} &= \frac{R_{TC3}(1 - R_{TC1}^2)(1 - R_{TC2}^2)}{S} \end{aligned} \tag{7}$$

$$S = R_{TC1}(1 - R_{TC2}^2)(1 - R_{TC3}^2) + R_{TC2}(1 - R_{1TC}^2)(1 - R_{TC3}^2) + R_{TC3}(1 - R_{TC1}^2)(1 - R_{TC2}^2)$$

R_{TCi} denotes Pearson correlation coefficients between product i and the SM truth using the TC approach.

According to Equation (7), w_{TC} represents the contribution each product makes to the combination, expressed as a function of the various correlations needed in its specification. When considering Equation (5) and Equation (7) together, however, this compatibility between the TC-based data-truth R and w_{TC} can be distorted for the following two reasons: 1) R's occurring with opposite signs; and 2) violations of the independence between the selected datasets (e.g., non-zero error cross-correlations). It should be noted that Equation (7) cannot directly be used to derive w_{TC} when any of the underlying TC assumptions is not satisfied. In practice, these factors can also affect the consistency between the conventional R and w_{con}. As such, the consistency between the optimal weights and the R values and the impacts of the error cross-correlations on their consistency were further explored by a set of synthetic experiments described in Appendix.

5.2.3. Evaluation Strategy

To identify the areas of relative performance for each parent product, evaluations and comparisons were performed at two levels: overall and conditioned by surface conditions. In addition, the consistency between the TC-derived data-truth R and the conventional R, and between the R values and the corresponding optimal weights in performance sequence has been separately assessed to increase the reliability of obtained results.

Specifically, we first compared the distributions of the conventional R and the TC-based R for all the parent products at a global scale. The characteristics of the temporal performance for each satellite-based SM dataset were further investigated across various static conditions including climate and land cover classes, and mean vegetation water contents. The reasons behind the similarity and difference of temporal performances described by the above two R types were studied and discussed. It should be stressed that the comparisons about the conventional R and the TC-based R are performed and summarized by applying three different reference products (in-situ measurements, ERA-Interim, and MERRA2 in supplementary material).

Given the SM retrievals derived from the discrete satellite observations over the studying period, the evaluation results for the remotely sensed products are prone to certain uncertainties resulting from sampling errors. Due to the strong autocorrelation in the SM data, the optimum block lengths commonly used in the moving block bootstrapping are almost identical to the study period while that of SM anomalies becomes to several days (Chen et al. 2018). Therefore, we calculated the confidence intervals (CI) by simply using randomly selected 100 of 1-year blocks over the study period by which the seasonality effects on the uncertainty can be ignored. For this we assumed that the interannual variability over the 3-year period is ignorable. Here, the CIs were still estimated using the differences between the 97.5th and 2.5th percentiles of the R values obtained in the repeated simulations to demonstrate the reliability of results.

In terms of the relative performance, the consistency between the R and its derived optimal weight was analyzed via a set of synthetic experiments and the results are available in Appendix. Additionally, the impacts of the error cross-correlations on the numerical values of the optimal weights were also investigated to improve the interpretation of the relationship between the optimal weight and the R. Using the reference products and the SM truth as the combination

objectives, the global distributions of the optimal weights calculated using the conventional R and the TC-based R were demonstrated and compared.

5.3. Data

As summarized in **Table 5.1**, various datasets covering the 3-year study period (April 1, 2015-March 31, 2018) have been used in this study. These include 1) three satellite-derived SM products as parent products: SMAP, SMOS-IC, and ASCAT; 2) two reanalysis products as references: ERA-Interim and MERRA2; 3) three static conditions: climate zone (CZ), land cover (LC) and mean vegetation water content (VWC). Additionally, in-situ SM measurements from 191 stations of the International Soil Moisture Network (ISMN) (Dorigo et al. 2011b) were also selected as the reference dataset (**Table 5.3** and **Figure 5.1**).

Table 5.1. Summary of datasets used in this study.

Data	Data source and/or product name	Reference	Resolution (temporal/spatial)	Unit
Satellite SM	SMAP L3 soil moisture product (Version 5)	O'Neill et al. (2018)	Daily, overpass (asc/des) at 6 PM/AM LST /36-km EASE-Grid	m^3/m^3
	SMOS-IC L3 soil moisture product (Version 105)	Fernandez-Moran et al. (2017a); Fernandez-Moran et al. (2017b)	Daily, overpass (asc/des) at 6 AM/PM LST /25-km EASE-Grid	m^3/m^3
	ASCAT L2 soil moisture index product (Version 5)	Wagner et al. (2013); Naeimi et al. (2009)	Daily, overpass (asc/des) at 9:30 PM/AM LST /25km	degree of saturation (0–100%)
Reference SM	MERRA2 topsoil layer soil moisture consent SFMC (M2T1NXLND) * in supplementary materials	Reichle et al. (2017)	Hourly(time-averaged)/0.5°× 0.625°	m^3/m^3
	ERA-Interim volumetric soil water layer 1	Dee et al. (2011)	6-hourly/25 km	m^3/m^3
	131 stations from the ISMN	Dorigo et al. (2011b)	Hourly/point	m^3/m^3
CZ	Updated world map of the Köppen-Geiger climate classification	Peel et al. (2007)	-/0.25°	-
LC	MODIS(MCD12Q1)	NASA-LP-DAAC (2012)	Yearly/0.05°	-
VWC	SMAP VWC (SPL3SMP_E)	Chan et al. (2013)	Daily/9km	kg/m^2

where asc/des: ascending/descending; LST: local solar time; EASE: Equal-Area Scalable Earth. SMAP, MERRA2 and MODIS products are available from the NASA Earthdata website (https://earthdata.nasa.gov/); SMOS-IC:https://www.catds.fr/Products/Available-products-from-CEC-SM/SMOS-IC; ASCAT: https://www.eumetsat.int/website/home/Data/index.html;

Note that a common spatial resolution of 36km × 36km was adopted for all datasets used in this study based on the global cylindrical 36km Equal-Area Scalable Earth, version 2 (EASEv2) (Brodzik et al. 2012). To this end, all data from the products where spatial resolutions do not conform with the common resolution, were resampled using bilinear interpolation unless otherwise stated. Additionally, the minimum number of corresponding observations over the studying period has been set as 50. For the SM data from the hourly updated reference datasets (i.e., ERA-Interim, MERRA2, and the in-situ observations), only the data that are closest to 6 a.m. local solar time (LST) were retained. Given that the ASCAT SM retrievals are based on the observations collected at 9:30 a.m. LST, the assessment of the performance of the ASCAT product is likely to be partially underestimated.

5.3.1. Satellite Soil Moisture Products

5.3.1.1. SMAP

With the aim of mapping the global SM and monitoring the landscape freeze/thaw conditions, the SMAP satellite was launched in January 2015 by National Aeronautics Space Administration (NASA) (Entekhabi et al. 2010a). To incorporate advantages of both active and passive microwave sensors, SMAP satellite carries a radar (1.26 GHz) and a radiometer (1.41 GHz) operated at the L-band frequency, but its radar stopped working in July 2015 due to a hardware malfunction. The nominal spatial resolution of the SMAP radiometer that observes the brightness temperature at the constant 40˚incidence angle. Orbiting in a near-polar sun-synchronous track at the altitude of 685km, the scanning time of the SMAP radiometer is around 6 a.m. LST

(descending) and 6 p.m. LST (ascending). In addition to surface SM products, the SMAP mission can also provide root-zone SM, freeze/thaw conditions, and the net exchange of carbon.

In this study, a descending SMAP L3 Radiometer Global Daily 36 km Equal-Area Scalable Earth (EASE) Grid Soil Moisture (SPL3SMP, hereafter simply referred to as SMAP) has been used. Using the similar criteria applied in previous studies (Zhang et al. 2019), the SMAP SM data retrieved under the following conditions were consistently filtered out: 1) the SM retrievals within the grids where the percentages of the open water area are larger than 10%; 2) SM data obtained from the areas where VWC higher than 5 kg/m^2; 3) the SM retrievals in the pixels where the fraction of the frozen condition (LST < 273.15 K) is larger than 10%; 4) not recommended by the retrieval quality flag.

5.3.1.2. SMOS-IC

The SMOS satellite launched by European Space Agency carries the first L-band (1.4 GHz) interferometric radiometer to provide multi-angular and dual-polarization measurements of brightness temperature for global SM and vegetation optical depth mapping (Kerr et al. 2001). The spatial resolution of the SMOS sensor is around 43km (Kerr et al. 2012). Operating in a sun-synchronous orbit, the crossing times at the equator of the satellite are 6 a.m. LST for ascending pass and 6 p.m. LST for descending pass (Fernandez-Moran et al. 2017a). Based on the L-band Microwave Emission of the Biosphere model (Wigneron et al. 2007), the gridded SM retrievals and vegetation optical depths can be estimated by a cost function to minimize the difference between the modeled and observed brightness temperatures (Wigneron et al. 2000).

The SMOS-IC product developed by INRA (Institute National de la Recherche Agronomique) and CESBIO (Centre d'Etudes Spatiales de la BIOsphere) provides surface SM retrievals posted

into the 25km EASE grids at a global scale (Fernandez-Moran et al. 2017b). There are four main differences in the processes of retrieving SM and vegetation optical depth data for the previous SMOS products (i.e. SMOS L2 and L3 datasets) and the SMOS-IC (Fernandez-Moran et al. 2017b). Firstly, the SMOS-IC algorithm assumes that each pixel is homogeneous instead of considering the LC details (4km × 4km) within the grid as the SMOS-L2 and L3 algorithms did. By retrieving the SM and vegetation optical depth over a homogeneous grid rather than deriving the representative pixel SM from the dominant fraction part, SMOS-IC retrievals become independent from the auxiliary datasets including European Centre for Medium-Range-Weather Forecast (ECMWF) SM and Moderate Resolution Imaging Spectroradiometer (MODIS) Leaf Area Index (LAI) to reduce the excess error propagations. Thirdly, new values of effective vegetation scattering albedo and soil roughness parameters (Parrens et al. 2016) based on the International Geosphere-Biosphere Programme (IGBP) classification have been used in the SMOS-IC algorithm after deliberated calibrations and validations (Fernandez-Moran et al. 2017a). Furthermore, the SMOS-IC algorithm selects the SMOS L3 brightness temperature product as the main input for SM and vegetation optical depth retrievals. The adoption of this optimized brightness temperature dataset does not account for the corrections of antenna patterns associated with the viewing angles and azimuth thus improving the computing efficiency (Fernandez-Moran et al. 2017a). To guarantee the quality of SMOS-IC SM retrievals, only the data with ‘recommended quality’ were retained and adopted here. In addition, the scene flags no more than 1 ($SF \leq 1$) and the TB-RMSE ≤ 8 K have also been applied to further remove uncertain SMOS-IC data possibly interfered by the local conditions and strong RFI (radio frequency interference) impacts.

5.3.1.3. ASCAT

ASCAT is a real aperture radar currently flown onboard the Meteorological Operational A and B satellites (MetOp-A and MetOp-B) which fly in a near-polar sun-synchronous orbit at the altitude of 817 km (Wagner et al. 2013). With a spatial footprint varying from 25 km to 34 km, ASCAT sensors are operated at the C-band (5.255 GHz) in a VV polarization. Three antennas (on each side of the satellite) illuminate a wide swath of 550km and receive quasi instantaneous backscatter power measurements at 3 different azimuth angles and two different incidence angles (Wagner et al. 2013). The equatorial overpass time of the ASCAT is 9:30 a.m. and 9:30 p.m. for the descending and ascending orbits, respectively. The retrieval of the relative SM (i.e. degree of saturation) from ASCAT measured backscatter coefficients is based on a semi-empirical change detection method (Hahn 2016; Naeimi et al. 2009). Here, the descending MetOp-B ASCAT soil moisture 25 km sampling Near Real-Time (Version 5) within the grids where the probability of snow and frozen ground below 10%, and the SM estimated retrieval errors below 50%, were selected (Chen et al. 2018).

5.3.1.4. Key Differences in Soil Moisture Retrieval Algorithms

Regarding the three candidates being compared, SMAP and SMOS-IC use brightness temperatures observed from the L-band passive microwave sensors and retrieve SM values using the zero-order radiative transfer model (Mo et al. 1982; Wigneron et al. 2017). ASCAT with active C-band radar can measure the backscatter coefficient and derive the relative SM content using a semi-empirical changing detection method (Wagner et al. 1999). As summarized in **Error! Reference source not found.**, in this study, the key differences in the retrieval algorithms of the SMAP, SMOS-IC and ASCAT are categorized into parameterizations of physical

temperature, surface roughness, vegetation, and dielectric mixing models used convert obtained dielectric constants to SM values.

Table 5.2. Summary of key differences among the SMAP L3, SMOS-IC, and ASCAT retrieval algorithms

Algorithm	SMAP L3 (SCA-V)	SMOS-IC (L-MEB)	ASCAT (TU Wien Change Detection)
Observation	T_B (V-polarization) at a fixed incidence angle of 40°	Multi-angular and dual polarization SMOS L3 T_B	Multi-angular and VV-pol backscatter coefficient σ^0
Physical Temperature	$T_s = T_c$, derived from GEOS-5 model surface and depth products. $C_T = 0.246$	$T_s(T_c)$ derived from ECMWF skin temperature (soil temperature of 0-7 cm topsoil layer). $C_T = (\frac{SM}{W_0})^{b_0}, W_0 = 0.3\ m^3/m^3;\ b_0 = 0.3$	
Roughness Parameter	H_R values from an IGBP class look-up table $N_R = 2$ $Q_R = 0$	H_R values from Parrens et al. (2016) $N_{RP} = -1$ (P = H, V) over low vegetation $N_{RV} = -1, N_{RH} = 1$ over forests $Q_R = 0$	Roughness conditions are assumed to be stable over time Slope σ' and curvature σ''
Vegetation	$\tau = b \cdot VWC$ ω and b vary with land cover classes (IGBP) VWC derived from MODIS NDVI values	τ simultaneously computed with SM and its initial value is 0.5 ω calibrated based on IGBP classifications	Vegetation effects are considered on a seasonal scale-cross over angles θ_d and θ_w Slope σ' and curvature σ'' on a time-window basis Land cover conditions are assumed to be stable over time
Dielectric mixing model	Mironov et al. (2009)	Mironov et al. (2009)	

T_B= brightness temperature; T_s=soil surface temperature; T_c=vegetation canopy temperature; GEOS-5: Goddard Earth Observing System Model, Version 5; C_T = parameters (Choudhury effective temperature scheme); W_0, b_0 = fitting parameters (Wigneron effective temperature scheme);τ=vegetation optical depth; b=vegetation parameter; Q_R = polarization mixing coefficient; H_R = roughness parameter; N_{RP} = roughness parameter accounting for polarization dependency; NDVI=normalized difference vegetation index; ω=effective scattering albedo; ECMWF: European Centre for Medium-Range-Weather Forecast; σ' = slope parameter; σ'' = curvature parameter; θ_d = reference angle for dry condition; θ_w = reference angle for wet condition.

5.3.2. ISMN Soil Moisture Data

The in-situ SM data of 131 stations from 11 networks were used in this study as a reference to calculate the conventional R and as a critical component of the triplets to estimate the TC-derived data-truth R. These stations are mostly distributed over North America, Alaska, Europe, and Asia. Following the same filtering procedures used in Zhang et al. (2019), the stations where in-situ SM measured at a shallowest depth smaller than 10 cm and symbolized as 'good' in the

quality flag were retained. Additionally, the pixels in which the proportions of the topographic complexity and wetland fraction are less than 10% were considered. To avoid an uneven high weight for the pixels, only in-situ SM data from one station was selected to represent SM variations in one grid. In light of this, the stations with the highest average conventional R with the satellite-based SM products considered here were used when multiple stations locate in the same grid (Dorigo et al. 2015). Given that Colliander et al. (2018) has shown that SM observations over 9km and 33km are approximated in depicting the SM temporal variations, the SM averages of in-situ measurements are assumed to be able to represent the areal SM variations at 36 km and have marginal effects on the estimated correlations with the satellite-based products.

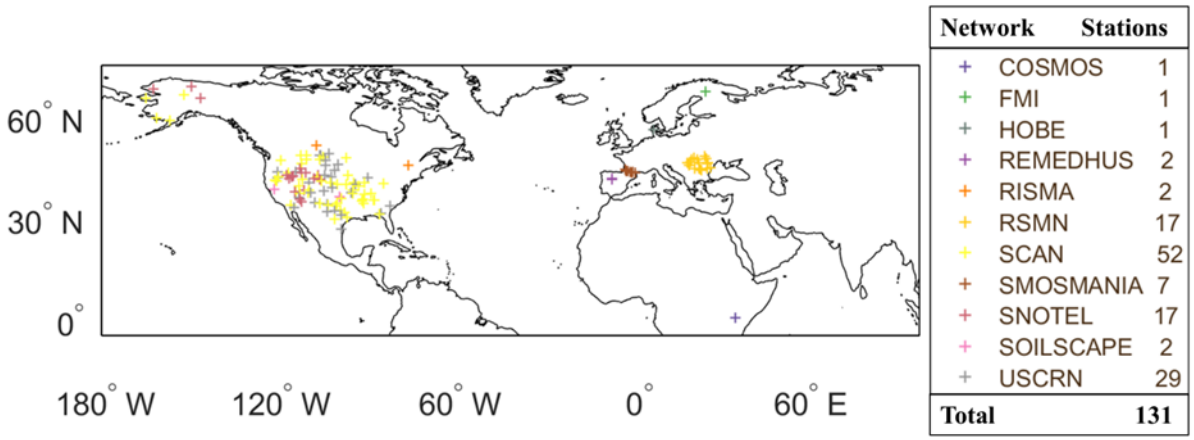


Figure 5.1. The distribution of the 131 ISMN stations was used for this study.

Table 5.3. Summary of in-situ SM used in this study.

Network	Country (Number of stations)	Reference
COSMOS	USA (1 station)	Zreda et al. (2008)
FMI	Finland (1 stations)	http://fmiarc.fmi.fi/
HOBE	Denmark (1 stations)	Bircher et al. (2012)
REMEDHUS	Spain (2 stations)	Sanchez et al. (2012)
RISMA	Canada (2 stations)	(Ojo et al. 2015)
RSMN	Romania (17 stations)	http://assimo.meteoromania.ro
SCAN	USA (52 stations)	http://www.wcc.nrcs.usda.gov/
SMOSMANIA	France (7 stations)	Albergel et al. (2008)

SNOTEL	USA (17 stations)	Leavesley et al. (2008)
SOILSCAPE	USA (2 stations)	Moghaddam et al. (2010)
USCRN	USA (29 stations)	Bell et al. (2013)
Total	131 stations	

** ISMN data are available from: https://ismn.geo.tuwien.ac.at. Where, Unit = m^3/m^3, Resolution (temporal/spatial) = Hourly/point.*

5.4. Results and Discussion

5.4.1. Comparisons of Conventional and TC-Based R over In-Situ Stations

The in-situ SM data were used as the reference to calculate the conventional R for each parent product. In addition, two triplets containing the in-situ SM data (i.e., SMAP, ASCAT, in-situ SM; SMOS-IC, ASCAT, in-situ SM) were also adopted to derive the TC-based data-truth R of the satellite-based products. The above two types of R values were referred to as 'in-situ conR' and 'in-situ TCR' to distinguish from the conventional R (conR) and the TC-based data-truth R (TCR) calculated based on the reanalysis products (ERA-Interim, here). Using the in-situ conR as the benchmark, the absolute values of the differences (ΔR) between different types of R values and the in-situ conR were separately computed over the pixels with the in-situ stations.

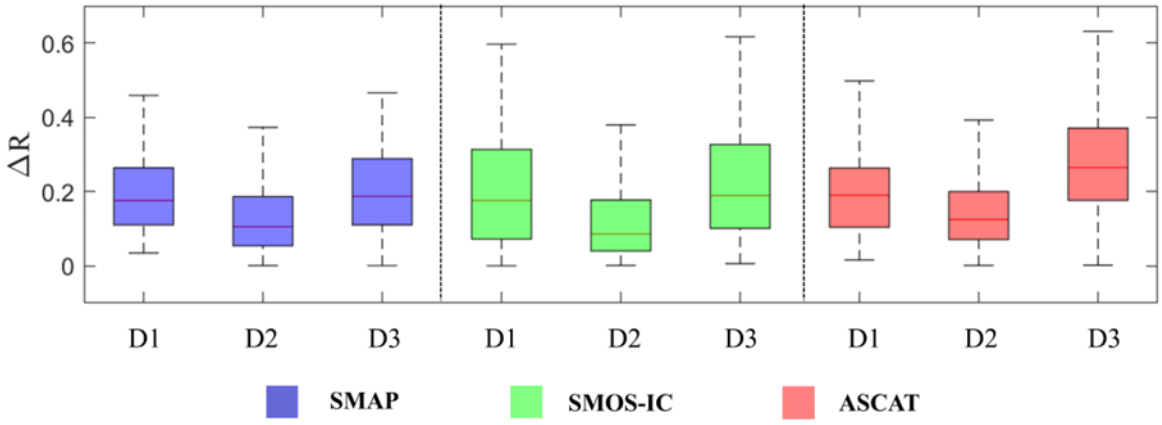


Figure 5.2. Boxplot of ΔR, the absolute values of differences between in-situ conR and the three types of R: in-situ TCR (D1=|in-situ TCR – in-situ conR|), conR (D2 = |conR- in-situ conR|) and TCR (D3 = |TCR – in-situ conR|).

As shown in **Figure 5.2**, ΔR values are generally distributed around 0.2 for each parent product. Despite the inclusion of the in-situ SM data in the triplet, it is to be noted that ΔR values for the in-situ TCR (D1 in **Figure 5.2**) tend to be similar or even higher than those for the TCR (D3 in **Figure 5.2**) across all SM products. The analogous performances of the in-situ TCR and the TCR could be attributed to the similar SM variations captured by the ERA-Interim and the in-situ SM data, which is also demonstrated by the smaller differences (~ 0.1) between the conR and the in-situ conR (D2). As shown in **Figure 5.2**, the conR is expected to be more reliable over the studying places due to the smaller bias with the in-situ conR while directly using the raw SM data.

The TCR is theoretically more reliable than the conR because the conR is highly dependent on the quality of the references (Gruber et al. 2016). However, the results in **Figure 5.2** indicate that the superiority of TCR is not always guaranteed in practice. The TCR based on a triplet can be also degraded because of several factors. In addition to violating some of the TC assumptions (e.g., zero error cross-correlation), different signs and considerably different qualities of the datasets collocated in the triplet could also degrade the quality of the TCR. Rather, if the quality of the references is good enough, the conR could be even better than the TCR as D2 < D1 and D3 in **Figure 5.2**.

5.4.2. Global Maps of Correlation

The global distributions of conR (**Figure 5.3**a–c) and TCR (**Figure 5.3**d–f) are generally consistent in their respective spatial patterns. In view of the areas where both types of R are commonly noted, all three products exhibit good temporal performance over North America, South America, the middle and southern parts of Africa, and most parts in Australia. Compared

to the passive remote sensing SM (SMAP and SMOS-IC), the ASCAT product has considerably low conR values (**Figure 5.3**c) over part of the Middle East where the TCR values of ASCAT are almost at the same magnitudes as SMAP and SMOS-IC's. When using the MERRA2 as the reference (**Figure S9** in supplementary material), similar distribution patterns of its conR and TCR can be found for those satellite-based products.

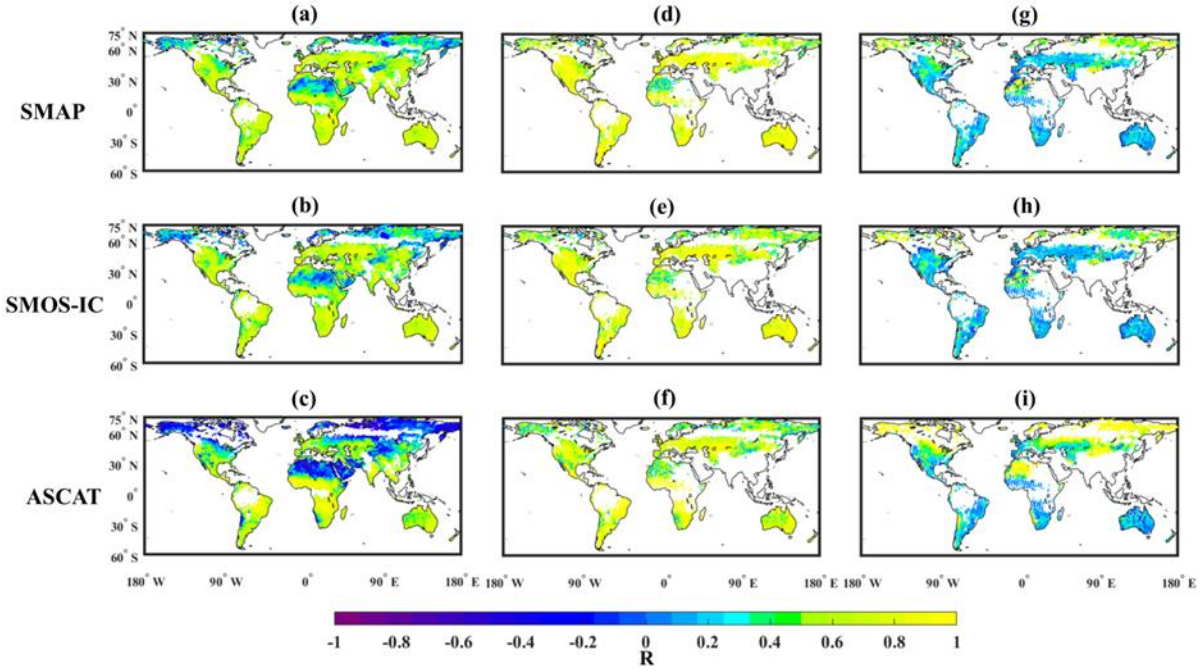


Figure 5.3. Global distribution of conventional R for SMAP, SMOS-IC, and ASCAT using ERA-Interim as reference (a-c), and the TC-based R using ERA-Interim as the component of triplets (d-f). The second column "minus" the first column (g-i).

However, regardless of the conR or the TCR, the temporal performance of SMAP, SMOS-IC, and ASCAT is relatively low in North Africa which is typically defined as arid and semi-arid environments. Over such regions, microwave sensors are likely to suffer challenges in observing relevant signals (Dorigo et al. 2010; Kim et al. 2018). One possible error is the inaccurate estimation of the radiating soil depth and the effective temperature since low-frequency microwave radiation could penetrate deeper soil layer (Holmes et al. 2006; Kim et al. 2018). Moreover, the SM variations within the extremely dry regions are too small to detect by the

scanning instrument or even interfere with the instrument noises. Compared to the SM retrieved from the passive observations, the ability of ASCAT to capture the SM temporal variability over arid and semi-arid areas is relatively lower (more blue areas). The opposing correlations between the modeling products and the active SM retrievals can be attributed to the backscatter returns dominated by the local sub-surface structure and collection geometry in the arid environment (Morrison and Wagner 2019).

In the high latitude areas (i.e., north part of North America, northern Europe, and northeastern Asia), the parent SM products do not have good agreements, especially for the conR values shown in **Figure 5.3**a–c. Similarly, the TCR values of SMAP, SMOS-IC, and ASCAT also moderately reduced. The relatively higher density of soil organic matter over the high latitude areas may account for the low correlations (De Lannoy et al. 2014). The temperature might not be the primary factor to reduce the correlations between the SM retrievals derived from passive microwave observations and the reference products because ECMWF and Goddard Earth Observing System Model, Version 5 (GEOS-5) surface temperature products have been validated with high temporal correlations with the in-situ measurements over almost all climate conditions (Ma et al. 2019). Moreover, the retrieval of SM from active microwave observations is less sensitive to surface temperature variations.

However, differences between the conR and the TCR commonly exist and are around positive 0.2 over most regions (**Figure 5.3**g–i). The larger deviations in the high-latitude areas and North Africa are possibly caused by the opposite signs of R values. In addition, the distributions of the products with the highest TCR and conR values are different overlapping pixels. For example, the proportion of grids in which the ASCAT product has a relatively superior temporal performance than SMAP and SMOS-IC are around 27% (for TCR) and 15% (for conR).

Moreover, the number of grids with effective TCR is only around 40% of that of pixels with the valid conventional R, which can also be clearly observed in **Figure 5.3**. Furthermore, the number of pixels with the highest TCR values of SMOS-IC is half of the number of grids with the highest conR values of the same product. Such discrepancies are partly caused by the reversed correlation signs of the TCR and the conR values, especially in the high-latitude regions for ASCAT. In light of this, for ASCAT, the spatial correlation between the TCR and conR is only 0.56 while those for SMAP and SMOS-IC are moderately higher (0.62 for SMAP; 0.71 for SMOS-IC).

The possible reasons that lead to the differences between the conR and TCR values can be summarized as the violation of TC assumptions, the sign effect, and sampling numbers. Theoretically, while the TC assumptions are fully satisfied, the TCR is expected to be more accurate because the use of conR involves added contribution from the reference product (Gruber et al. 2016). In practice, however, the TC related assumptions are more or less violated in particular when the data from satellite-observation datasets and the land surface modeling products are directly adopted. In terms of calculation equations related to TCR (Gruber et al. 2016; McColl et al. 2014), the non-zero error cross-correlations can straightforwardly affect the R values (Yilmaz and Crow 2014).

The sign compatibility between the observation products and the truth needs to be carefully considered. Additionally, when the SM temporal variations described by one dataset are largely different from the other two products, the unreliable estimations of the TCR, such as the square TCR values larger than 1, would occur. Although the pixels with either negative or more than 1 square TCR values are commonly filtered out, the cases where two products are negatively correlated with the third one likely exist and generate unsound TCR.

In addition, the reduction of the sample numbers due to the data collocation can also affect the TCR. The effective time-series data of the satellite-based SM product have to be curtailed to accommodate the TC requirements. Nevertheless, the previous study has shown that the temporal sampling frequency can also influence evaluating the quality of SM retrievals (Al-Yaari et al. 2019).

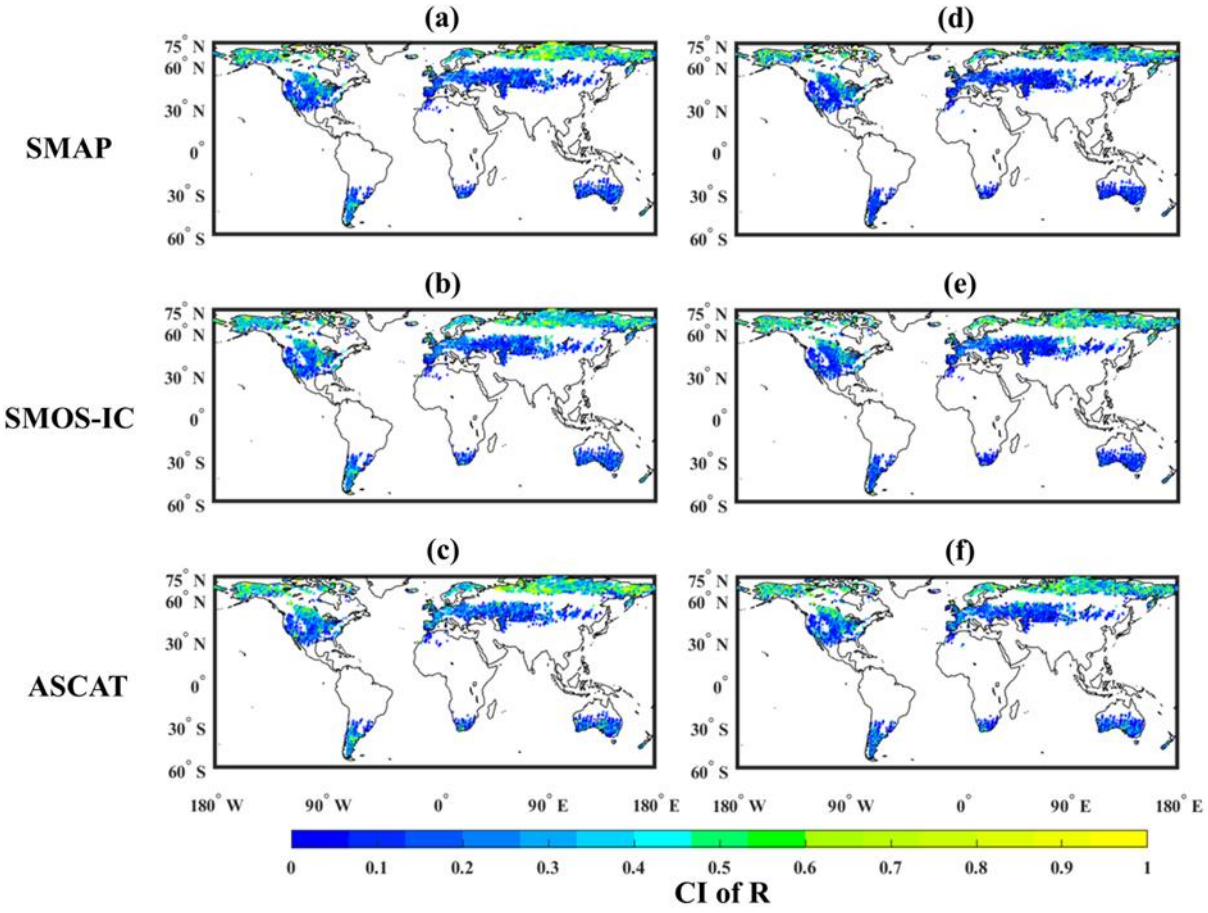


Figure 5.4. Global distribution of the confidence intervals (CI) of the conventional R using ERA-Interim for (a) SMAP, (b) SMOS-IC, and (c) ASCAT, as well as the TC-based R for (d) SMAP, (e) SMOS-IC, and (f) ASCAT via resampling. To demonstrate the reliability of the conR and the TCR derived here in terms of the impacts from the sampling, the CI were estimated and analyzed by estimating the CI via the resampling based on the randomly one-year blocks. As

shown in **Figure 5.4**, the CIs for the conR and the TCR in this study are mostly around 0.1 except the high-latitude regions. Compared to the passive remote-sensing SM products, the CIs of the ASCAT are relatively larger. Generally, the conR and the TCR values used here are affected by sampling procedures to a limited extent when the minimum observations over one-year period is set as 30.

5.4.3. Correlations across Selected Conditions

Since the performances of the SM retrievals from remotely sensed observations vary under different spatial settings, an investigation of the temporal performances of different datasets over specific regions is necessary. In addition to recognizing the accuracy of SMAP, SMOS-IC, and ASCAT in a particular class of areas, this facilitates the exploration of reasons leading to the performance difference. Given that the different amplitudes of variations of the R values and the optimal weights may complicate the understanding of the evaluation results, the main text only focuses on the TCR and the conR. The distributions of their derived highest optimal weights are consistent with the R values and are provided in the Supplementary Materials. Here, three important static factors related to the algorithm parameters and SM variations are considered, which are climate zone (CZ), land cover (LC), and mean vegetation water content (VWC).

5.4.3.1. Climate Zone

Firstly, global CZ is provided by an updated global map of the Koppen-Geiger climate classification (Peel et al. 2007). The climate classifications in this map involve five major climatic types including Tropical, Arid, Temperate, Cold, and Polar, and can be further categorized into 30 sub-classes according to the local status of temperature and precipitation. The secondary classification of CZ has been adopted here: Tropical (Af: rainforest, Am: tropical

monsoon, Aw: tropical savannah); Arid (BW: arid desert; BS: arid steppe); Temperate (Cs: temperate with dry summer; Cw: temperate with dry winter; Cf: temperate without dry season); Cold (Ds: cold with dry summer; Dw: cold with dry winter; Df: cold without dry season); Polar (ET: polar tundra, EF: polar frost). **Figure 5.5Error! Reference source not found.** describes the distributions of R values under different climate conditions. Note that the rainforest regions (Af) are only shown for the conventional R because of the small number of samples in the TC-based R.

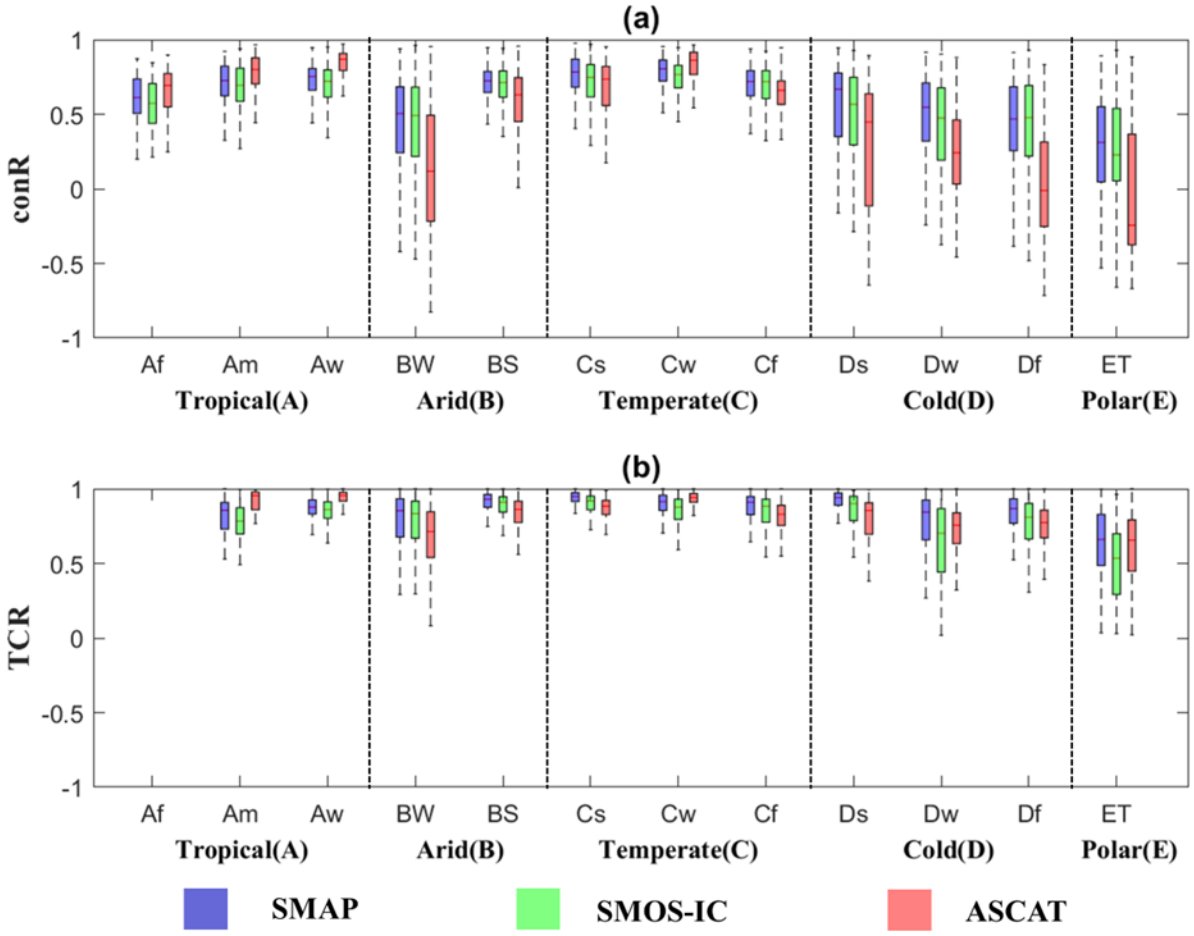


Figure 5.5. Boxplots show variations in the conventional R (a) against ERA-interim and TC-based R (b) under different climate zones. Tropical (Af: rainforest, Am: tropical monsoon, Aw: tropical savannah). Arid (BW: arid desert; BS: arid steppe). Temperate (Cs: temperate with dry summer; Cw: temperate with dry winter; Cf: temperate without dry season). cold and polar (Ds: cold with dry summer; Dw: cold with dry winter; Df: cold without dry season; ET: polar tundra).

In general, the TCR values (**Figure 5.5**b) are overestimated compared to those from conR (**Figure 5.5**a), but their ranking patterns are similar to each other. It should be noted that negative signs in the conR broadly appear while the signs of the TCR are generally positive. In particular, the interquartile range of ASCAT is fairly well distributed in the regions exhibiting negative values over Arid, Cold and Polar climate conditions. Therefore, together with estimating the magnitude of R, the sign of the TCR should be reasonably determined by which the optimal weight can be estimated more accurately.

In case of the conR (**Figure 5.5**a), the SMAP and SMOS-IC products display comparable and better R values higher than 0.6 over most climatic regions. Nevertheless, all three SM products present relatively low correlations in the BW (arid desert) and ET (polar tundra) regions. ASCAT has marginally but consistently better performance than SMOS-IC and SMAP products over tropical areas. This result is consistent with the previous study that demonstrates ASCAT has better correlations with the in-situ measurements over the Equatorial climate (Al-Yaari et al. 2019). For the BW condition (desert), SMOS-IC and SMAP can better capture the SM dynamics, which conforms with the conclusions of Wagner et al. (2013). Although the derivation of SMOS-IC relies on the multi-angular and dual-polarization brightness temperatures and has potential of error propagation from the measurement inaccuracy (O'Neill et al. 2015), over the temperate and cold regions, SMAP and SMOS-IC exhibit overall comparable and better performances than ASCAT except for Cw (temperate with dry winter). Furthermore, the temporal performance of ASCAT is sharply degraded over the cold and polar regions.

5.4.3.2. Land Cover

Figure 5.6Error! Reference source not found. presents the distribution of R values of SMAP, SMOS-IC, and ASCAT under various LC conditions using boxplots. Here, the map of vegetation

types is represented by an MCD12C1 (Version051) product (Friedl et al. 2010) based on the 17-class IGBP classification (Loveland and Belward 1997) from MODIS observations. Specifically, the IGBP classification includes: ENF (evergreen needleleaf forests); EBF (evergreen broadleaf forests); DNF (deciduous needleleaf forests); DBF (deciduous broadleaf forests); MF (mixed forests); CS (closed shrublands); OS (open shrublands) (OS); WS (woody savannas); S (savannas); PW (permanent wetlands); G (Grassland); C (croplands); UBL (urban and built-up lands); CNV (cropland and natural vegetation mosaics); SI (snow and ice); BSV (barren and sparsely vegetated areas); W (water).

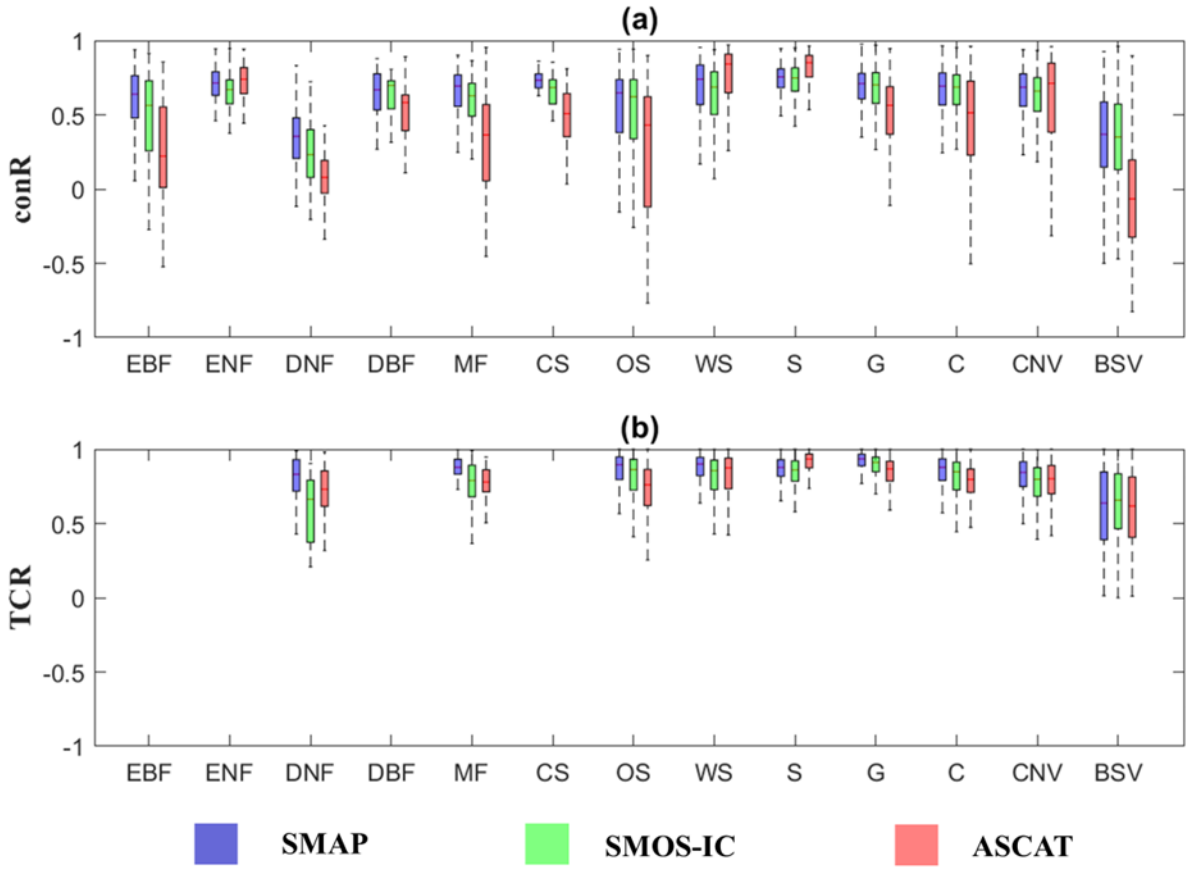


Figure 5.6. Boxplots show variations in the conventional R (a) against ERA-interim and TC-based R (b) under different land cover conditions. EBF (evergreen broadleaf forests); ENF (evergreen needleleaf forests); DNF (deciduous needleleaf forests); DBF (deciduous broadleaf forests); MF (mixed forests); CS (closed shrublands); OS (open shrublands) (OS); WS (woody savannas); S (savannas); G (grassland); C (croplands); CNV (cropland and natural vegetation mosaics); BSV (barren and sparsely vegetated areas).

Similarly, the characteristics of the temporal performances represented by the TCR and conR, are overall consistent. Due to the sufficient samples spanning across various LC types, more forest regions are included in **Figure 5.6**a. In general, SMAP and SMOS-IC have superior performance than ASCAT over most LC conditions. However, ASCAT exhibits advantages in savannas regions (WS and S) that typically correspond to the tropical zone. The strength of ASCAT in savannas can be attributed to the high sensitivity of the active microwave instrument to SM over sparsely vegetated areas (Wagner et al. 2013). It seems that all products have difficulty to capture the temporal SM variations over the DNF (deciduous needleleaf forest) and BSV (barren and sparsely vegetated) areas. The poor temporal performances in the regions of DNF may partly account for the lower R values in the high-latitude areas. In contrast to the SM retrievals from ASCAT of which R values greatly fluctuate with varying LC types, the temporal performances of the passive remote sensing SM data are relatively stable and comparable. Dong et al. (2020) showed that SMOS-IC has a stronger ability in capturing the temporal variations of the SM climatology than several land surface modeling products over cropland. In light of this, the conR values of SMOS-IC and even SMAP over certain LC terrains may be slightly underestimated when selecting the modeling products as references.

5.4.3.3. Mean Vegetation Water Content

Figure 5.7 illustrates the distribution of R values of the satellite-based observation products across various VWC intervals. The average VWC values over the study period have been computed and mapped using the VWC data from the SMAP ancillary datasets (Chan et al. 2013).

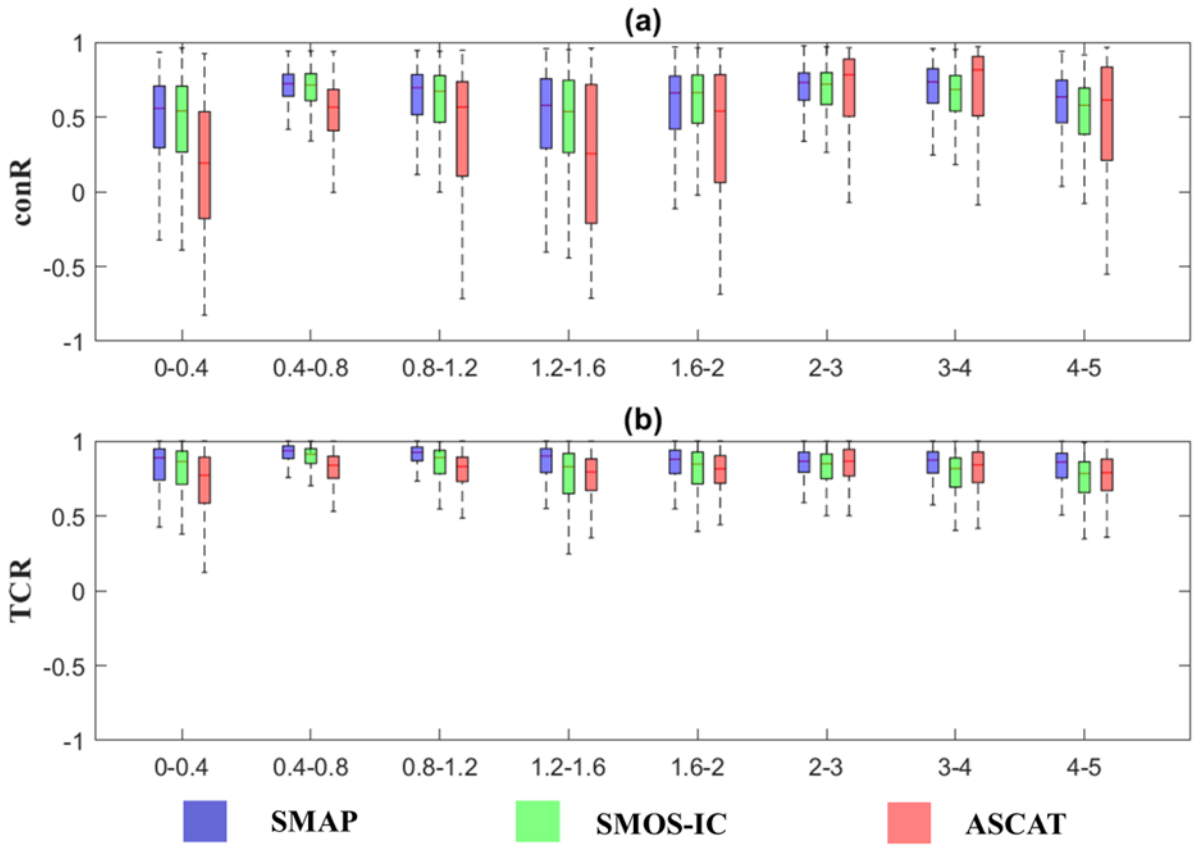


Figure 5.7. Boxplots show variations in the conventional R (a) against ERA-interim and TC-based R (b) under different mean VWC ranges in kg/m^2.

SMAP and SMOS-IC have better temporal performances than ASCAT product from 0–2 kg/m^2. However, the superior magnitudes change with the variations of VWC. The largest discrepancies between the passive and active remote sensing SM data occur in the VWC intervals of 0–0.4 kg/m^2 and 1.2–1.6 kg/m^2. Such abrupt performance degradation of ASCAT over the moderate vegetation regions (1.2–1.6 kg/m^2) is consistent with Al-Yaari et al. (2019) which indicated that the performance of ASCAT would degrade over the vegetation range with Leaf Aare Index (LAI) from 0 to 3 m^2/m^2. Here, 3 m^2/m^2 of LAI can be approximated to 1.5 kg/m^2 of VWC (Wigneron et al. 2006), corresponding to the ASCAT decline interval in this study. However, ASCAT is more advantageous than SMAP and SMOS-IC over regions with 2–5 kg/m^2 of VWC. This result matches Kim et al. (2018) claiming that ASCAT data have the capacity to accurately reproduce

the SM temporal pattern in areas with high vegetation density, considering vegetation optical depth (VOD) from 0 to 0.6. Here, the VOD range from 0 to 0.6 can be approximately as the VWC ranging from 0 to 6 kg/m^2. Lastly, the R values of SMAP, SMOS-IC, and ASCAT become closer over 4–5 kg/m^2 of VWC.

5.4.4. Global Maps of Optimal Weight

Figure 5.8 illustrates the spatial distribution of the best product with the highest optimal weight (w) in terms of maximizing R using ERA-Interim (w_{con}) and the SM truth (w_{TC}) as the objective products. The line plots in the right-hand panel present the variations of zonal mean and one standard deviation of the optimal weights along with the latitude for each parent product. It should be stressed that the w_{TC} was calculated based on Equation (5) rather than Equation (7) since the underlying TC assumptions cannot be fully satisfied while directly using the SM raw data.

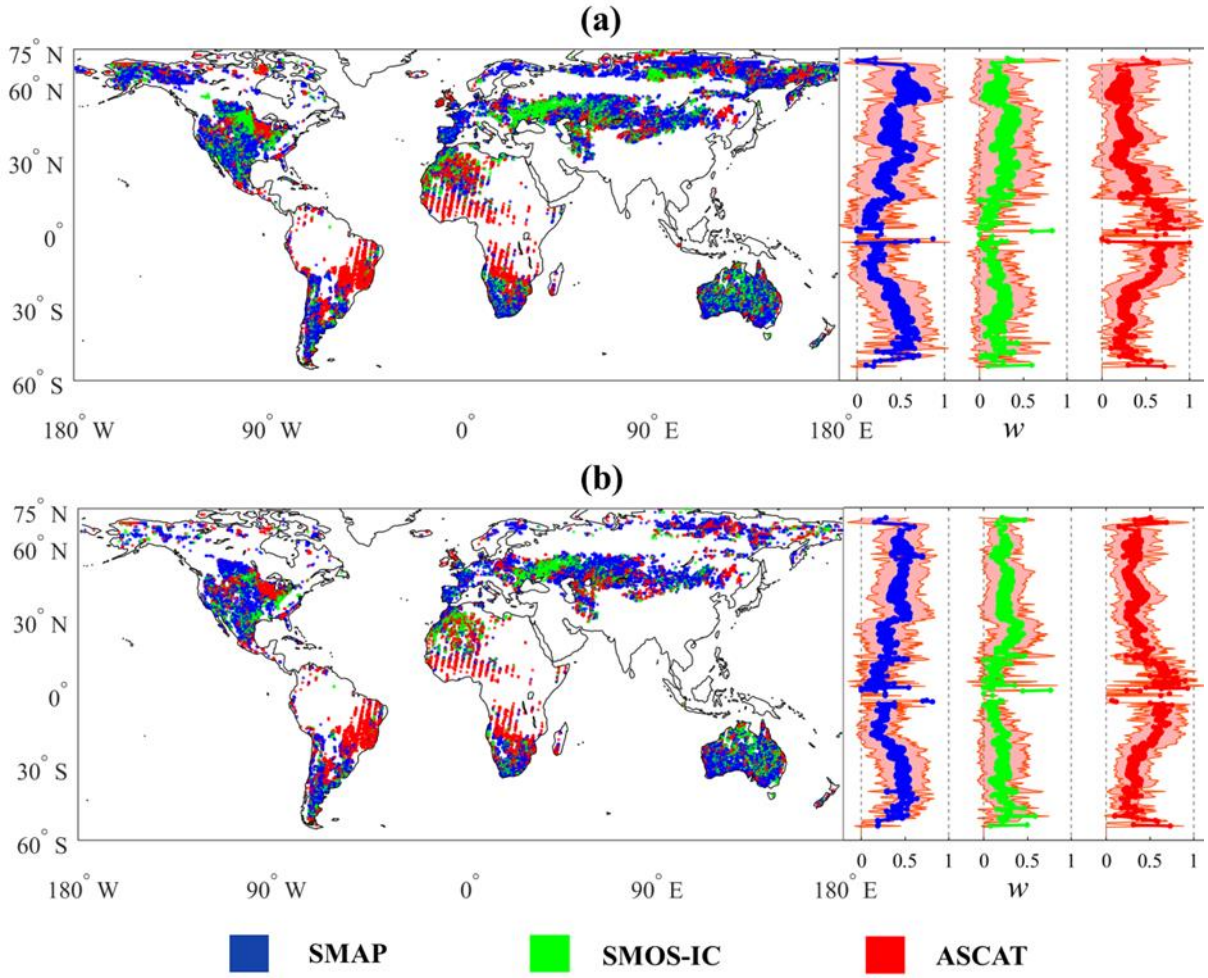


Figure 5.8. Global maps indicating areas exhibiting the highest optimal weight (w) for SMAP, SMOS-IC and ASCAT based on (a) w_{con} and (b) w_{TC}. The line plots in the right-hand panel present variations of zonal mean optimal weights with $\pm$ 1 standard deviation.

Before analyzing their relative temporal performances, the consistency between the optimal weights and the R values has been examined. The Spearman correlations between the w_{TC} (w_{con}) and the TCR (conR) are 0.83 (0.71), 0.75 (0.72), and 0.77 (0.65) for SMAP, SMOS-IC, and ASCAT. Compared to the results from the synthetic experiments in the Appendix, the correlations have slightly declined. In terms of the relatively lower Spearman correlations under the reference condition, this could be caused by the simultaneous use of the positive and negative conR values, resulting in the unpredictable performance sequences reflected through the w_{con}. Moreover, the correspondences between the highest R and the highest optimal weights are

separately 93% (w_{TC} and TCR) and 86% (w_{con} and conR). Therefore, the highest optimal weight can represent the best dataset with the highest contribution to capture the temporal variability of SM.

The complementarity among the SM products can be observed across the geographical areas. The spatial patterns of the highest w_{con} and w_{TC} are overall consistent. The advantageous areas of SMAP include the middle part of North America, most areas in Eurasia, western South America, and southern Africa while SMOS-IC has relatively better performance over eastern North America, several regions in Asia as well as middle Australia. Additionally, ASCAT shows superior performance in eastern South America and most regions in Africa. The latter result is consistent with Al-Yaari et al. (2019) indicating that ASCAT has higher correlations against the in-situ measurements (9 stations) in Africa than SMAP and SMOS products. Furthermore, the advantageous areas of SMAP and ASCAT showing in **Error! Reference source not found.** are similar to Kim et al. (2018) that compared the performance of SMAP, AMSR2, and ASCAT using the TC approach.

According to **Table 5.4**, SMAP and ASCAT have more the highest w_{TC} than SMOS-IC. In terms of the w_{con}, the amounts of grids where the SMAP, SMOS-IC, and ASCAT have the highest optimal weights are comparable. The high proportion of ASCAT with the highest optimal weights can be partially attributed to its negatively high R values distributed over the high-latitude regions and North Africa. In light of this, containing certain areas where all three products have poor or contrasted performances is likely to affect the overall performance evaluation of a dataset. In a view of visual inspection, however, the advantageous regions of SMAP product cover more diverse areas on a global scale.

Table 5.4. Proportions (%) of highest weights of three parent products by different references, where the numbers in brackets are the number of pixels.

Reference or its corresponding triplets	SMAP	SMOS-IC	ASCAT	Total
ERA-Interim (w_{con})	34 (18614)	32 (17495)	34 (18653)	100 (54762)
TC (ERA-Interim) (w_{TC})	47 (8935)	21 (3953)	32 (6128)	100 (19016)
MERRA2 (w_{con})	30 (16243)	31 (16879)	39 (21651)	100 (54773)
TC (MERRA2) (w_{TC})	41 (6619)	18 (2899)	42 (6788)	100 (16306)
Average	38	26	37	100

If similar performances of a product (e.g., SMAP) over a certain geographical area or an equivalent static condition were found using different products (ERA-Interim and the SM truth), the results are expected to be more reliable. Here, similar performances can be roughly evaluated from the degree of proximity in the obtained R or optimal weight, and their rank-order variations or one dataset compared to the other two products. Both the conR and TCR values derived based on the ERA-Interim (**Figure 5.3**) and MERRA2 (**Figure S9**) are also close. In terms of the zonal mean optimal weights (the right panels of **Figure 5.8** and **Figure S13**), the varying patterns of the zonal optimal weights and their standard deviations are similar in the two cases that applied MERRA2 and ERA-Interim as the references.

When it comes to rank-order variations using MERRA2, the advantaged regions of SMOS-IC are further extended to partial Asia and western Australia (**Figure S13**: both conR and TCR calculated using the MERRA2). Given the highly consistent spatial distributions of w_{con} and w_{TC} derived using the same reanalysis product (reference or triplet component), such dissimilar performances are mainly due to the inherent differences in the two reanalysis products. However, for each parent product, the degree of differences between the advantageous regions described by the w_{con} and w_{TC} calculated using the ERA-Interim and MERRA2, is expected to be acceptable. SMAP and SMOS-IC displayed highly comparable and stable temporal performances at a global

scale and therefore, the rank order of their approximate R values possibly altered when applying different reanalysis products.

5.4.5. Limitations and Outlook

It should be noted that there are some limitations to this study. Firstly, the above results and discussions are based on SM datasets over a full three-year period from April 1, 2015, to March 31, 2018. SM data covering longer periods are required to evaluate the performance of remotely sensed SM products more accurately and prove the consistency of the results, especially for the TC estimation.

In addition, the SMOS-IC and ASCAT products being compared were first resampled to the SMAP standard spatial resolution (36-km EASE here), and then all products rescaled to the references by a linear normalization approach (Draper et al. 2009). The implementations of these processes inevitably impact the results of inter-comparisons. Furthermore, for the references, only the data nearest to 6 a.m. were selected and used. Hence, the performance of ASCAT in which ascending SM retrievals are normally obtained at 9:30 a.m. may be slightly degraded.

Lastly, this paper only investigated the impacts on the temporal performances of parent products for static conditions without considering influences caused by dynamic factors. Static conditions are mostly defined by average values and cannot reflect the dynamic quality of SM retrievals precisely (Zhang et al. 2019). Given this, the effects on temporal correlations and their derived optimal weights by dynamic factors such as land surface temperature and soil wetness should be investigated in future studies. Moreover, for each grid, the numerical relationship between the optimal weights and dynamic conditions has the potential to combine near-real-time merged products that effectively leverage the strengths of various SM products over certain pixels.

5.5. Summary and Conclusions

This study compared three advanced satellite-based SM products (SMAP, SMOS-IC, and ASCAT) at a global scale in terms of temporal correlations against the reference products (in-situ SM, ERA-Interim, and MERRA2) using the conventional R, TC-based data-truth R values, and their derived optimal weights (w_{con} and w_{TC}). Compared to the R values that independently reflect the temporal performances of the satellite-based SM products with respect to reference products, the optimal weights represent the relative performance of one dataset dependent on the other considered parent products in a comparable manner using normalization. In addition to demonstrating the temporal performances on the rank order, the numerical values of the optimal weights show a wide range of change with the unique information contained in the observation products. Specifically, the high error cross-correlation between two observation products with approximate R values can purposely reduce the contribution from the product with poorer temporal performance to inhibit the impacts from the overlapped errors.

Firstly, the conventional R and the TC-based R calculated using the in-situ measurements and ERA-Interim were compared and the differences between both types of R commonly exist. Theoretically, the TC-based R is expected to be able to reflect the correlations between the satellite-based SM datasets and the SM truth and thus more accurate than the conventional R strongly dependent on the quality of the references. In practice, however, the accuracy of the TC-based R could be influenced by several factors including the violation of TC assumptions, the sign effects as well as the sampling procedures. As such, the conR is expected to be more reliable than the TCR and in-situ TCR over the pixels with the in-situ stations due to the smaller bias with the in-situ conR. Despite that, for each satellite-based SM product considered here, its

temporal performances exhibited by the conR and the TCR are highly consistent in terms of the spatial distribution even adopting different references (ERA-Interim and MERRA2).

Regarding the temporal performance, three datasets suffer challenges in retrieving SM over the arid environment and high-latitude regions. Meanwhile, there are some areas where the temporal performances of the studied products are commonly good, such as the most available regions in North America, South America, and Australia, as well as the middle and south parts of Africa. Compared to ASCAT, the SM retrievals from SMAP and SMOS-IC have shown good abilities to reproduce the SM variations over large areas of Eurasia.

Furthermore, global assessment and inter-comparison of the temporal performances under various conditions including climate, land cover, and mean vegetation water content, were conducted. Again, the conventional R (conventional R derived optimal weight) and the TC-based R values (TC-based R derived optimal weight) have overall consistent distribution patterns across diverse static circumstances. In most climate zones, passive remote sensing SM retrievals exhibit stable and higher R values than ASCAT. Specifically, SMAP and SMOS-IC data present substantial advantages in the desert and polar regions. However, ASCAT shows extremely high R values over tropical zones. Concerning the LC, SMOS-IC and SMAP SM retrievals have strong capabilities to reproduce SM variations within most LC types except the DNF (deciduous needleleaf forest) and BSV (barren and sparsely vegetated regions). Despite the high R values of SMAP and SMOS-IC products over WS (woody savannas) and S (savannas) regions, ASCAT exhibits even better temporal performances. Furthermore, the temporal performances of passive remote sensing SM retrievals have superiority in the regions with the VWC intervals of 0–2kg/m^2, especially in 0–0.4kg/m^2 and 1.2–1.6kg/m^2 whereas ASCAT generally outperforms SMAP and SMOS-IC over the areas with the higher vegetation density (2–5kg/m^2).

The violation of the underlying TC assumptions, the optimization strategies as well as the simultaneous use of R values with opposite signs can account for the decreased matching between the R and optimal weight. Therefore, a set of synthetic experiments has been designed with the aim of investigating the impacts of the error cross-correlations. According to the results of the synthetic experiments, moderate and high error cross-correlations can overly extend the distances between the optimal weights of two products with similar R values. Such adjustments can avoid incorporating the overlapped error information contained in the parent products but possibly distort and underestimate the contribution from the signal portion of the product. However, in terms of rank order, the consistency between the R and the optimal weight is rarely affected by the error cross-correlations, which is demonstrated by the Spearman correlations continuously more than 0.9. In practice, the Spearman correlations between the optimal weights and the R values are slightly reduced to around 0.8. Therefore, the highest optimal weights have sufficient capacity to represent the performance of the product with the highest R values. As for the advantageous areas of relative performance expressed by the highest optimal weights, the comparably global complementarity can be noticed even though the advantageous areas of SMAP appears to cover more diverse areas from visual observations.

Through the integrated evaluation of the temporal performances of the three SM products, this study can be a guideline for data users to select proper SM datasets for their research. In addition, understanding the optimal weights in-depth could be helpful to linearly combine the product with improved temporal representations.

5.6. Appendix: Results of Synthetic Experiments

For a further insight on the compatibility between the R and the optimal weight, we implemented an experiment using synthetic datasets. Specifically, a group of time-series data representing the truth values (T) was first generated using a lag-1 auto-correlation model of which autocorrelation coefficient is randomly drawn from a uniform distribution from 0 to 1. Then, three sets of observation time-series data (Obs) were created by multiplying with random coefficients (α) from 0 to 1 and adding zero-mean Gaussian random errors (ε) to the truth data ($Obs = \alpha T + \varepsilon$). The variances of the errors were limited to fall in the interval 0 to 0.5. Here, we applied a range of error cross-correlation coefficients from 0 to 0.9 increasing by 0.1 for generating random errors while sequentially increasing dependence among the observations. Since the error cross-correlation coefficients can only be controlled by a product pair, there is an unknown error cross-correlation left when the three sets of error time-series data are generated using the two controlled error cross-correlations. The degree of compatibility degrees between the data-truth R and the optimal weight were estimated using the Spearman rank-order correlation. Furthermore, the highest values (H) of the R and the optimal weight as well as the differences (ΔD) between the highest R (optimal weight) and the second highest R (optimal weight) were compared and discussed. To assess the uncertainty in the results, the synthetic experiment was repeated 1000 times.

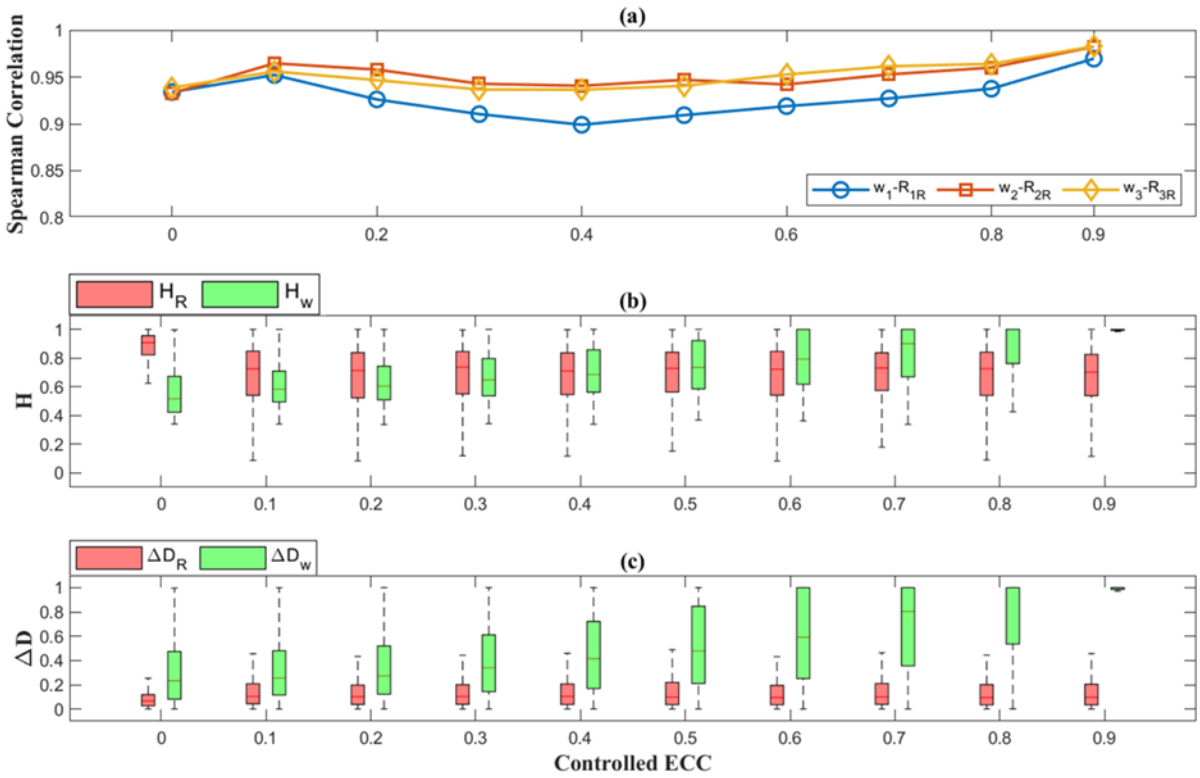


Figure A5.1. Results of the synthetic experiments under the controlled error cross-correlations (ECC): (a) Spearman correlations between the R and the optimal weight (b) boxplot of the highest R (H_R) and the highest optimal weights (H_w) (c) the differences (ΔD) between the highest and the second highest R values (ΔD_R) or optimal weight values (ΔD_w).

According to the **Figure A5.1**a, the Spearman correlations between the R and the optimal weight are generally higher than 0.9 regardless of the increasing error cross-correlations. Therefore, the optimal weight has the ability to represent the relative temporal performance of each parent product consistent with the R values. However, the numerical values of the optimal weight can be impacted by error cross-correlations. With the increase of the error cross-correlations, the highest optimal weights rise to closer to 1 no matter what the highest R values are (**Figure A5.1**b). Additionally, ΔD values of the optimal weights are inclined to rapidly expand once the controlled error cross-correlations are larger than 0.3 (**Figure A5.1**c). In light of this, the optimal weights and the combined product are likely to excessively rely on the parent product with the best performance. While the controlled error cross-correlations are near 0.9, the optimal weights

are mostly concentrated to either 1 or 0. The values under such extreme conditions indicate that the optimal weights can accordingly adjust to avoid overlapping error information from different parent products. Given that, the signal portions of the parent products are also likely to have certain impacts on the optimal weights in a similar manner but at a greatly reduced magnitude. However, the exaggerated values of the highest optimal weights are likely to block the signal portions of the other parent products and underestimate their respective contributions. As a result, the optimal weight can describe the relative temporal performance consistent with the R, avoid the overlapping error information from different parent products in the combination, and have the potential to perform as an indicator of the error cross-correlations.

Chapter 6. Summary and Conclusions

The ultimate objective of this thesis is to improve the scientific quality and utility of soil moisture estimations from passive microwave remote sensing by addressing the identified drawbacks of the current satellite-based soil moisture retrievals from the NASA SMAP mission. Preliminary studies mainly focus on the retrieval quality of L-band soil moisture retrievals over the organic-rich soils and water-included footprints as well as on filling the SMAP temporal gaps through precipitation information. Specifically, the performances of nine advanced dielectric models have been compared over the organic-rich soils, and the evaluation results suggest the separate use of Mironov 2009 and Mironov 2019 in mineral and organic soils could be an optimal option for better accuracy. Meanwhile, the study of evaluating the water temperature of inland water bodies built the foundation for further research in mitigating the effects of water contaminations within the radiometer's instantaneous field-of-view and the associated wet bias when the dynamic water fractions become available.

In terms of temporal continuity, the gaps between SMAP observations have been supplemented through water balance budgeting. Accurately capturing each rainfall-induced soil moisture peak is a major motivation behind this study whilst the optimized hydrologic parameters and the quantified hydrologic loss are also critical. Based on these, future studies will focus on the spatial disaggregation of the satellite-based precipitation product, quality improvement with the combined use of uncertainty analysis and advanced mathematic techniques, as well as an attempt to establish a long-term soil moisture estimation at a global scale using one state-of-the-art satellite-based precipitation product, Multi-Source Weighted-Ensemble Precipitation (MSWEP).

In order to acquire more information on soil water content from radiometer observations, one ongoing research is concentrated on learning the layered radiative transfer model (Appendix from p.141 to p.164) that considers naturally inhomogeneous vertical soil profiles and forward simulations of thermal radiations of near-surface soils. In addition to the expectation of more accurate soil moisture estimates and validation benefits, the greater promise lies in the inversed application of this radiative transfer model for the multifrequency dual-polarization measurements to infer vertical gradients of soil moisture and soil temperatures from various sensors.

It is expected that the outcomes of this thesis can change how radiometer observations are interpreted in the future while the newly yielded soil moisture data with temporal continuity and higher accuracy would help researchers to quantitatively understand the linkages between water balance components and deepen the interpretation of the terrestrial-atmosphere interactions in the context of climate change.

Appendix: Near-Surface Soil Moisture Profile Estimation Through a Four-Layer Radiative Transfer Model

A1. Introduction

Root-zone soil moisture, which represents the water content present in the soil down to tens of centimeters below the land surface, plays an important role in several disciplines, such as ecohydrology, agriculture and meteorology (Akbar et al. 2018; Babaeian et al. 2019; Njoku and Kong 1977). Water content from the deeper layers is directly associated with the transpiration rate and photosynthetic activity of plant and affects large-scale weather forecasting and climate modeling (Dirmeyer 2000; Fatichi et al. 2016; Koster and Suarez 2001; Reich et al. 2018). Unlike readily available surface soil moisture, however, observation-based root-zone soil moisture is scarce primarily because of its greater depths and high spatiotemporal variability (Brakhasi et al. 2023).

As an alternative, researchers have developed a variety of techniques that link surface soil moisture to deeper soil moisture, including exponential filtering, data assimilation, multiple regression, statistical approach, and the use of maximum entropy principle (Baldwin et al. 2017; Carranza et al. 2021; Karthikeyan and Mishra 2021; Mishra et al. 2018; Mishra et al. 2020; Qiu et al. 2010; Walker et al. 2001; Xia et al. 2022). For example, global scale root-zone soil moisture datasets like Soil Moisture Active Passive (SMAP) Level 4 product and the Global Land Data Assimilation System (GLDAS) result from the assimilation of surface soil moisture into land surface models (Brakhasi et al. 2023). However, the lack of priori information of root-zone soil moisture measurements often prevents constructing the relationship between the soil

moisture at the surface and deeper layers. Moreover, the soil moisture evolution with depth is prone to the non-linear variations, largely modulated by atmospheric forcing, as well as vegetation cover (Brakhasi et al. 2023; Hirschi et al. 2014). Such an intricate association is unlikely to be explained by a few simple regression expressions or exponential formulations while the inclusion of multiple inputs of the remaining methods might introduce unanticipated uncertainties.

Microwave radiometry offers a remote sensing tool that is well-suited to the determination of soil moisture since the thermal emission of soils is sensitive to the dielectric behaviors of wet soils in the microwave range (Njoku and Kong 1977). The satellite-based L-band (1.41 GHz) soil moisture products have been widely applied in various hydrological applications and meteorological predictions while the validity of P-band (750 MHz) sensors in monitoring soil moisture variations within a 10 cm depth has also been verified (Shen et al. 2020). Despite this, soil moisture retrievals from the above radiometers typically assume mean soil moisture within a 5 cm or 10 cm soil layer with uniform soil temperature. Naturally occurring soil systems feature heterogeneous subsurface moisture and temperature profiles. Therefore, either the applicability of a constant emission depth across the globe or the appropriateness of hypothesizing a homogeneous soil property under various terrains remains to be further studied.

Based on a stratified non-scattering half-space medium, brightness temperature responses to subsurface moisture and temperature variations can be described by both the sophisticated incoherent model and coherent model (Burke et al. 1979; Njoku and Kong 1977; Schmugge and Choudhury 1981; Wilheit 1978). Excluding the perturbing effects of surface roughness, canopy cover, and subsurface scattering, the inversion schemes of the above models enable the derivation of gradient moisture and temperature in the soil. For instance, Brakhasi et al. (2023)

combined the brightness temperatures observed at L-band and P-band frequencies and estimated the soil moisture profiles under flat bare soil using a coherent approach. Theoretically, the coherent model is considered more accurate than the radiative transfer approach (incoherent) as it accounts for the phase-interference oscillation and thereby performs well over regions with a rapidly vertical moisture variations (Njoku and Kong 1977; Schmugge and Choudhury 1981; Ulaby et al. 2014). However, experimental tests indicated that both types of models have the ability to estimate values of brightness temperatures with reasonable accuracy (Schmugge and Choudhury 1981). Additionally, in nature, the disturbance of surface roughness and horizontal heterogeneity in the soil on the coherent reflections cannot be entirely ignored (Schmugge and Choudhury 1981). Regarding these and the utilization simplicity of the incoherent model, the applicability investigation of the radiative transfer model in the derivation of soil moisture and temperature profiles is necessary.

The ultimate aim of this research is to estimate the soil moisture profiles by means of a four-layered radiative transfer model based on the available multi-frequency dual-polarization satellite-based microwave brightness temperatures. Such a project will start the forward modeling brightness temperatures from a four-layer soil medium. Specifically, layered soil moisture and soil temperature from the ECMWF ERA5 Land (Muñoz-Sabater et al. 2021b) represent averages from 0 ~ 7 cm, 7 ~ 28 cm, 28 ~ 100 cm, and 100 ~ 289 cm will be used as the input for the four-layer radiative transfer model to simulate the brightness temperatures at a global scale. Three scenarios will be considered, which are 1) radiation from the uniform soil moisture and temperature from a soil layer of 0 ~ 7 cm, 2) the aggregated radiation from four soil layers with the uniform soil moisture but nonuniform soil temperature through Lv's scheme (Lv et al. 2014), and 3) the aggregated radiation from four soil layers characterized by varying soil

moisture and temperature using a series of analytical formulations. A comprehensive analysis will be conducted to compare the discrepancies among the simulated brightness temperatures and to distinguish the effects of various inputs. In addition, the analytical formulation of the brightness temperatures potentially builds an effective scheme for estimating the representative radiating temperature. Such a comparison of global diurnal microwave emissions provides the basis for the radiating contributions of each layer across various geographic regions before the model inversion.

A2. Methodology

A2.1. Banded Matrix Solution of Total Brightness Temperature

A2.1.1. Theoretical Formulation: A 3-Layer Stratified Medium

Consider the brightness temperature observed at an angle of θ_0.

The radiative transfer equations that describe the upward intensity $I_{u_1}(z)$ and downward intensity $I_{d_1}(z)$ in layer 1 are:

$$cos\theta_1 \frac{dI_{u_1}(z)}{dz} = -\kappa_{a_1} I_{u_1}(z) + \kappa_{a_1} C_1 T_1 \quad (1)$$

$$-cos\theta_1 \frac{dI_{d_1}(z)}{dz} = -\kappa_{a_1} I_{d_1}(z) + \kappa_{a_1} C_1 T_1 \quad (2)$$

$$C_1 = \frac{KRe(\epsilon_1)}{\lambda^2 \epsilon_0} \quad (3)$$

where K is the Boltzmann's constant $1.38 \times 10^{-23}\, J/K$; λ is the wavelength propagated in the medium; ϵ_1 and ϵ_0 separately represent the absolute dielectric permittivity of layer 1 and free space, and their ratio reflects the dielectric constant of layer 1; θ_1 indicates the propagation direction within layer 1 and can be determined through Snell's law; $Re()$ means the real part of

the term within the parenthesis; z denotes the depth within layer 1; κ_{a_1} is the absorption coefficient of layer 1; and T_1 is the physical soil temperature of layer 1.

After integration, the above equations can be written as:

$$\frac{I_{u_1}(z)}{C_1} = A_1 e^{-\kappa_{a_1} sec\theta_1 z} + T_1 \tag{4}$$

$$\frac{I_{d_1}(z)}{C_1} = B_1 e^{\kappa_{a_1} sec\theta_1 z} + T_1 \tag{5}$$

where A_1 and B_1 are coefficients to be determined.

Similarly, the upward intensity $I_{u_2}(z)$ and downward intensity $I_{d_2}(z)$ in layer 2 are:

$$cos\theta_2 \frac{dI_{u_2}(z)}{dz} = -\kappa_{a_2} I_{u_2}(z) + \kappa_{a_2} C_2 T_2 \tag{6}$$

$$-cos\theta_2 \frac{dI_{d_2}(z)}{dz} = -\kappa_{a_2} I_{d_2}(z) + \kappa_{a_2} C_2 T_2 \tag{7}$$

$$C_2 = \frac{KRe(\epsilon_2)}{\lambda^2 \epsilon_0} \tag{8}$$

with $I_{u_2}(z)$ and $I_{d_2}(z)$ given by:

$$\frac{I_{u_2}(z)}{C_2} = A_2 e^{-\kappa_{a_2} sec\theta_2 z} + T_2 \tag{9}$$

$$\frac{I_{d_2}(z)}{C_2} = B_2 e^{\kappa_{a_2} sec\theta_2 z} + T_2 \tag{10}$$

For layer 3, which is also the last semi-infinite layer, only the upward intensity $I_{u_3}(z)$ contributes to the observed brightness temperature:

$$\frac{I_{u_3}(z)}{C_3} = T_3 \tag{11}$$

$$C_3 = \frac{KRe(\epsilon_3)}{\lambda^2 \epsilon_0} \tag{12}$$

The derivation of Equation 11 will be shown in the later part.

The above equations describe the upward and downward intensities in each layer as a function of z, and their values at the interfaces must match those of the adjacent layers above and below through the reflection and transmission coefficients. For example, r_{12} and t_{12} denote the reflection and transmission coefficients at the bottom boundary of layer 1 (or at the top boundary of layer 2), where $r_{12} = 1 - t_{12}$. One important assumption here is that reflectivity values from below and from above the same boundary are identical, such as $r_{12} = r_{21}$.

For this 3-layer stratified medium, there are 4 unknown coefficients (A_1, B_1, A_2 and B_2) and 3 interfaces along the vertical axis at $z = -d_0 (d_0 = 0), -d_1$, and $-d_2$.

For the first interface at $z = 0$:

$$\frac{I_{d_1}}{C_1}(z = -d_0) = r_0 \frac{I_{u_1}}{C_1}(z = -d_0) + (1 - r_0)\frac{I_{d_0}}{C_0}(z = -d_0) \tag{13}$$

Note that $\frac{I_{d_0}}{C_0} = 0$ because it is assumed that there is no downward intensity coming down from the air.

Therefore,

$$\frac{I_{d_1}}{C_1}(z = -d_0) = r_0 \frac{I_{u_1}}{C_1}(z = -d_0) \tag{14}$$

At the boundary $z = -d_1$ (the bottom of layer 1 and the top of layer 2)

$$\frac{I_{u_1}}{C_1}(z=-d_1) = r_1\frac{I_{d_1}}{C_1}(z=-d_1) + (1-r_1)\frac{I_{u_2}}{C_2}(z=-d_1) \tag{15}$$

$$\frac{I_{d_2}}{C_2}(z=-d_1) = r_1\frac{I_{u_2}}{C_2}(z=-d_1) + (1-r_1)\frac{I_{d_1}}{C_1}(z=-d_1) \tag{16}$$

At the boundary $z = -d_2$ (the bottom of layer 2 and the top of the last layer)

$$\frac{I_{u_2}}{C_2}(z=-d_2) = r_2\frac{I_{d_2}}{C_2}(z=-d_2) + (1-r_2)\frac{I_{u_3}}{C_3}(z=-d_2) \tag{17}$$

Using the expression provided in Equations (4) – (5), and (9) – (11), Equations (14) – (17) could be converted into the form of:

$$B_1 + T_1 = r_0(A_1 + T_1) \tag{18}$$

$$A_1 e^{\kappa_{a_1} sec\theta_1 d_1} + T_1 = r_1(B_1 e^{-\kappa_{a_1} sec\theta_1 d_1} + T_1) + (1-r_1)(A_2 e^{\kappa_{a_2} sec\theta_2 d_1} + T_2) \tag{19}$$

$$B_2 e^{-\kappa_{a_2} sec\theta_2 d_1} + T_2 = r_1(A_2 e^{\kappa_{a_2} sec\theta_2 d_1} + T_2) + (1-r_1)(B_1 e^{-\kappa_{a_1} sec\theta_1 d_1} + T_1) \tag{20}$$

$$A_2 e^{\kappa_{a_2} sec\theta_2 d_2} + T_2 = r_2(B_2 e^{-\kappa_{a_2} sec\theta_2 d_2} + T_2) + (1-r_2)T_3 \tag{21}$$

Re-arrange Equations (18) – (21)

$$r_0 A_1 + (-1)B_1 = (1-r_0)T_1 \tag{22}$$

$$e^{\kappa_{a_1} sec\theta_1 d_1} A_1 + (-r_1)e^{-\kappa_{a_1} sec\theta_1 d_1} B_1 + (r_1-1)e^{\kappa_{a_2} sec\theta_2 d_1} A_2 = (1-r_1)(T_2 - T_1) \tag{23}$$

$$(1-r_1)e^{-\kappa_{a_1} sec\theta_1 d_1} B_1 + r_1 e^{\kappa_{a_2} sec\theta_2 d_1} A_2 + (-1)e^{-\kappa_{a_2} sec\theta_2 d_1} B_2 = (1-r_1)(T_2 - T_1) \tag{24}$$

$$e^{\kappa_{a_2} sec\theta_2 d_2} A_2 + (-r_2)e^{-\kappa_{a_2} sec\theta_2 d_2} B_2 = (1-r_2)(T_3 - T_2) \tag{25}$$

The above equations can then be cast into the banded-matrix equations:

$$\begin{bmatrix} r_0 & -1 & 0 & 0 \\ e^{\kappa_{a_1} sec\theta_1 d_1} & -r_1 e^{-\kappa_{a_1} sec\theta_1 d_1} & (r_1-1)e^{\kappa_{a_2} sec\theta_2 d_1} & 0 \\ 0 & (1-r_1)e^{-\kappa_{a_1} sec\theta_1 d_1} & r_1 e^{\kappa_{a_2} sec\theta_2 d_1} & -e^{-\kappa_{a_2} sec\theta_2 d_1} \\ 0 & 0 & e^{\kappa_{a_2} sec\theta_2 d_2} & -r_2 e^{-\kappa_{a_2} sec\theta_2 d_2} \end{bmatrix} \begin{bmatrix} A_1 \\ B_1 \\ A_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} (1-r_0)T_1 \\ (1-r_1)(T_2 - T_1) \\ (1-r_1)(T_2 - T_1) \\ (1-r_2)(T_3 - T_2) \end{bmatrix}$$

Since the total brightness temperature of the whole stratified medium $T_B = \frac{I_{u_0}}{C_0}(z = -d_0)$,

$$\frac{I_{u_0}}{C_0}(z = -d_0) = r_0\frac{I_{d_0}}{C_0}(z = -d_0) + (1 - r_0)\frac{I_{u_1}}{C_1}(z = -d_0) \tag{26}$$

Recall $\frac{I_{d_0}}{C_0} = 0$

$$\frac{I_{u_0}}{C_0}(z = -d_0) = (1 - r_0)\frac{I_{u_1}}{C_1}(z = -d_0) \tag{27}$$

$$\frac{I_{u_1}}{C_1}(z = -d_0) = A_1 e^{\kappa_{a_1} sec\theta_1 d_0} + T_1 \tag{28}$$

Given $d_0 = 0$

$$\frac{I_{u_1}}{C_1}(z = -d_0) = A_1 + T_1$$

Therefore,

$$T_B = \frac{I_{u_0}}{C_0}(z = -d_0) = (1 - r_0)(A_1 + T_1) \tag{29}$$

As long as the banded matrix solutions become available, in particular A_1, the total brightness temperature T_B can be simply obtained. The Fresnel reflectivity at the boundary $z = -d_0$ could be derived with the knowledge of polarization, the refraction indexes ($n_l^2 = \frac{\epsilon_l}{\epsilon_0}$, l means the layer number), as well as the θ_0 and θ_1 calculated using the Snell's law as mentioned above.

A2.1.2. Theoretical Formulation: An N-Layer Stratified Medium

In general, for an N-layer stratified medium, there are $2N - 2$ equations to solve for $2N - 2$ unknown coefficients.

$$Zx = b \tag{30}$$

For the diagonal elements:

$$Z_{1,1} = r_0 \tag{31}$$
$$Z_{2n,2n} = -r_n e^{-\kappa_{a_n} sec\theta_n d_n} \text{ for } n = 1, ..., N-1 \tag{32}$$
$$Z_{2n-1,2n-1} = r_{n-1} e^{\kappa_{a_n} sec\theta_n d_{n-1}} \text{ for } n = 2, ..., N-1 \tag{33}$$

For the upper off-diagonal elements:

$$Z_{1,2} = -1 \tag{34}$$
$$Z_{2n,2n+1} = (r_n - 1) e^{\kappa_{a_n+1} sec\theta_{n+1} d_n} \text{ for } n = 1, ..., N-2 \tag{35}$$
$$Z_{2n-1,2n} = -e^{\kappa_{a_n} sec\theta_n d_{n-1}} \text{ for } n = 2, ..., N-1 \tag{36}$$

For the lower off-diagonal elements:

$$Z_{2n,2n-1} = e^{\kappa_{a_n} sec\theta_n d_n} \text{ for } n = 1, ..., N-1 \tag{37}$$
$$Z_{2n-1,2n-2} = (1 - r_{n-1}) e^{-\kappa_{a_n-1} sec\theta_n - 1 d_{n-1}} \text{ for } n = 2, ..., N-2 \tag{38}$$

For the matrix x:

$$x_{2n-1} = A_{2n-1} \text{ for } n = 1, ..., N-1 \tag{39}$$
$$x_{2n-2} = B_{2n-1} \text{ for } n = 1, ..., N-1 \tag{40}$$

For the right-hand side:

$$b_1 = (1 - r_0) T_1 \tag{41}$$
$$b_{2n} = (1 - r_n)(T_{n+1} - T_n) \text{ for } n = 1, ..., N-1 \tag{42}$$
$$b_{2n-1} = (1 - r_{n-1})(T_n - T_{n-1}) \text{ for } n = 2, ..., N-1 \tag{43}$$

Once the unknown coefficients are solved, Equation (28) can be applied to calculate the total brightness temperature.

A2.2. Individual Contributions from Sub-Layers to Total Brightness Temperature

In this study, two methods have been adopted to reflect the contributions from individual layers within the stratified medium. The first approach attempts to divide the total brightness temperature in a form of summations of brightness temperatures contributed by each layer, namely $T_B = T_{B_1} + T_{B_2} + T_{B_3} + \cdots + T_{B_N}$. In addition, the total brightness temperature of a stratified system being expressed in terms of layered physical temperatures was also investigated, which is $T_B = S_1T_1 + S_2T_2 + S_3T_3 + \cdots + S_NT_N$. Here, a 3-layer system has been applied as an illustrative example to introduce the derivation processes of the above two methods to partition the total brightness temperature.

A2.2.1. Brightness Temperatures from Individual Layers within a 3-Layer Medium

During this process, the immediate results (i.e., B_n) obtained from the banded matrix solutions have been used. Based on previous derivations, Equation (29) (i.e., $T_B = (1 - r_0)(A_1 + T_1)$) proves that the solution or the way to present A_1 is critical. The integration equation obtained for layer 1 and matching boundary condition at the interface between layer 1 and layer 2 (i.e., $z = -d_1$) provides:

$$\frac{I_{u_1}}{C_1}(z = -d_1) = A_1 e^{\kappa_{a_1} sec\theta_1 d_1} + T_1 \tag{44}$$

$$\frac{I_{u_1}}{C_1}(z = -d_1) = r_1 \frac{I_{d_1}}{C_1}(z = -d_1) + (1 - r_1)\frac{I_{u_2}}{C_2}(z = -d_1) \tag{45}$$

Therefore,

$$A_1 e^{\kappa_{a_1} sec\theta_1 d_1} + T_1 = r_1 \frac{I_{d_1}}{C_1}(z = -d_1) + (1 - r_1)\frac{I_{u_2}}{C_2}(z = -d_1) \quad (46)$$

$$A_1 = -T_1 e^{-\kappa_{a_1} sec\theta_1 d_1} + e^{-\kappa_{a_1} sec\theta_1 d_1} r_1 \frac{I_{d_1}}{C_1}(z = -d_1) + e^{-\kappa_{a_1} sec\theta_1 d_1}(1 - r_1)\frac{I_{u_2}}{C_2}(z = -d_1) \quad (47)$$

As our objective term is $\frac{I_{u_1}}{C_1}(z = -d_0)$ with its integration equation:

$$A_1 + T_1 = \frac{I_{u_1}}{C_1}(z = -d_0) = A_1 e^{\kappa_{a_1} sec\theta_1 d_0} + T_1 \quad (48)$$

Equation (48) can be re-organized by replacing A_1 by Equation (47)

$$\frac{I_{u_1}}{C_1}(z = -d_0) = T_1(1 - e^{-\kappa_{a_1} sec\theta_1 (d_1 - d_0)}) + e^{-\kappa_{a_1} sec\theta_1 (d_1 - d_0)} r_1 \frac{I_{d_1}}{C_1}(z = -d_1) + e^{-\kappa_{a_1} sec\theta_1 (d_1 - d_0)}(1 - r_1)\frac{I_{u_2}}{C_2}(z = -d_1) \quad (49)$$

Recall $T_B = (1 - r_0)\frac{I_{u_1}}{C_1}(z = -d_0)$ where $\frac{I_{u_1}}{C_1}(z = -d_0)$ represents that the aggregated upward radiation from all the layers below the bottom boundary $(z = -d_0)$ of layer 0 (i.e., air). Likewise, $\frac{I_{u_2}}{C_2}(z = -d_1)$ in Equation (49) reflects the total upward radiation from all the layers below the bottom boundary $(z = -d_1)$ of layer 1. Hence, the first two terms on the right-hand side of Equation (49) can be considered as the partial total brightness temperature contributed by layer 1 alone.

For those two terms related to layer 1, $\frac{I_{u_1}}{C_1}(z = -d_0)$ can be adjusted by the immediate constant B_1. At the top boundary $(z = -d_0)$ and bottom boundary $(z = -d_1)$ of layer 1, the integration of downward intensity returns:

$$\frac{I_{d_1}}{C_1}(z=-d_1) = B_1 e^{-\kappa_{a_1} sec\theta_1 d_1} + T_1 \tag{50}$$

$$\frac{I_{d_1}}{C_1}(z=-d_0) = B_1 e^{-\kappa_{a_1} sec\theta_1 d_0} + T_1 \tag{51}$$

$$B_1 = e^{\kappa_{a_1} sec\theta_1 d_0}\frac{I_{d_1}}{C_1}(z=-d_0) - e^{\kappa_{a_1} sec\theta_1 d_0} T_1 \tag{52}$$

Put the above expression of B_1 into Equation (50)

$$\frac{I_{d_1}}{C_1}(z=-d_1) = T_1(1-e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}) + \frac{I_{d_1}}{C_1}(z=-d_0)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} \tag{53}$$

Substitute $\frac{I_{u_1}}{C_1}(z=-d_1)$ in Equation (49) by Equation (53), Equation (49) can be updated in a form of:

$$\frac{I_{u_1}}{C_1}(z=-d_0) = T_{B_{L_1}} + (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}\frac{I_{u_2}}{C_2}(z=-d_1) \tag{54}$$

where

$$\begin{aligned} T_{B_{L_1}} &= T_1(1-e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}) + r_1 e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}[T_1(1-e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}) + e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}\frac{I_{d_1}}{C_1}(z=-d_0)] \\ &= T_1(1-e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}) + r_1 e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_1(1-e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}) + r_1 e^{-2\kappa_{a_1} sec\theta_1 (d_1-d_0)}\frac{I_{d_1}}{C_1}(z=-d_0) \end{aligned}$$

$T_{B_{L_1}}$ can be considered as the part of the total upward emission below the top boundary of layer 1 $(z=-d_0)$, which is only contributed by the layer 1.

Following a similar process, the derivation of $\frac{I_{u_2}}{C_2}(z=-d_1)$ is dependent on the layer constant A_2. A_2 can be expressed by the transmitted upward radiation from layer 3 and the reflected downward emission from layer 2 at the bottom boundary of layer 2 $(z=-d_2)$ because of the continuous boundary conditions.

$$\frac{I_{u_2}}{C_2}(z=-d_2) = A_2 e^{\kappa_{a_2} sec\theta_2 d_2} + T_2 \tag{55}$$

$$\frac{I_{u_2}}{C_2}(z=-d_2) = r_2 \frac{I_{d_2}}{C_2}(z=-d_2) + (1-r_2)\frac{I_{u_3}}{C_3}(z=-d_2) \tag{56}$$

Therefore

$$A_2 e^{\kappa_{a_2} sec\theta_2 d_2} + T_2 = r_2 \frac{I_{d_2}}{C_2}(z=-d_2) + (1-r_2)\frac{I_{u_3}}{C_3}(z=-d_2) \tag{57}$$

$$A_2 = -T_2 e^{-\kappa_{a_2} sec\theta_2 d_2} + r_2 e^{-\kappa_{a_2} sec\theta_2 d_2}\frac{I_{d_2}}{C_2}(z=-d_2) + (1-r_2)e^{-\kappa_{a_2} sec\theta_2 d_2}\frac{I_{u_3}}{C_3}(z=-d_2) \tag{58}$$

As our objective term is $\frac{I_{u_2}}{C_2}(z=-d_1)$ with its integration equation:

$$\frac{I_{u_2}}{C_2}(z=-d_1) = A_2 e^{\kappa_{a_2} sec\theta_2 d_1} + T_2 \tag{59}$$

Equation (59) can be re-organized by replacing A_2 by Equation (58)

$$\begin{aligned}\frac{I_{u_2}}{C_2}(z=-d_1) = {} & T_2(1-e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}) + e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} r_2 \frac{I_{d_2}}{C_2}(z=-d_2) \\ & + e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}(1-r_2)\frac{I_{u_3}}{C_3}(z=-d_2)\end{aligned} \tag{60}$$

Again, given that $\frac{I_{u_3}}{C_3}(z=-d_2)$ represents all the upward radiation from the layers below the bottom boundary of layer 2 $(z=-d_2)$, the first two terms on right-hand side of Equation (60) could be attributed to layer 2 alone, and $\frac{I_{d_2}}{C_2}(z=-d_2)$ can be adjusted by the immediate constant B_2. At the top boundary $(z=-d_1)$ and bottom boundary $(z=-d_2)$ of layer 2, the integration of downward intensity returns:

$$\frac{I_{d_2}}{C_2}(z=-d_2) = B_2 e^{-\kappa_{a_2} sec\theta_2 d_2} + T_2 \tag{61}$$

$$\frac{I_{d_2}}{C_2}(z=-d_1) = B_2 e^{-\kappa_{a_2} sec\theta_2 d_1} + T_2 \tag{62}$$

$$B_2 = e^{\kappa_{a_2} sec\theta_2 d_1} \frac{I_{d_2}}{C_2}(z=-d_1) - e^{\kappa_{a_2} sec\theta_2 d_1} T_2 \tag{63}$$

Put the above expression of B_2 into Equation (61)

$$\frac{I_{d_2}}{C_2}(z=-d_2) = T_2(1-e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}) + \frac{I_{d_2}}{C_2}(z=-d_1) e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} \tag{64}$$

Substitute $\frac{I_{d_2}}{C_2}(z=-d_2)$ in Equation (60) by Equation (64), Equation (60) can be updated in a form of:

$$\frac{I_{u_2}}{C_2}(z=-d_1) = T_{B_{L_2}} + (1-r_2) e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} \frac{I_{u_3}}{C_3}(z=-d_2) \tag{65}$$

where

$$\begin{aligned} T_{B_{L_2}} &= T_2(1-e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}) + r_2 e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} [T_2(1-e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}) + e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} \frac{I_{d_2}}{C_2}(z=-d_1)] \\ &= T_2(1-e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}) + r_2 e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} T_2(1-e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}) + r_2 e^{-2\kappa_{a_2} sec\theta_2 (d_2-d_1)} \frac{I_{d_2}}{C_2}(z=-d_1) \end{aligned}$$

$T_{B_{L_2}}$ can be considered as the part of the total upward emission below the top boundary of layer 2 $(z=-d_1)$, which is only contributed by the layer 2. In light of these, the total upward radiation generated from any sub-layer n alone below its top boundary can be illustrated by:

$$\begin{aligned} T_{B_{L_n}} &= T_n(1-e^{-\kappa_{a_n} sec\theta_n (d_n-d_{n-1})}) + r_n e^{-\kappa_{a_n} sec\theta_n (d_n-d_{n-1})} T_n(1-e^{-\kappa_{a_n} sec\theta_n (d_n-d_{n-1})}) \\ &\quad + r_n e^{-2\kappa_{a_n} sec\theta_n (d_n-d_{n-1})} \frac{I_{d_n}}{C_n}(z=-d_{n-1}) \end{aligned} \tag{66}$$

The form of Equation (66) is expected to be always valid because the boundary equations and those integration equations of specific intensity are all similar and consistent for all the sub-layers. However, Burke et al. (1979) indicates that

$$T_{B_{L_n}} = T_n(1 - e^{-\kappa_{a_n} sec\theta_n (d_n - d_{n-1})}) + r_n e^{-\kappa_{a_n} sec\theta_n (d_n - d_{n-1})} T_n(1 - e^{-\kappa_{a_n} sec\theta_n (d_n - d_{n-1})}) \quad (67)$$

For both Equation (66) and (67), the first term on the right side of equation is upwelling radiation emitted by the layer n itself. The second term on the right side of equation is the upward radiation reflected by the downward radiation originated from the layer n. The unique third term on the right-hand side of Equation (66) represents the upward reflection of the downward radiations at the top boundary of layer n. Such downward radiations could be composed of the transmitted downward radiation from the upper layer (n-1) and the reflected upward radiation from the layer n, and firstly depart from the top boundary ($z = -d_{n-1}$), reflected at the boundary ($z = -d_{n-1}$) and move upward back to the top boundary of layer n. Unlike the first two terms, the third term could therefore be considered as a second-order process, and likely have a small magnitude. Nevertheless, it may be not suitable to completely ignore it in terms of continuity in the derivation processes.

In addition, for a general N-layer system, the individual upward radiation from the last layer is

$$\begin{aligned} T_{B_{L_N}} = {} & T_N(1 - e^{-\kappa_{a_N} sec\theta_N (d_N - d_{N-1})}) + r_N e^{-\kappa_{a_N} sec\theta_N (d_N - d_{N-1})} T_N(1 - e^{-\kappa_{a_N} sec\theta_N (d_N - d_{N-1})}) \\ & + r_N e^{-2\kappa_{a_N} sec\theta_N (d_N - d_{N-1})} \frac{I_{d_N}}{C_N}(z = -d_{N-1}) \end{aligned} \quad (68)$$

Since $d_N = -\infty$ and $e^{-\infty} \approx 0$, the above equation can be further simplified as

$$T_{B_{L_N}} = T_N \quad (69)$$

Furthermore, it is known that

$$\frac{I_{d_n}}{C_n}(z=-d_{n-1}) = B_n e^{-\kappa_{a_n} sec\theta_n d_{n-1}} + T_n \tag{70}$$

For simplicity, Equation (66) can be expressed as

$$\begin{aligned} T_{B_{L_n}} &= T_n(1-e^{-\kappa_{a_n} sec\theta_n (d_n - d_{n-1})})(1+r_n e^{-\kappa_{a_n} sec\theta_n (d_n - d_{n-1})}) \\ &+ r_n e^{-2\kappa_{a_n} sec\theta_n (d_n - d_{n-1})}(B_n e^{-\kappa_{a_n} sec\theta_n d_{n-1}} + T_n) \end{aligned} \tag{71}$$

Therefore, in a 3-layer stratified medium, we can obtain that

$$\begin{aligned} T_{B_{L_1}} &= T_1(1-e^{-\kappa_{a_1} sec\theta_1 (d_1 - d_0)})(1+r_1 e^{-\kappa_{a_1} sec\theta_1 (d_1 - d_0)}) \\ &+ r_1 e^{-2\kappa_{a_1} sec\theta_1 (d_1 - d_0)}(B_1 e^{-\kappa_{a_1} sec\theta_1 d_0} + T_1) \end{aligned} \tag{72}$$

$$\begin{aligned} T_{B_{L_2}} &= T_2(1-e^{-\kappa_{a_2} sec\theta_2 (d_2 - d_1)})(1+r_2 e^{-\kappa_{a_2} sec\theta_2 (d_2 - d_1)}) \\ &+ r_2 e^{-2\kappa_{a_2} sec\theta_2 (d_2 - d_1)}(B_2 e^{-\kappa_{a_2} sec\theta_2 d_1} + T_2) \end{aligned} \tag{73}$$

$$\begin{aligned} T_{B_{L_3}} &= T_3(1-e^{-\kappa_{a_3} sec\theta_3 (d_3 - d_2)})(1+r_3 e^{-\kappa_{a_3} sec\theta_3 (d_3 - d_2)}) \\ &+ r_3 e^{-2\kappa_{a_3} sec\theta_3 (d_3 - d_2)}(B_3 e^{-\kappa_{a_3} sec\theta_3 d_2} + T_3) \\ &= T_3 \end{aligned} \tag{74}$$

For the layer 3 ($d_3 = -\infty$), based on Equation (54) and (65), we can derive that

$$\begin{aligned} \frac{I_{u_3}}{C_3}(z=-d_2) &= T_{B_{L_3}} + (1-r_3)e^{-\kappa_{a_3} sec\theta_2 (d_3 - d_2)} \frac{I_{u_4}}{C_4}(z=-d_3) \\ &= T_{B_{L_3}} \\ &= T_3 \end{aligned} \tag{75}$$

Recall the total brightness temperature of a 3-layer medium,

$$
\begin{aligned}
T_B &= (1-r_0)\frac{I_{u_1}}{C_1}(z=-d_0) \\
&= (1-r_0)(T_{B_{L_1}} + (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}\frac{I_{u_2}}{C_2}(z=-d_1)) \\
&= (1-r_0)T_{B_{L_1}} + (1-r_0)(1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}(T_{B_{L_2}} + (1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}\frac{I_{u_3}}{C_3}(z=-d_2)) \\
&= (1-r_0)T_{B_{L_1}} + (1-r_0)(1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}T_{B_{L_2}} \\
&+ (1-r_0)(1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}T_3
\end{aligned} \tag{76}
$$

Given the above derivation process, even in a N-layer system, the contributions from the individual layers to the total brightness temperature can be achieved by continuously partitioning the last term in the equation using the integration equations with parameters A_n and B_n and matching boundary equations until the last layer, along with the accurate products of the transmission and absorption coefficients.

If we placed the $T_{B_{L_1}}$, and $T_{B_{L_2}}$ by Equation (72) and (73), we will get

$$
\begin{aligned}
T_B &= (1-r_0)[T_1(1-e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)})(1+r_1 e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}) + r_1 e^{-2\kappa_{a_1} sec\theta_1 (d_1-d_0)}(B_1 e^{-\kappa_{a_1} sec\theta_1 d_0} + T_1)] \\
&+ (1-r_0)(1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}[T_2(1-e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)})(1+r_2 e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}) \\
&+ r_2 e^{-2\kappa_{a_2} sec\theta_2 (d_2-d_1)}(B_2 e^{-\kappa_{a_2} sec\theta_2 d_1} + T_2)] \\
&+ (1-r_0)(1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}T_3
\end{aligned} \tag{77}
$$

Similarly, using the explicit terms as shown in Equation (71), the total brightness temperature of an N-layer stratified medium can be generalized as:

$$
\begin{aligned}
T_B &= \sum_{i=1}^{N}[T_i(1-e^{-\kappa_{a_i} sec\theta_i (d_i-d_{i-1})})(1+r_i e^{-\kappa_{a_i} sec\theta_i (d_i-d_{i-1})}) \\
&+ r_i e^{-2\kappa_{a_i} sec\theta_i (d_i-d_{i-1})}(B_i e^{-\kappa_{a_i} sec\theta_i d_{i-1}} + T_i)] \\
&\cdot \prod_{j=1}^{i}(1-r_{j-1})e^{-[\sum_{j=2}^{\forall i}\kappa_{a_{j-1}} \sec\theta_{j-1}(d_{j-1}-d_{j-2})]}
\end{aligned} \tag{78}
$$

where $\forall i$ denotes $i \geq j \geq 2$, indicating that the exponential term will only be calculated when $i \geq j \geq 2$

A2.2.2. Derivation of the Relationship between Total Brightness Temperature and Layered Physical Temperature in a 3-Layer Medium

Additionally, the total brightness temperature would also be expressed by the physical temperatures of sub-layers in a form of $T_B = S_1T_1 + S_2T_2 + S_3T_3 + \cdots + S_NT_N$. Instead of directly using the immediate coefficients (such as A_N and B_N), this can be achieved by explicitly describing the A_1 term with the information from each layer as well as inversely recurrence procedures.

For a 3-layer medium, recall Equations (18) – (21), a group of equivalent Equations show below:

$$r_0A_1 - B_1 = (1 - r_0)T_1 \tag{79}$$

$$e^{\kappa_{a_1} sec\theta_1 d_1} A_1 - r_1 e^{-\kappa_{a_1} sec\theta_1 d_1} B_1 + (r_1 - 1)e^{\kappa_{a_2} sec\theta_2 d_1} A_2 = (1 - r_1)(T_2 - T_1) \tag{80}$$

$$(1 - r_1)e^{-\kappa_{a_1} sec\theta_1 d_1} B_1 + r_1 e^{\kappa_{a_2} sec\theta_2 d_1} A_2 - e^{-\kappa_{a_2} sec\theta_2 d_1} B_2 = (1 - r_1)(T_2 - T_1) \tag{81}$$

$$e^{\kappa_{a_2} sec\theta_2 d_2} A_2 - r_2 e^{-\kappa_{a_2} sec\theta_2 d_2} B_2 = (1 - r_2)(T_3 - T_2) \tag{82}$$

Based on Equation (79),

$$B_1 = r_0A_1 - (1 - r_0)T_1 \tag{83}$$

Then we replaced B_1 term in Equation (80) by Equation (83)

$$\begin{aligned} A_2 = {} & \frac{1}{(1 - r_1)} \frac{(1 - r_0 r_1 e^{-2\kappa_{a_1} sec\theta_1 d_1})}{e^{-\kappa_{a_1} sec\theta_1 d_1}} e^{-\kappa_{a_2} sec\theta_2 d_1} A_1 \\ & + \frac{r_1(1 - r_0)}{(1 - r_1)} e^{-\kappa_{a_1} sec\theta_1 d_1} e^{-\kappa_{a_2} sec\theta_2 d_1} T_1 \\ & + e^{-\kappa_{a_2} sec\theta_2 d_1}(T_1 - T_2) \end{aligned} \tag{84}$$

Similarly, write B_2 in terms of A_1 from Equations (83) and (84)

$$
\begin{aligned}
B_2 = {} & r_0(1-r_1)e^{-\kappa_{a_1} sec\theta_1 d_1} e^{\kappa_{a_2} sec\theta_2 d_1} A_1 \\
& + \frac{r_1}{(1-r_1)} \frac{(1-r_0 r_1 e^{-2\kappa_{a_1} sec\theta_1 d_1})}{e^{-\kappa_{a_1} sec\theta_1 d_1}} e^{\kappa_{a_2} sec\theta_2 d_1} A_1 \\
& - (1-r_0)(1-r_1)e^{-\kappa_{a_1} sec\theta_1 d_1} e^{\kappa_{a_2} sec\theta_2 d_1} T_1 \\
& + \frac{r_1^2(1-r_0)}{(1-r_1)} e^{-\kappa_{a_1} sec\theta_1 d_1} e^{\kappa_{a_2} sec\theta_2 d_1} T_1 \\
& + e^{\kappa_{a_2} sec\theta_2 d_1}(T_1 - T_2)
\end{aligned} \tag{85}
$$

Using Equations (84) and (85), A_1 can be obtained via solving Equation (82)

$$
\begin{aligned}
A_1 = {} & \{[-r_1(1-r_0)g_{10} - g_{10}g_{20}(1-r_0)r_2(1-r_1)^2 - g_{11}(1-r_2(1-r_1)g_{20})]T_1 \\
& + [g_{11}(1-r_2(1-r_1)g_{20}) - g_{11}g_{21}]T_2 + g_{11}g_{21}T_3\}/Q_0
\end{aligned} \tag{86}
$$

where

$$
\begin{aligned}
g_{10} &= \frac{1}{e^{2\kappa_{a_1} sec\theta_1 (d_1 - d_0)} - r_0 r_1} \\
g_{11} &= g_{10} e^{\kappa_{a_1} sec\theta_1 (d_1 - d_0)} (1 - r_1) \\
g_{20} &= \frac{1}{e^{2\kappa_{a_2} sec\theta_2 (d_2 - d_1)} - r_1 r_2} \\
g_{21} &= g_{20} e^{\kappa_{a_2} sec\theta_2 (d_2 - d_1)} (1 - r_2) \\
Q_0 &= 1 - r_0 r_2 (1-r_1)^2 g_{10} g_{20}
\end{aligned}
$$

If we assumed that

$$
Q_1 = -\frac{r_1(1-r_0)g_{10}}{Q_0} - \frac{g_{10}g_{20}(1-r_0)r_2(1-r_1)^2}{Q_0} - \frac{g_{11}(1-r_2(1-r_1)g_{20})}{Q_0} \tag{87}
$$

$$
Q_2 = \frac{g_{11}(1-r_2(1-r_1)g_{20})}{Q_0} - \frac{g_{11}g_{21}}{Q_0} \tag{88}
$$

$$
Q_3 = \frac{g_{11}g_{21}}{Q_0} \tag{89}
$$

$$
A_1 = Q_1 T_1 + Q_2 T_2 + Q_3 T_3 \tag{90}
$$

Recall Equation (29): $T_B = (1 - r_0)(A_1 + T_1)$

$$
T_B = (1-r_0)[(1+Q_1)T_1 + Q_2 T_2 + Q_3 T_3] \tag{91}
$$

A2.2.3. Discrepancy between the above Two Methods

Section 5.2.2.1 and Section 5.2.2.2 introduce two fashions for the separation of the total brightness temperature and thereby linking it with information from each sub-layer. The investigation on the discrepancies between the above methods has been conducted here. The 3-layer stratified medium will continuously use as the illustrative example. For the first method, the total brightness temperature is

$$T_B = T_{B_1} + T_{B_2} + T_{B_3} \tag{92}$$

where

$$\begin{aligned}
T_{B_1} &= (1-r_0)[T_1(1-e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)})(1+r_1 e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}) \\
&\quad + r_1 e^{-2\kappa_{a_1} sec\theta_1 (d_1-d_0)}(B_1 e^{-\kappa_{a_1} sec\theta_1 d_0} + T_1)] \\
T_{B_2} &= (1-r_0)(1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}[T_2(1-e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)})(1+r_2 e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)}) \\
&\quad + r_2 e^{-2\kappa_{a_2} sec\theta_2 (d_2-d_1)}(B_2 e^{-\kappa_{a_2} sec\theta_2 d_1} + T_2)] \\
T_{B_3} &= (1-r_0)(1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} T_3
\end{aligned}$$

In terms of the second approach, the total brightness temperature is expressed by

$$T_B = S_1 T_1 + S_2 T_2 + S_3 T_3 \tag{93}$$

where

$$\begin{aligned}
S_1 &= (1-r_0)(1+Q_1) \\
S_2 &= (1-r_0)Q_2 \\
S_3 &= (1-r_0)Q_3
\end{aligned}$$

The difference between the above two expressions can be studied by representing B_1 and B_2 terms by the explicitly layered coefficients and temperatures (i.e., Equations (83) and (85)).

If we assumed that

$$x_1 = (1-e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)})(1+r_1 e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}) \tag{94}$$

$$y_1 = r_0 r_1 e^{-2\kappa_{a_1} sec\theta_1 (d_1-d_0)} \tag{95}$$

Recall Equation (92), now T_{B_1} can be expressed as:

$$\begin{aligned} T_{B_1} &= (1-r_0)[x_1T_1 + r_1e^{-2\kappa_{a_1}sec\theta_1(d_1-d_0)}(B_1e^{-\kappa_{a_1}sec\theta_1 d_0} + T_1)] \\ &= (1-r_0)[x_1T_1 + r_1e^{-2\kappa_{a_1}sec\theta_1(d_1-d_0)}(B_1 + T_1)] \end{aligned} \tag{96}$$

The expression of B_1 in terms of A_1 has been provided in Equation (83), therefore, $B_1 + T_1$ is equal to:

$$\begin{aligned} B_1 + T_1 &= r_0A_1 - (1-r_0)T_1 + T_1 \\ &= r_0(A_1 + T_1) \\ &= r_0[(1+Q_1)T_1 + Q_2T_2 + Q_3T_3] \end{aligned} \tag{97}$$

Equation (96) can be further presented as:

$$\begin{aligned} T_{B_1} &= (1-r_0)\{x_1T_1 + r_1e^{-2\kappa_{a_1}sec\theta_1(d_1-d_0)}r_0[(1+Q_1)T_1 + Q_2T_2 + Q_3T_3]\} \\ &= (1-r_0)\{[x_1 + y_1(1+Q_1)]T_1 + y_1Q_2T_2 + y_1Q_3T_3\} \end{aligned} \tag{98}$$

Following a similar process conducted from Equation (94) to Equation (98), we can firstly obtain:

$$x_2 = (1 - e^{-\kappa_{a_2}sec\theta_2(d_2-d_1)})(1 + r_2e^{-\kappa_{a_2}sec\theta_2(d_2-d_1)}) \tag{99}$$

$$y_2 = r_1r_2e^{-2\kappa_{a_2}sec\theta_2(d_2-d_1)} \tag{100}$$

$$T_{B_2} = (1-r_0)(1-r_1)e^{-\kappa_{a_1}sec\theta_1(d_1-d_0)}[x_2T_2 + r_2e^{-2\kappa_{a_2}sec\theta_2(d_2-d_1)}(B_2e^{-\kappa_{a_2}sec\theta_2 d_1} + T_2)] \tag{101}$$

Recall Equation (82), B_2 can be expressed in term of A_2.

$$B_2 = \frac{e^{2\kappa_{a_2}sec\theta_2 d_2}}{r_2}A_2 + \frac{(1-r_2)e^{\kappa_{a_2}sec\theta_2 d_2}}{r_2}T_2 - \frac{(1-r_2)e^{\kappa_{a_2}sec\theta_2 d_2}}{r_2}T_3 \tag{102}$$

Put Equation (82) where A_2 was represented in terms of A_1 into the Equation (102), and multiply with $e^{-\kappa_{a_2}\sec\theta_2 d_1}$

$$B_2 e^{-\kappa_{a_2} sec\theta_2 d_1} + T_2 = \frac{(1-y_1)e^{\kappa_{a_1} sec\theta_1 d_1} e^{2\kappa_{a_2} sec\theta_2 (d_2-d_1)}}{(1-r_1)r_2} A_1 + \frac{(1-r_0)r_1 e^{-\kappa_{a_1} sec\theta_1 d_1} e^{2\kappa_{a_2} sec\theta_2 (d_2-d_1)}}{(1-r_1)r_2} T_1 + \frac{e^{2\kappa_{a_2} sec\theta_2 (d_2-d_1)}}{r_2}(T_1 - T_2) + \frac{(1-r_2)e^{\kappa_{a_2} sec\theta_2 (d_2-d_1)}}{r_2}(T_2 - T_3) + T_2 \quad (103)$$

As a result,

$$\begin{aligned}\frac{T_{B_2}}{(1-r_0)} &= (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} x_2 T_2 + (1-y_1)A_1 + (1-r_0)r_1 e^{-2\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_1 \\ &+ (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_1 - (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_2 \\ &+ (1-r_1)(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_2 \\ &- (1-r_1)(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_3 \\ &+ (1-r_1)r_2 e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} e^{-2\kappa_{a_2} sec\theta_2 (d_2-d_1)} T_2 \end{aligned} \quad (104)$$

When we substitute A_1 by $Q_1T_1 + Q_2T_2 + Q_3T_3$. Equation (104) can be simplified (here, we separate the derivations for coefficients before T_1, T_2, and T_3 as the background references).

(1) T_1's coefficient:

$$\begin{aligned}&(1-y_1)Q_1T_1 + (1-r_0)r_1 e^{-2\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_1 + (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_1 \\ =&[(1-y_1)Q_1 + r_1 e^{-2\kappa_{a_1} sec\theta_1 (d_1-d_0)} + (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} - y_1]T_1 \\ =&[(1-y_1)Q_1 + (1-y_1) - (1-r_1 e^{-2\kappa_{a_1} sec\theta_1 (d_1-d_0)} - (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)})]T_1 \\ =&[(1-y_1)(1+Q_1) - x_1]T_1\end{aligned}$$

(2) T_2's coefficient:

$$\begin{aligned}&(1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} x_2 T_2 + (1-y_1)Q_2T_2 - (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_2 \\ &+(1-r_1)(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_2 + (1-r_1)r_2 e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} e^{-2\kappa_{a_2} sec\theta_2 (d_2-d_1)} T_2 \\ =&[(1-y_1)Q_2 - (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} + (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} + (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} x_2 \\ &+(1-r_1)(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} + (1-r_1)r_2 e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} e^{-2\kappa_{a_2} sec\theta_2 (d_2-d_1)}]T_2 \\ =&\{(1-y_1)Q_2 - (1-r_1)e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}[1-(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} - r_2 e^{-2\kappa_{a_2} sec\theta_2 (d_2-d_1)} - x_2]\}T_2 \\ =&(1-y_1)Q_2T_2\end{aligned}$$

(3) T_3's coefficient:

$$\begin{aligned}&(1-y_1)Q_3T_3 - (1-r_1)(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)} T_3 \\ &= [(1-y_1)Q_3 - (1-r_1)(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}]T_3\end{aligned}$$

Based on the above Equations

$$T_{B_2} = (1-r_0)\{[(1-y_1)(1+Q_1)-x_1]T_1 + (1-y_1)Q_2T_2 + [(1-y_1)Q_3 - (1-r_1)(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}]T_3\} \quad (105)$$

To summary, contributions from sub-layers to the total brightness temperature derived from the approach proposed in Section 5.2.2.1 can be expressed as

$$T_{B_1} = F_1T_1 + F_2T_2 + F_3T_3 \quad (106)$$

where

$$F_1 = (1-r_0)[y_1(1+Q_1)+x_1]$$
$$F_2 = (1-r_0)y_1Q_2$$
$$F_3 = (1-r_0)y_1Q_3$$

$$T_{B_2} = G_1T_1 + G_2T_2 + G_3T_3 \quad (107)$$

where

$$G_1 = (1-r_0)[(1-y_1)(1+Q_1)-x_1]$$
$$G_2 = (1-r_0)(1-y_1)Q_2$$
$$G_3 = (1-r_0)[(1-y_1)Q_3 - (1-r_1)(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}]$$

$$T_{B_3} = M_1T_1 + M_2T_2 + M_3T_3 \quad (108)$$

where

$$M_1 = 0;$$
$$M_2 = 0;$$
$$M_3 = (1-r_0)(1-r_1)(1-r_2)e^{-\kappa_{a_2} sec\theta_2 (d_2-d_1)} e^{-\kappa_{a_1} sec\theta_1 (d_1-d_0)}$$

According to Equation (93), the coefficients of the above two approaches having the following relationship:

$$S_1 = F_1 + G_1 + M_1 \quad (109)$$
$$S_2 = F_2 + G_2 + M_2 \quad (110)$$
$$S_3 = F_3 + G_3 + M_3 \quad (111)$$

Given the identical total brightness temperature on which the above two methods are constructed, such a relationship is expected to exist for any N-layer stratified medium ($N > 1$) in forms of the above two fashions.

Reference

Akbar, R., Short Gianotti, D., McColl, K.A., Haghighi, E., Salvucci, G.D., & Entekhabi, D. (2018). Hydrological storage length scales represented by remote sensing estimates of soil moisture and precipitation. *Water Resources Research, 54*, 1476-1492

Al-Yaari, A., Wigneron, J.-P., Dorigo, W., Colliander, A., Pellarin, T., Hahn, S., Mialon, A., Richaume, P., Fernandez-Moran, R., & Fan, L. (2019). Assessment and inter-comparison of recently developed/reprocessed microwave satellite soil moisture products using ISMN ground-based measurements. *Remote sensing of environment, 224*, 289-303

Al-Yaari, A., Wigneron, J.-P., Ducharne, A., Kerr, Y., De Rosnay, P., De Jeu, R., Govind, A., Al Bitar, A., Albergel, C., & Munoz-Sabater, J. (2014). Global-scale evaluation of two satellite-based passive microwave soil moisture datasets (SMOS and AMSR-E) with respect to Land Data Assimilation System estimates. *Remote sensing of environment, 149*, 181-195

Al-Yaari, A., Wigneron, J.-P., Kerr, Y., Rodriguez-Fernandez, N., O'Neill, P., Jackson, T., De Lannoy, G.J.M., Al Bitar, A., Mialon, A., & Richaume, P. (2017). Evaluating soil moisture retrievals from ESA's SMOS and NASA's SMAP brightness temperature datasets. *Remote sensing of environment, 193*, 257-273

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., & Martin, E. (2008). From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations

Babaeian, E., Sadeghi, M., Jones, S.B., Montzka, C., Vereecken, H., & Tuller, M. (2019). Ground, proximal, and satellite remote sensing of soil moisture. *Reviews of Geophysics, 57*, 530-616

Baldwin, D., Manfreda, S., Keller, K., & Smithwick, E. (2017). Predicting root zone soil moisture with soil properties and satellite near-surface moisture data across the conterminous United States. *Journal of Hydrology, 546*, 393-404

Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., Van Dijk, A.I., Huffman, G.J., Adler, R.F., & Wood, E.F. (2019a). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences, 23*, 207-224

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., Van Dijk, A.I., McVicar, T.R., & Adler, R.F. (2019b). MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment. *Bulletin of the American meteorological Society, 100*, 473-500

Bell, J.E., Palecki, M.A., Baker, C.B., Collins, W.G., Lawrimore, J.H., Leeper, R.D., Hall, M.E., Kochendorfer, J., Meyers, T.P., & Wilson, T. (2013). US Climate Reference Network soil moisture and temperature observations. *Journal of Hydrometeorology, 14*, 977-988

Bertoldi, G., Della Chiesa, S., Notarnicola, C., Pasolli, L., Niedrist, G., & Tappeiner, U. (2014). Estimation of soil moisture patterns in mountain grasslands by means of SAR RADARSAT2 images and hydrological modeling. *Journal of Hydrology, 516*, 245-257

Bircher, S., Andreasen, M., Vuollet, J., Vehviläinen, J., Rautiainen, K., Jonard, F., Weihermüller, L., Zakharova, E., Wigneron, J.-P., & Kerr, Y.H. (2016). Soil moisture sensor calibration for organic soil surface layers. *Geoscientific instrumentation, methods and data systems, 5*, 109-125

Bircher, S., Skou, N., Jensen, K.H., Walker, J.P., Rasmussen, L., & Verhoest, N. (2012). A soil moisture and temperature network for SMOS validation in Western Denmark. *Hydrology & Earth System Sciences, 16*

Brakhasi, F., Walker, J.P., Ye, N., Wu, X., Shen, X., Yeo, I.-Y., Boopathi, N., Kim, E., Kerr, Y., & Jackson, T. (2023). Towards soil moisture profile estimation in the root zone using L-and P-band radiometer observations: A coherent modelling approach. *Science of Remote Sensing*, 100079

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., & Levizzani, V. (2014). Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. *Journal of Geophysical Research: Atmospheres, 119*, 5128-5141

Brocca, L., Filippucci, P., Hahn, S., Ciabatta, L., Massari, C., Camici, S., Schüller, L., Bojkov, B., & Wagner, W. (2019). SM2RAIN–ASCAT (2007–2018): global daily satellite rainfall data from ASCAT soil moisture observations. *Earth System Science Data, 11*, 1583-1601

Brocca, L., Melone, F., Moramarco, T., & Morbidelli, R. (2009a). Soil moisture temporal stability over experimental areas in Central Italy. *Geoderma, 148*, 364-374

Brocca, L., Melone, F., Moramarco, T., & Singh, V. (2009b). Assimilation of observed soil moisture data in storm rainfall-runoff modeling. *Journal of Hydrologic Engineering, 14*, 153-165

Brodzik, M.J., Billingsley, B., Haran, T., Raup, B., & Savoie, M.H. (2012). EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets. *ISPRS International Journal of Geo-Information, 1*, 32-45

Broll, G., Brauckmann, H.J., Overesch, M., Junge, B., Erber, C., Milbert, G., Baize, D., & Nachtergaele, F. (2006). Topsoil characterization—recommendations for revision and expansion of the FAO-Draft (1998) with emphasis on humus forms and biological features. *Journal of Plant Nutrition and Soil Science, 169*, 453-461

Burke, W., Schmugge, T., & Paris, J. (1979). Comparison of 2.8-and 21-cm microwave radiometer observations over soils with emission model calculations. *Journal of Geophysical Research: Oceans, 84*, 287-294

Caldwell, T.G., Bongiovanni, T., Cosh, M.H., Jackson, T.J., Colliander, A., Abolt, C.J., Casteel, R., Larson, T., Scanlon, B.R., & Young, M.H. (2019). The Texas soil observation network: A comprehensive soil moisture dataset for remote sensing and land surface model validation. *Vadose Zone Journal, 18*, 1-20

Carranza, C., Nolet, C., Pezij, M., & van der Ploeg, M. (2021). Root zone soil moisture estimation with Random Forest. *Journal of Hydrology, 593*, 125840

Carrea, L., Embury, O., & Merchant, C.J. (2015). Datasets related to in-land water for limnology and remote sensing applications: distance-to-land, distance-to-water, water-body identifier and lake-centre co-ordinates. *Geoscience data journal, 2*, 83-97

Carrea, L., & Merchant, C. (2019). GloboLakes: Lake Surface Water Temperature (LSWT) v4.0 (1995-2016). Centre for Environmental Data Analysis. In

Carrea, L., & Merchant, C. (2020a). Copernicus Global Land Operations Cryosphere and Water C-GLOPS2 Framework Service Contract N 199496 (JRC) ALGORIHM THEORETICAL BASIS DOCUMENT LAKE SURFACE WATER TEMPERATURE 1KM PRODUCTS. In

Carrea, L., & Merchant, C. (2020b). Copernicus Global Land Operations Cryosphere and Water C-GLOPS2 Framework Service Contract N 199496 (JRC) Product User Manual LAKE SURFACE WATER TEMPERATURE 1KM PRODUCTS. In

Carrea, L., & Merchant, C. (2020c). Copernicus Global Land Operations Cryosphere and Water C-GLOPS Framework Service Contract N 199496 (JRC) QUALITY ASSESSMENT REPORT LAKE SURFACE WATER TEMPERATURE 1KM PRODUCTS. In

Chan, S., Hunt, R., Bindlish, R., Njoku, E., Kimball, J., & Jackson, T. (2013). SMAP ancillary data report: vegetation water content. *Jet Propulsion Lab., California Inst. Technol., Pasadena, CA, USA, JPL D-53058*

Chan, S.K., Bindlish, R., O'Neill, P., Jackson, T., Njoku, E., Dunbar, S., Chaubell, J., Piepmeier, J., Yueh, S., & Entekhabi, D. (2018). Development and assessment of the SMAP enhanced passive soil moisture product. *Remote sensing of environment, 204*, 931-941

Chan, S.K., Bindlish, R., O'Neill, P.E., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., & Piepmeier, J. (2016). Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and Remote Sensing, 54*, 4994-5007

Chen, F., Crow, W.T., Bindlish, R., Colliander, A., Burgin, M.S., Asanuma, J., & Aida, K. (2018). Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. *Remote Sensing of Environment, 214*, 1-13

Clewley, D., Whitcomb, J.B., Akbar, R., Silva, A.R., Berg, A., Adams, J.R., Caldwell, T., Entekhabi, D., & Moghaddam, M. (2017). A method for upscaling in situ soil moisture measurements to satellite footprint scale using random forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*, 2663-2673

Colliander, A., Jackson, T.J., Bindlish, R., Chan, S., Das, N., Kim, S., Cosh, M., Dunbar, R., Dang, L., & Pashaian, L. (2017). Validation of SMAP surface soil moisture products with core validation sites. *Remote sensing of environment, 191*, 215-231

Colliander, A., Jackson, T.J., Chan, S., O'Neill, P., Bindlish, R., Cosh, M., Caldwell, T., Walker, J., Berg, A., & McNairn, H. (2018). An assessment of the differences between spatial resolution and grid size for the SMAP enhanced soil moisture product over homogeneous sites. *Remote sensing of environment, 207*, 65-70

Crosman, E.T., & Horel, J.D. (2009). MODIS-derived surface temperature of the Great Salt Lake. *Remote sensing of environment, 113*, 73-81

Crow, W.T., Berg, A.A., Cosh, M.H., Loew, A., Mohanty, B.P., Panciera, R., de Rosnay, P., Ryu, D., & Walker, J.P. (2012). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics, 50*

Crow, W.T., & Wood, E.F. (2002). The value of coarse-scale soil moisture observations for regional surface energy balance modeling. *Journal of Hydrometeorology, 3*, 467-482

Dai, A., Trenberth, K.E., & Qian, T. (2004). A global dataset of Palmer Drought Severity Index for 1870–2002: Relationship with soil moisture and effects of surface warming. *Journal of Hydrometeorology, 5*, 1117-1130

Das, N.N., & O'Neill, P. (2020). Soil Moisture Active Passive (SMAP) Ancillary Data Report, Soil Attributes, August 15, 2020, JPL D-53058, Version B, Jet Propulsion Laboratory, Pasadena, CA. In

De Jeu, R.A., Wagner, W., Holmes, T., Dolman, A., Van De Giesen, N., & Friesen, J. (2008). Global soil moisture patterns observed by space borne microwave radiometers and scatterometers. *Surveys in Geophysics, 29*, 399-420

De Jeu, r.d., & Owe, M. (2003). Further validation of a new methodology for surface moisture and vegetation optical depth retrieval. *International Journal of Remote Sensing, 24*, 4559-4578

De Lannoy, G.J., Koster, R.D., Reichle, R.H., Mahanama, S.P., & Liu, Q. (2014). An updated treatment of soil texture and associated hydraulic properties in a global land modeling system. *Journal of Advances in Modeling Earth Systems, 6*, 957-979

Dee, D.P., Uppala, S.M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., & Bauer, d.P. (2011). The ERA‐Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society, 137*, 553-597

Dirmeyer, P.A. (2000). Using a global soil wetness dataset to improve seasonal climate simulation. *Journal of Climate, 13*, 2900-2922

Dobson, M.C., Ulaby, F.T., Hallikainen, M.T., & El-Rayes, M.A. (1985). Microwave dielectric behavior of wet soil-Part II: Dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing, GE-23*, 35-46

Dong, J., Crow, W.T., Tobin, K.J., Cosh, M.H., Bosch, D.D., Starks, P.J., Seyfried, M., & Collins, C.H. (2020). Comparison of microwave remote sensing and land surface modeling for surface soil moisture climatology estimation. *Remote sensing of environment, 242*, 111756

Dorigo, W., Gruber, A., De Jeu, R., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., & Parinussa, R. (2015). Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote sensing of environment, 162*, 380-395

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W., Xaver, A., Annor, F., & Ardö, J. (2021). The International Soil Moisture Network: serving Earth system science for over a decade. *Hydrology and earth system sciences, 25*, 5749-5804

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., & Gruber, A. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote sensing of environment, 203*, 185-215

Dorigo, W., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., & van Oevelen, P. (2011a). The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences, 15*, 1675-1698

Dorigo, W., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiova, A., Sanchis-Dufau, A., Zamojski, D., Cordes, C., Wagner, W., & Drusch, M. (2013). Global automated quality control of in situ soil moisture data from the International Soil Moisture Network. *Vadose Zone Journal, 12*

Dorigo, W.A., Scipal, K., Parinussa, R.M., Liu, Y.Y., Wagner, W., De Jeu, R.A., & Naeimi, V. (2010). Error characterisation of global active and passive microwave soil moisture datasets. *Hydrology and Earth System Sciences, 14*, 2605

Dorigo, W.A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., & Jackson, T. (2011b). The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences, 15*, 1675-1698

Draper, C.S., Walker, J.P., Steinle, P.J., De Jeu, R.A., & Holmes, T.R. (2009). An evaluation of AMSR–E derived soil moisture over Australia. *Remote sensing of environment, 113*, 703-710

ElSaadani, M., Habib, E., Abdelhameed, A.M., & Bayoumi, M. (2021). Assessment of a spatiotemporal deep learning approach for soil moisture prediction and filling the gaps in between soil moisture observations. *Frontiers in artificial intelligence, 4*, 636234

Entekhabi, D., Asrar, G.R., Betts, A.K., Beven, K.J., Bras, R.L., Duffy, C.J., Dunne, T., Koster, R.D., Lettenmaier, D.P., & McLaughlin, D.B. (1999). An agenda for land surface hydrology research and a call for the second international hydrological decade. *Bulletin of the American meteorological Society, 80*, 2043-2058

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., & Johnson, J. (2010a). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE, 98*, 704-716

Entekhabi, D., Reichle, R.H., Koster, R.D., & Crow, W.T. (2010b). Performance metrics for soil moisture retrievals and application requirements. *Journal of Hydrometeorology, 11*, 832-840

Entekhabi, D., Rodriguez-Iturbe, I., & Castelli, F. (1996). Mutual interaction of soil moisture state and atmospheric processes. *Journal of Hydrology, 184*, 3-17

Entekhabi, D., Yueh, S., O'Neill, P., Kellogg, K., Allen, A., Bindlish, R., Brown, M., Chan, S., Colliander, A., & Crow, W. (2014). SMAP Handbook Soil Moisture Active Passive. *Mapping Soil Moisture Freeze/Thaw from Space, Pasadena, CA*

Fang, B., Kansara, P., Dandridge, C., & Lakshmi, V. (2021a). Drought monitoring using high spatial resolution soil moisture data over Australia in 2015–2019. *Journal of Hydrology, 594*, 125960

Fang, B., Lakshmi, V., Bindlish, R., & Jackson, T.J. (2018). AMSR2 soil moisture downscaling using temperature and vegetation data. *Remote Sensing, 10*, 1575

Fang, B., Lakshmi, V., Bindlish, R., Jackson, T.J., Cosh, M., & Basara, J. (2013). Passive microwave soil moisture downscaling using vegetation index and skin surface temperature. *Vadose Zone Journal, 12*, 1-19

Fang, B., Lakshmi, V., Bindlish, R., Jackson, T.J., & Liu, P.-W. (2020). Evaluation and validation of a high spatial resolution satellite soil moisture product over the Continental United States. *Journal of Hydrology, 588*, 125043

Fang, B., Lakshmi, V., Cosh, M., & Hain, C. (2021b). Very High Spatial Resolution Downscaled SMAP Radiometer Soil Moisture in the CONUS using VIIRS/MODIS data. *Journal of Selected Topics in Journal of Applied Earth Observations and Remote Sensing, In Press*

Fang, B., Lakshmi, V., Cosh, M., Liu, P.W., Bindlish, R., & Jackson, T.J. (2022). A global 1-km downscaled SMAP soil moisture product based on thermal inertia theory. *Vadose Zone Journal, 21*, e20182

Fang, B., Lakshmi, V., Cosh, M.H., & Hain, C. (2021c). Very high spatial resolution downscaled SMAP radiometer soil moisture in the CONUS using VIIRS/MODIS data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14*, 4946-4965

Fatichi, S., Pappas, C., & Ivanov, V.Y. (2016). Modeling plant–water interactions: an ecohydrological overview from the cell to the global scale. *Wiley Interdisciplinary Reviews: Water, 3*, 327-368

Fernandez-Moran, R., Al-Yaari, A., Mialon, A., Mahmoodi, A., Al Bitar, A., De Lannoy, G., Rodriguez-Fernandez, N., Lopez-Baeza, E., Kerr, Y., & Wigneron, J.-P. (2017a). SMOS-IC: An alternative SMOS soil moisture and vegetation optical depth product. *Remote Sensing, 9*, 457

Fernandez-Moran, R., Wigneron, J.-P., De Lannoy, G., Lopez-Baeza, E., Parrens, M., Mialon, A., Mahmoodi, A., Al-Yaari, A., Bircher, S., & Al Bitar, A. (2017b). A new calibration of the effective scattering albedo and soil roughness parameters in the SMOS SM retrieval algorithm. *International Journal of Applied Earth Observation and Geoinformation, 62*, 27-38

Findell, K.L., Gentine, P., Lintner, B.R., & Kerr, C. (2011). Probability of afternoon precipitation in eastern United States and Mexico enhanced by high evaporation. *Nature geoscience, 4*, 434-439

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment, 114*, 168-182

Ge, Y., Thomasson, J.A., & Sui, R. (2011). Remote sensing of soil properties in precision agriculture: A review. *Frontiers of Earth Science, 5*, 229-238

Gruber, A., Su, C.H., Zwieback, S., Crow, W., Dorigo, W., & Wagner, W. (2016). Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied Earth Observation and Geoinformation, 45, Part B*, 200-211

Guillod, B.P., Orlowsky, B., Miralles, D.G., Teuling, A.J., & Seneviratne, S.I. (2015). Reconciling spatial and temporal soil moisture effects on afternoon rainfall. *Nature communications, 6*, 6443

Gupta, P., Christopher, S.A., Wang, J., Gehrig, R., Lee, Y., & Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment, 40*, 5880-5892

Hagan, D.F.T., Wang, G., Kim, S., Parinussa, R.M., Liu, Y., Ullah, W., Bhatti, A.S., Ma, X., Jiang, T., & Su, B. (2020). Maximizing temporal correlations in long-term global satellite soil moisture data-merging. *Remote Sensing, 12*, 2164

Hahn, S. (2016). Algorithm Theoretical Baseline Document (ATBD) Surface Soil Moisture ASCAT Data Record Time Series. *Tech. Rep. Doc. No: SAF/HSAF/CDOP2/ATBD, v0.4*

Hallikainen, M.T., Ulaby, F.T., Dobson, M.C., El-Rayes, M.A., & Wu, L.-K. (1985). Microwave dielectric behavior of wet soil-part 1: Empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, 25-34

He, K., Zhao, W., Brocca, L., & Quintana-Seguí, P. (2022). SMPD: A soil moisture-based precipitation downscaling method for high-resolution daily satellite precipitation estimation. *Hydrology and Earth System Sciences Discussions*, 1-29

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., & Bauer-Marschallinger, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one, 12*, e0169748

Hirschi, M., Mueller, B., Dorigo, W., & Seneviratne, S.I. (2014). Using remotely sensed soil moisture for land–atmosphere coupling diagnostics: The role of surface vs. root-zone soil moisture variability. *Remote sensing of environment, 154*, 246-252

Hogg, R.V., Tanis, E.A., & Zimmerman, D.L. (2010). *Probability and statistical inference*. Pearson/Prentice Hall Upper Saddle River, NJ, USA:

Holmes, T., De Rosnay, P., De Jeu, R., Wigneron, R.P., Kerr, Y., Calvet, J.C., Escorihuela, M., Saleh, K., & Lemaître, F. (2006). A new parameterization of the effective temperature for L band radiometry. *Geophysical research letters, 33*

Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K., & Iguchi, T. (2014). The global precipitation measurement mission. *Bulletin of the American meteorological Society, 95*, 701-722

Huang, P., Patel, M., & Bobet, A. (2008). FHWA/IN/JTRP-2008/2 Classification of Organic Soils. In: School of Civil Engineering Purdue University, West Lafayette, Indiana

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J., Sorooshian, S., Tan, J., & Xie, P. (2019). NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). *Algorithm Theoretical Basis Document (ATBD) Version, 6*

Jackson, T.J. (1993). III. Measuring surface soil moisture using passive microwave remote sensing. *Hydrological processes, 7*, 139-152

Karthikeyan, L., & Mishra, A.K. (2021). Multi-layer high-resolution soil moisture estimation using machine learning over the United States. *Remote sensing of environment, 266*, 112706

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D.N., & Wood, E.F. (2017). Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms. *Advances in Water Resources, 109*, 106-120

Kellner, E., & Lundin, L.-C. (2001). Calibration of time domain reflectometry for water content in peat soil. *Hydrology Research, 32*, 315-332

Kerr, Y.H. (2007). Soil moisture from space: Where are we? *Hydrogeology journal, 15*, 117-120

Kerr, Y.H., Waldteufel, P., Richaume, P., Wigneron, J.P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., & Juglea, S.E. (2012). The SMOS soil moisture retrieval algorithm. *IEEE Transactions on Geoscience and Remote Sensing, 50*, 1384-1403

Kerr, Y.H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., & Gruhier, C. (2010). The SMOS mission: New tool for monitoring key elements ofthe global water cycle. *Proceedings of the IEEE, 98*, 666-687

Kerr, Y.H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., & Berger, M. (2001). Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing, 39*, 1729-1735

Kerr, Y.H., Wigneron, J.-P., Al Bitar, A., Mialon, A., & Srivastava, P. (2016). Soil moisture from space: Techniques and limitations. *Satellite Soil Moisture Retrieval* (pp. 3-27): Elsevier

Kim, H., & Lakshmi, V. (2018). Use of Cyclone Global Navigation Satellite System (CYGNSS) observations for estimation of soil moisture. *Geophysical Research Letters, 45*, 8272-8282

Kim, H., Parinussa, R., Konings, A.G., Wagner, W., Cosh, M.H., Lakshmi, V., Zohaib, M., & Choi, M. (2018). Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products. *Remote Sensing of Environment, 204*, 260-275

Kim, S., Parinussa, R.M., Liu, Y.Y., Johnson, F.M., & Sharma, A. (2015). A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation. *Geophysical Research Letters, 42*, 6662-6670

Kim, S., Parinussa, R.M., Liu, Y.Y., Johnson, F.M., & Sharma, A. (2016). Merging Alternate Remotely-Sensed Soil Moisture Retrievals Using a Non-Static Model Combination Approach. *Remote Sensing, 8*, 518

Kim, S., Zhang, R., Pham, H., & Sharma, A. (2019). A review of satellite-derived soil moisture and its usage for flood estimation. *Remote Sensing in Earth Systems Sciences, 2*, 225-246

Koster, R.D., Crow, W.T., Reichle, R.H., & Mahanama, S.P. (2018). Estimating basin‐scale water budgets with SMAP soil moisture data. *Water Resources Research, 54*, 4228-4244

Koster, R.D., Dirmeyer, P.A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C., Kanae, S., Kowalczyk, E., & Lawrence, D. (2004). Regions of strong coupling between soil moisture and precipitation. *Science, 305*, 1138-1140

Koster, R.D., Mahanama, S., Yamada, T., Balsamo, G., Berg, A., Boisserie, M., Dirmeyer, P., Doblas‐Reyes, F., Drewitt, G., & Gordon, C. (2010). Contribution of land surface initialization to subseasonal forecast skill: First results from a multi‐model experiment. *Geophysical Research Letters, 37*

Koster, R.D., & Suarez, M.J. (2001). Soil moisture memory in climate models. *Journal of Hydrometeorology, 2*, 558-570

Lakshmi, V., Wood, E.F., & Choudhury, B.J. (1997). A soil‐canopy‐atmosphere model for use in satellite microwave remote sensing. *Journal of Geophysical Research: Atmospheres, 102*, 6911-6927

Layden, A., Merchant, C., & MacCallum, S. (2015). Global climatology of surface water temperatures of large lakes by remote sensing. *International Journal of Climatology, 35*, 4464-4479

Leavesley, G., David, O., Garen, D., Lea, J., Marron, J., Pagano, T., Perkins, T., & Strobel, M. (2008). A modeling framework for improved agricultural water supply forecasting. In, *AGU Fall Meeting Abstracts* (p. 0497)

Li, X., Wigneron, J.-P., Frappart, F., De Lannoy, G., Fan, L., Zhao, T., Gao, L., Tao, S., Ma, H., & Peng, Z. (2022). The first global soil moisture and vegetation optical depth product retrieved from fused SMOS and SMAP L-band observations. *Remote sensing of environment, 282*, 113272

Lieberherr, G., & Wunderle, S. (2018). Lake surface water temperature derived from 35 years of AVHRR sensor data for European lakes. *Remote Sensing, 10*, 990

Lin, Y., & Mitchell, K.E. (2005). 1.2 the NCEP stage II/IV hourly precipitation analyses: Development and applications. In, *Proceedings of the 19th Conference Hydrology, American Meteorological Society, San Diego, CA, USA*: Citeseer

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., & Loumagne, C. (2014). When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrology and Earth System Sciences, 18*, 575-594

Loveland, T.R., & Belward, A. (1997). The IGBP-DIS global 1km land cover data set, DISCover: first results. *International Journal of Remote Sensing, 18*, 3289-3295

Lv, S., Wen, J., Zeng, Y., Tian, H., & Su, Z. (2014). An improved two-layer algorithm for estimating effective soil temperature in microwave radiometry using in situ temperature and soil moisture measurements. *Remote sensing of environment, 152*, 356-363

Ma, H., Zeng, J., Chen, N., Zhang, X., Cosh, M.H., & Wang, W. (2019). Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations. *Remote Sensing of Environment, 231*, 111215

Maberly, S.C., O'Donnell, R.A., Woolway, R.I., Cutler, M.E., Gong, M., Jones, I.D., Merchant, C.J., Miller, C.A., Politi, E., & Scott, E.M. (2020). Global lake thermal regions shift under climate change. *Nature communications, 11*, 1-9

MacCallum, S.N., & Merchant, C.J. (2012). Surface water temperature observations of large lakes by optimal estimation. *Canadian Journal of Remote Sensing, 38*, 25-45

Magan, B., Kim, S., Wasko, C., Barbero, R., Moron, V., Nathan, R., & Sharma, A. (2020). Impact of atmospheric circulation on the rainfall-temperature relationship in Australia. *Environmental Research Letters, 15*, 094098

Malicki, M., Plagge, R., & Roth, C. (1996). Improving the calibration of dielectric TDR soil moisture determination taking into account the solid soil. *European Journal of Soil Science, 47*, 357-366

Manrique-Suñén, A., Nordbo, A., Balsamo, G., Beljaars, A., & Mammarella, I. (2013). Representing land surface heterogeneity: Offline analysis of the tiling method. *Journal of Hydrometeorology, 14*, 850-867

McColl, K.A., Alemohammad, S.H., Akbar, R., Konings, A.G., Yueh, S., & Entekhabi, D. (2017). The global distribution and dynamics of surface soil moisture. *Nature geoscience, 10*, 100-104

McColl, K.A., Vogelzang, J., Konings, A.G., Entekhabi, D., Piles, M., & Stoffelen, A. (2014). Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophysical Research Letters, 41*, 6229-6236

Mialon, A., Richaume, P., Leroux, D., Bircher, S., Al Bitar, A., Pellarin, T., Wigneron, J.-P., & Kerr, Y.H. (2015). Comparison of Dobson and Mironov dielectric models in the SMOS soil moisture retrieval algorithm. *IEEE Transactions on Geoscience and Remote Sensing, 53*, 3084-3094

Miralles, D.G., Teuling, A.J., Van Heerwaarden, C.C., & Vilà-Guerau de Arellano, J. (2014). Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature geoscience, 7*, 345-349

Mironov, D. (2008). Parameterization of Lakes in Numerical Weather Prediction. Description of a Lake Model. In (p. 44). COSMO Technical Report

Mironov, D., Heise, E., Kourzeneva, E., Ritter, B., Schneider, N., & Terzhevik, A. (2010). Implementation of the lake parameterisation scheme FLake into the numerical weather prediction model COSMO. *Boreal Environment Research, 15*, 218-230

Mironov, V., Kerr, Y., Wigneron, J.-P., Kosolapova, L., & Demontoux, F. (2012). Temperature- and texture-dependent dielectric model for moist soils at 1.4 GHz. *IEEE Geoscience and Remote Sensing Letters, 10*, 419-423

Mironov, V.L., Kosolapova, L.G., & Fomin, S.V. (2009). Physically and mineralogically based spectroscopic dielectric model for moist soils. *IEEE transactions on Geoscience and remote sensing, 47*, 2059-2070

Mironov, V.L., Kosolapova, L.G., Fomin, S.V., & Savin, I.V. (2019). Experimental analysis and empirical model of the complex permittivity of five organic soils at 1.4 GHz in the temperature range from− 30° C to 25° C. *IEEE Transactions on Geoscience and Remote Sensing, 57*, 3778-3787

Mishra, V., Cruise, J.F., Hain, C.R., Mecikalski, J.R., & Anderson, M.C. (2018). Development of soil moisture profiles through coupled microwave–thermal infrared observations in the<? xmltex\break?> southeastern United States. *Hydrology and Earth System Sciences, 22*, 4935-4957

Mishra, V., Ellenburg, W.L., Markert, K.N., & Limaye, A.S. (2020). Performance evaluation of soil moisture profile estimation through entropy-based and exponential filter models. *Hydrological Sciences Journal, 65*, 1036-1048

Mo, T., Choudhury, B., Schmugge, T., Wang, J., & Jackson, T. (1982). A model for microwave emission from vegetation‐covered fields. *Journal of Geophysical Research: Oceans, 87*, 11229-11237

Moghaddam, M., Entekhabi, D., Goykhman, Y., Li, K., Liu, M., Mahajan, A., Nayyar, A., Shuman, D., & Teneketzis, D. (2010). A wireless soil moisture smart sensor web using physics-based optimal control: Concept and initial demonstrations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3*, 522-535

Morrison, K., & Wagner, W. (2019). Explaining Anomalies in SAR and Scatterometer Soil Moisture Retrievals From Dry Soils With Subsurface Scattering. *IEEE Transactions on Geoscience and Remote Sensing, 58*, 2190-2197

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., & Hersbach, H. (2021a). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data Discussions*, 1-50

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., & Hersbach, H. (2021b). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data, 13*, 4349-4383

Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S., & Wagner, W. (2009). An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. *IEEE Transactions on Geoscience and Remote Sensing, 47*, 1999-2013

NASA-LP-DAAC (2012). Land Cover Type Yearly L3 Global 0.05Deg CMG. In: NASA LP DACC, USGS EROS Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov)

Njoku, E.G., & Entekhabi, D. (1996). Passive microwave remote sensing of soil moisture. *Journal of Hydrology, 184*, 101-129

Njoku, E.G., Jackson, T.J., Lakshmi, V., Chan, T.K., & Nghiem, S.V. (2003). Soil moisture retrieval from AMSR-E. *IEEE transactions on geoscience and remote sensing, 41*, 215-229

Njoku, E.G., & Kong, J.A. (1977). Theory for passive microwave remote sensing of near-surface soil moisture. *Journal of Geophysical Research, 82*, 3108-3118

O'Neill, P., Bindlish, R., Chan, S., Chaubell, J., Colliander, A., Njoku, E., & Jackson, T. (2021a). Algorithm Theoretical Basis Document Level 2 & 3 Soil Moisture (Passive) Data Products, revision G, October 12, 2021, SMAP Project, JPL D-66480, Jet Propulsion Laboratory, Pasadena, CA. . In

O'Neill, P., Chan, S., Njoku, E., Jackson, T., Bindlish, R., & Chaubell, J. (2021b). L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 8. Soil Moisture Retrieval Data. Boulder, Colorado, USA. In, . NASA National Snow and Ice Data Center Distributed Active Archive Center

O'Neill, P., Chan, S., Njoku, E., Jackson, T., Bindlish, R., Chaubell, J., & Colliander, A. (2021c). SMAP Enhanced L3 Radiometer Global and Polar Grid Daily 9 km EASE-Grid Soil Moisture, Version 5. Soil Moisture. Boulder, Colorado USA. In. NASA Snow and ICE Data Center Distributed Active Archive Center

O'Neill, P., & Jackson, T. (1990). Observed effects of soil organic matter content on the microwave emissivity of soils. *Remote sensing of environment, 31*, 175-182

O'Neill, P.E., Chan, S., Njoku, E.G., Jackson, T., & Bindlish, R. (2018). SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 5 (SPL3SMP). *In. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center*

O'Reilly, C.M., Sharma, S., Gray, D.K., Hampton, S.E., Read, J.S., Rowley, R.J., Schneider, P., Lenters, J.D., McIntyre, P.B., & Kraemer, B.M. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters, 42*, 10,773-710,781

O'Neill, P., Njoku, E., Jackson, T., Chan, S., & Bindlish, R. (2015). SMAP algorithm theoretical basis document: Level 2 & 3 soil moisture (passive) data products. *Jet Propulsion Lab., California Inst. Technol., Pasadena, CA, USA, JPL D-66480*

Ojo, E.R., Bullock, P.R., L'Heureux, J., Powers, J., McNairn, H., & Pacheco, A. (2015). Calibration and evaluation of a frequency domain reflectometry sensor for real‐time soil moisture monitoring. *Vadose Zone Journal, 14*, 1-12

Pallant, J. (2011). Survival manual. *A step by step guide to data analysis using SPSS, 4*

Paquet, J., Caron, J., & Banton, O. (1993). In situ determination of the water desorption characteristics of peat substrates. *Canadian Journal of Soil Science, 73*, 329-339

Park, C.-H., Behrendt, A., LeDrew, E., & Wulfmeyer, V. (2017). New approach for calculating the effective dielectric constant of the moist soil for microwaves. *Remote Sensing, 9*, 732

Park, C.-H., Berg, A., Cosh, M.H., Colliander, A., Behrendt, A., Manns, H., Hong, J., Lee, J., Zhang, R., & Wulfmeyer, V. (2021). An inverse dielectric mixing model at 50 MHz that considers soil organic carbon. *Hydrology and Earth System Sciences, 25*, 6407-6420

Park, C.H., Montzka, C., Jagdhuber, T., Jonard, F., De Lannoy, G., Hong, J., Jackson, T.J., & Wulfmeyer, V. (2019). A dielectric mixing model accounting for soil organic matter. *Vadose Zone Journal, 18*, 190036

Parrens, M., Wigneron, J.-P., Richaume, P., Mialon, A., Al Bitar, A., Fernandez-Moran, R., Al-Yaari, A., & Kerr, Y.H. (2016). Global-scale surface roughness effects at L-band as estimated from SMOS observations. *Remote sensing of environment, 181*, 122-136

Peel, M.C., Finlayson, B.L., & McMahon, T.A. (2007). Updated world map of the Köppen-Geiger climate classification

Pellarin, T., Wigneron, J.-P., Calvet, J.-C., Berger, M., Douville, H., Ferrazzoli, P., Kerr, Y.H., Lopez-Baeza, E., Pulliainen, J., & Simmonds, L.P. (2003). Two-year global simulation of L-band brightness temperatures over land. *IEEE Transactions on Geoscience and Remote Sensing, 41*, 2135-2139

Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M.H., Crow, W.T., Dabrowska-Zielinska, K., Dadson, S., & Davidson, M.W. (2021). A roadmap for high-resolution satellite soil moisture applications–confronting product characteristics with user requirements. *Remote sensing of environment, 252*, 112162

Peplinski, N.R., Ulaby, F.T., & Dobson, M.C. (1995). Dielectric properties of soils in the 0.3-1.3-GHz range. *IEEE Transactions on Geoscience and Remote Sensing, 33*, 803-807

Petropoulos, G. (2013). *Remote sensing of energy fluxes and soil moisture content*. CRC press

Petropoulos, G.P., Ireland, G., & Barrett, B. (2015). Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth, Parts A/B/C, 83*, 36-56

Pham, H.T., Kim, S., Marshall, L., & Johnson, F. (2019). Using 3D robust smoothing to fill land surface temperature gaps at the continental scale. *International journal of applied earth observation and geoinformation, 82*, 101879

Politi, E., MacCallum, S., Cutler, M., Merchant, C., Rowan, J., & Dawson, T. (2016). Selection of a network of large lakes and reservoirs suitable for global environmental change analysis using Earth Observation. *International Journal of Remote Sensing, 37*, 3042-3060

Pribyl, D.W. (2010). A critical review of the conventional SOC to SOM conversion factor. *Geoderma, 156*, 75-83

Qiu, Y., Fu, B.-j., Wang, J., Chen, L., Meng, Q., & Zhang, Y. (2010). Spatial prediction of soil moisture content using multiple-linear regressions in a gully catchment of the Loess Plateau, China. *Journal of arid environments, 74*, 208-220

Reich, P.B., Sendall, K.M., Stefanski, A., Rich, R.L., Hobbie, S.E., & Montgomery, R.A. (2018). Effects of climate warming on photosynthesis in boreal tree species depend on soil moisture. *Nature, 562*, 263-267

Reichle, R.H., Draper, C.S., Liu, Q., Girotto, M., Mahanama, S.P., Koster, R.D., & De Lannoy, G.J. (2017). Assessment of MERRA-2 land surface hydrology estimates. *Journal of Climate, 30*, 2937-2960

Reichle, R.H., Koster, R.D., Dong, J., & Berg, A.A. (2004). Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. *Journal of Hydrometeorology, 5*, 430-442

Robock, A., & Li, H. (2006). Solar dimming and CO2 effects on soil moisture trends. *Geophysical Research Letters, 33*

Roth, C., Malicki, M., & Plagge, R. (1992). Empirical evaluation of the relationship between soil dielectric constant and volumetric water content as the basis for calibrating soil moisture measurements by TDR. *Journal of Soil Science, 43*, 1-13

Ruf, C.S., Chew, C., Lang, T., Morris, M.G., Nave, K., Ridley, A., & Balasubramaniam, R. (2018). A new paradigm in earth environmental monitoring with the cygnss small satellite constellation. *Scientific reports, 8*, 1-13

Sabater, J.M. (2019). ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). In

Sadeghi, M., Nguyen, P., Naeini, M.R., Hsu, K., Braithwaite, D., & Sorooshian, S. (2021). PERSIANN-CCS-CDR, a 3-hourly 0.04° global precipitation climate data record for heavy precipitation studies. *Scientific Data, 8*, 1-11

Salvucci, G.D. (2001). Estimating the moisture dependence of root zone water loss using conditionally averaged precipitation. *Water Resources Research, 37*, 1357-1365

Samuelsson, P., Kourzeneva, E., & Mironov, D. (2010). The impact of lakes on the European climate as simulated by a regional climate model. *Boreal Environment Research, 15*, 113 - 129

Sanchez, N., Martínez-Fernández, J., Scaini, A., & Perez-Gutierrez, C. (2012). Validation of the SMOS L2 soil moisture data in the REMEDHUS network (Spain). *IEEE Transactions on Geoscience and Remote Sensing, 50*, 1602-1611

Schaefer, G.L., Cosh, M.H., & Jackson, T.J. (2007). The USDA natural resources conservation service soil climate analysis network (SCAN). *Journal of Atmospheric and Oceanic Technology, 24*, 2073-2077

Schaefer, G.L., & Paetzold, R.F. (2001). SNOTEL (SNOwpack TELemetry) and SCAN (soil climate analysis network). *Automated Weather Stations for Applications in Agriculture and Water Resources Management: Current Use and Future Perspectives, 1074*, 187-194

Schaeffer, B.A., Iiames, J., Dwyer, J., Urquhart, E., Salls, W., Rover, J., & Seegers, B. (2018). An initial validation of Landsat 5 and 7 derived surface water temperature for US lakes, reservoirs, and estuaries. *International Journal of Remote Sensing, 39*, 7789-7805

Schmugge, T., & Choudhury, B. (1981). A comparison of radiative transfer models for predicting the microwave emission from soils. *Radio Science, 16*, 927-938

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., & Teuling, A.J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews, 99*, 125-161

Sharifi, E., Saghafian, B., & Steinacker, R. (2019). Downscaling satellite precipitation estimates with multiple linear regression, artificial neural networks, and spline interpolation techniques. *Journal of Geophysical Research: Atmospheres, 124*, 789-805

Shen, X., Walker, J.P., Ye, N., Wu, X., Boopathi, N., Yeo, I.-Y., Zhang, L., & Zhu, L. (2020). Soil moisture retrieval depth of P-and L-band radiometry: Predictions and observations. *IEEE Transactions on Geoscience and Remote Sensing, 59*, 6814-6822

Skierucha, W. (2000). Accuracy of soil moisture measurement by TDR technique. *International agrophysics, 14*

Srivastava, P.K., Han, D., Ramirez, M.A.R., & Islam, T. (2013). Appraisal of SMOS soil moisture at a catchment scale in a temperate maritime climate. *Journal of Hydrology, 498*, 292-304

Srivastava, P.K., Han, D., Rico-Ramirez, M.A., O'Neill, P., Islam, T., Gupta, M., & Dai, Q. (2015). Performance evaluation of WRF-Noah Land surface model estimated soil moisture for hydrological application: Synergistic evaluation using SMOS retrieved soil moisture. *Journal of Hydrology, 529*, 200-212

Stoffelen, A. (1998). Toward the true near‐surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research: Oceans (1978–2012), 103*, 7755-7766

Taylor, C.M., de Jeu, R.A., Guichard, F., Harris, P.P., & Dorigo, W.A. (2012). Afternoon rain more likely over drier soils. *Nature, 489*, 423-426

Topp, G.C., Davis, J., & Annan, A.P. (1980). Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resources Research, 16*, 574-582

Ulaby, F.T., Long, D.G., Blackwell, W.J., Elachi, C., Fung, A.K., Ruf, C., Sarabandi, K., Zebker, H.A., & Van Zyl, J. (2014). *Microwave radar and radiometric remote sensing*. University of Michigan Press Ann Arbor, MI, USA

Ulaby, F.T., Moore, R.K., & Fung, A.K. (1986a). *Radar Remote Sensing and Surface Scattering and Emission Theory*. Norwood, MA 02062: Artech House

Ulaby, F.T., Moore, R.K., & Fung, A.K. (1986b). *Volume Scattering and Emission Theory, Advanced Systems, and Applications*. Norwood, MA 02062: Artech House

Van den Hurk, B.J., Graham, L.P., & Viterbo, P. (2002). Comparison of land surface hydrology in regional climate simulations of the Baltic Sea catchment. *Journal of Hydrology, 255*, 169-193

Vereecken, H., Huisman, J., Pachepsky, Y., Montzka, C., Van Der Kruk, J., Bogena, H., Weihermüller, L., Herbst, M., Martinez, G., & Vanderborght, J. (2014). On the spatio-temporal dynamics of soil moisture at the field scale. *Journal of Hydrology, 516*, 76-96

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., & Schneider, S. (2013). The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift, 22*, 5-33

Wagner, W., Lemoine, G., & Rott, H. (1999). A method for estimating soil moisture from ERS scatterometer and soil data. *Remote sensing of environment, 70*, 191-207

Wagner, W., Scipal, K., Pathe, C., Gerten, D., Lucht, W., & Rudolf, B. (2003). Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. *Journal of Geophysical Research: Atmospheres, 108*

Walker, J.P., Willgoose, G.R., & Kalma, J.D. (2001). One-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: A simplified soil moisture model and field application. *Journal of Hydrometeorology, 2*, 356-373

Wan, W., Li, H., Xie, H., Hong, Y., Long, D., Zhao, L., Han, Z., Cui, Y., Liu, B., & Wang, C. (2017). A comprehensive data set of lake surface water temperature over the Tibetan Plateau derived from MODIS LST products 2001–2015. *Scientific Data, 4*, 1-10

Wang, G., Garcia, D., Liu, Y., De Jeu, R., & Dolman, A.J. (2012). A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations. *Environmental Modelling & Software, 30*, 139-142

Wang, J.R., & Schmugge, T.J. (1980). An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Transactions on Geoscience and Remote Sensing, GE-18*, 288-295

Wasko, C., & Sharma, A. (2014). Quantile regression for investigating scaling of extreme precipitation with temperature. *Water Resources Research, 50*, 3608-3614

Wigneron, J.-P., Calvet, J.-C., De Rosnay, P., & Kerr, Y. (2006). L-MEB: A simple model at L-band for the continental areas-Application to the simulation of a half-degree resolution and global scale data set. In: IEE Electromagnetic Waves series

Wigneron, J.-P., Jackson, T., O'neill, P., De Lannoy, G., de Rosnay, P., Walker, J., Ferrazzoli, P., Mironov, V., Bircher, S., & Grant, J. (2017). Modelling the passive microwave signature from land surfaces: A review of recent results and application to the L-band SMOS & SMAP soil moisture retrieval algorithms. *Remote sensing of environment, 192*, 238-262

Wigneron, J.-P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.-J., Richaume, P., Ferrazzoli, P., De Rosnay, P., Gurney, R., & Calvet, J.-C. (2007). L-band microwave emission of the biosphere (L-MEB) model: Description and calibration against experimental data sets over crop fields. *Remote sensing of environment, 107*, 639-655

Wigneron, J.-P., Schmugge, T., Chanzy, A., Calvet, J.-C., & Kerr, Y.H. (1998). Use of passive microwave remote sensing to monitor soil moisture. *Agronomie, 18*, 27-43

Wigneron, J.-P., Waldteufel, P., Chanzy, A., Calvet, J.-C., & Kerr, Y. (2000). Two-dimensional microwave interferometer retrieval capabilities over land surfaces (SMOS mission). *Remote sensing of environment, 73*, 270-282

Wilheit, T.T. (1978). Radiative transfer in a plane stratified dielectric. *IEEE Transactions on Geoscience Electronics, 16*, 138-143

Woolway, R.I., & Merchant, C.J. (2019). Worldwide alteration of lake mixing regimes in response to climate change. *Nature geoscience, 12*, 271-276

Xia, Y., Watts, J.D., Machmuller, M.B., & Sanderman, J. (2022). Machine learning based estimation of field-scale daily, high resolution, multi-depth soil moisture for the Western and Midwestern United States. *PeerJ, 10*, e14275

Yi, Y., Kimball, J.S., Chen, R.H., Moghaddam, M., & Miller, C.E. (2019). Sensitivity of active-layer freezing process to snow cover in Arctic Alaska. *The Cryosphere, 13*, 197-218

Yilmaz, M.T., & Crow, W.T. (2013). The optimality of potential rescaling approaches in land data assimilation. *Journal of Hydrometeorology, 14*, 650-660

Yilmaz, M.T., & Crow, W.T. (2014). Evaluation of Assumptions in Soil Moisture Triple Collocation Analysis. *Journal of Hydrometeorology, 15*, 1293-1302

Zanella, A., Jabiol, B., Ponge, J.-F., Sartori, G., De Waal, R., Van Delft, B., Graefe, U., Cools, N., Katzensteiner, K., & Hager, H. (2011). European humus forms reference base

Zhang, Q., Yuan, Q., Li, J., Wang, Y., Sun, F., & Zhang, L. (2021a). Generating seamless global daily AMSR2 soil moisture (SGD-SM) long-term products for the years 2013–2019. *Earth System Science Data, 13*, 1385-1401

Zhang, R., Kim, S., & Sharma, A. (2019). A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes. *Remote sensing of environment, 223*, 82-94

Zhang, R., Kim, S., Sharma, A., & Lakshmi, V. (2021b). Identifying relative strengths of SMAP, SMOS-IC, and ASCAT to capture temporal variability. *Remote sensing of environment, 252*, 112126

Zhao, G., Gao, H., & Cai, X. (2020). Estimating lake temperature profile and evaporation losses by leveraging MODIS LST data. *Remote sensing of environment, 251*, 112104

Zreda, M., Desilets, D., Ferré, T., & Scott, R.L. (2008). Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophysical research letters, 35*

Supplementary Information

Supplementary Figures

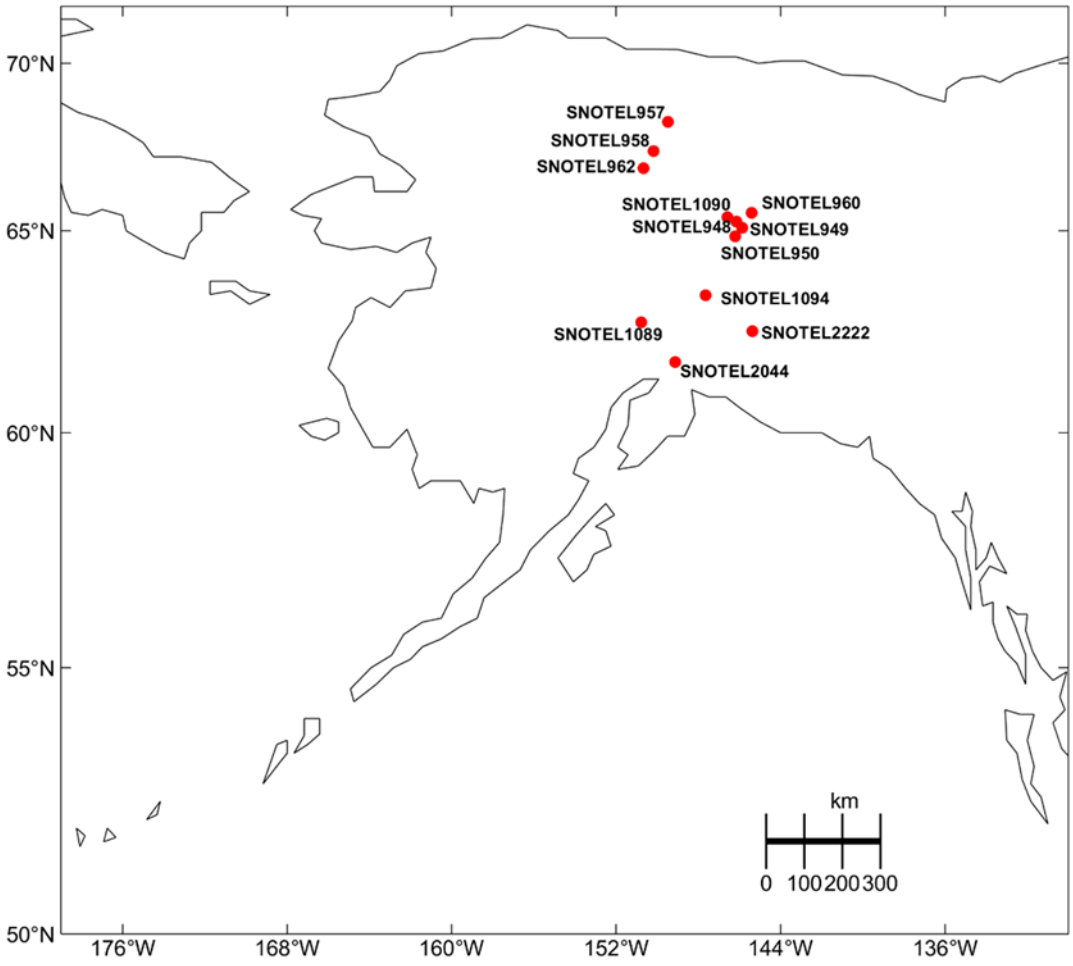
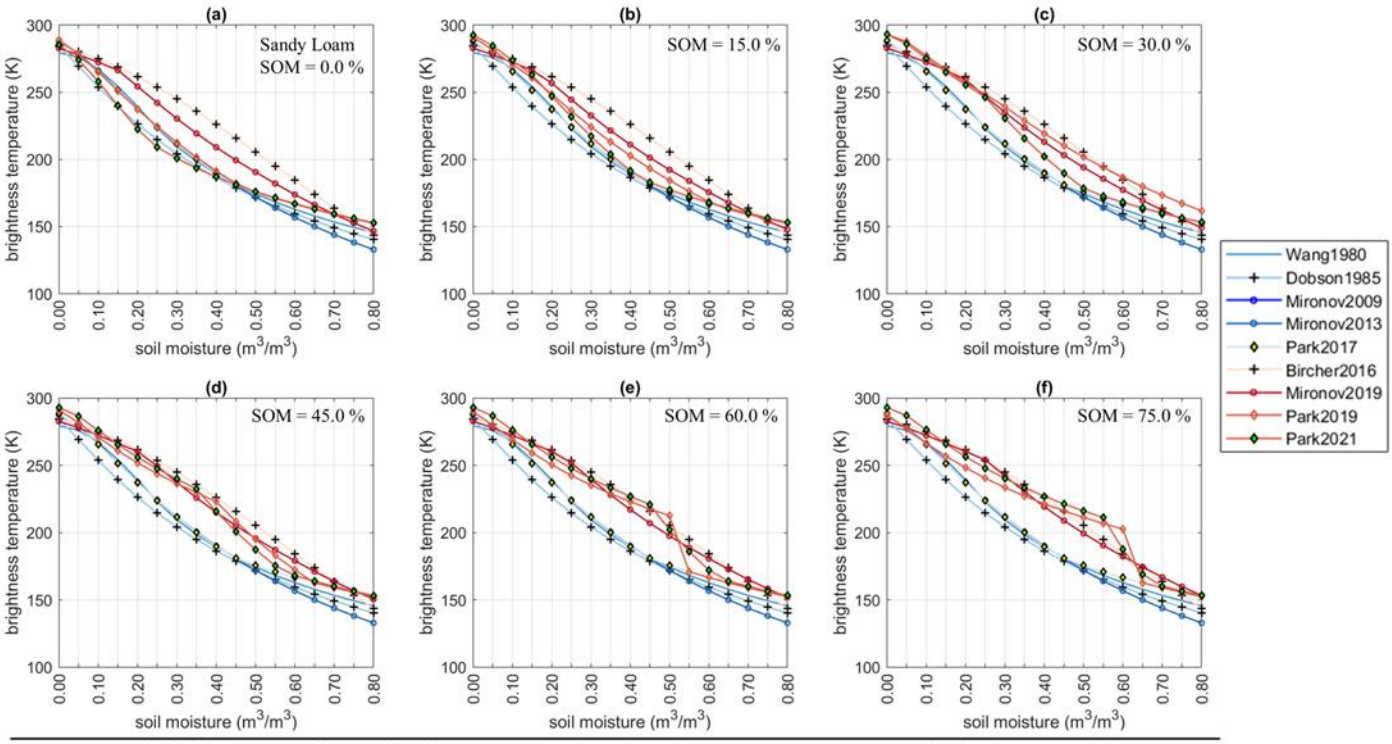


Figure S1. The geographical distributions of all the 12 stations finally used for validation.



Model Inputs	Gravimetric Sand Fraction (%)	Gravimetric Clay Fraction (%)	Gravimetric Silt Fraction (%)	Bulk Density (g/cm³)	Salinity (‰)	Soil Physical Temperature (°C)	Frequency (GHz)	Soil Organic Matter (%)	Volumetric Soil Moisture (m³/m³)	Incidence Angle (°)
Simulated Values	51.5	35.1	13.4	1.54	0.0	20.0	1.4	0.0, 15.0, 30.0, 45.0, 60.0, 75.0	0.0 – 0.8	40.0

Figure S2. Simulated brightness temperature of a sandy loam with various soil organic matter, and the accompanied table displays all the input values where most of soil parameters are directly taken from the sample of sandy loam used in Hallikainen et al., (1985).

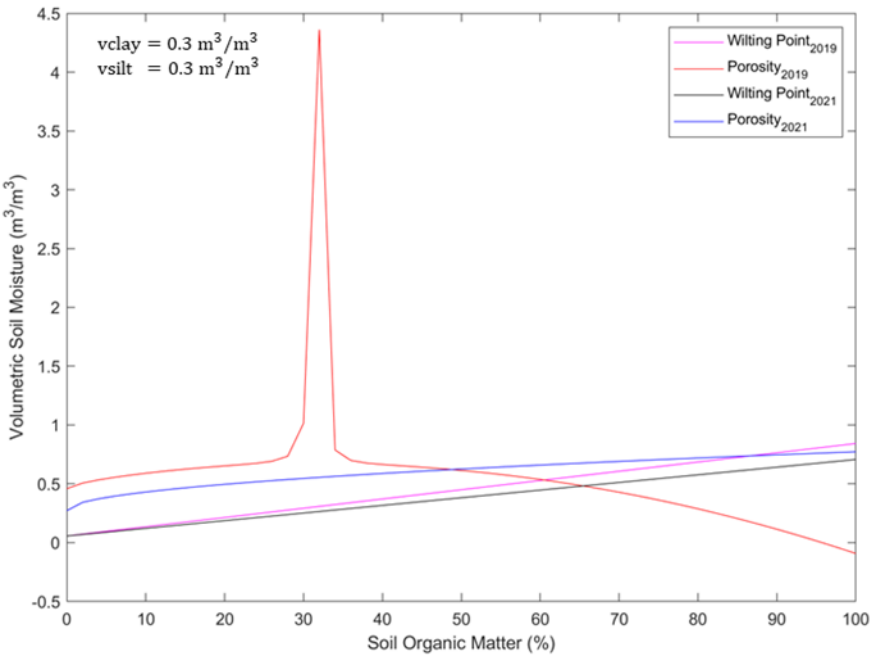


Figure S3. Variations of wilting point and porosity estimated from Park 2019 and Park 2021 with increasing soil organic matter with assumed volumetric textural compositions.

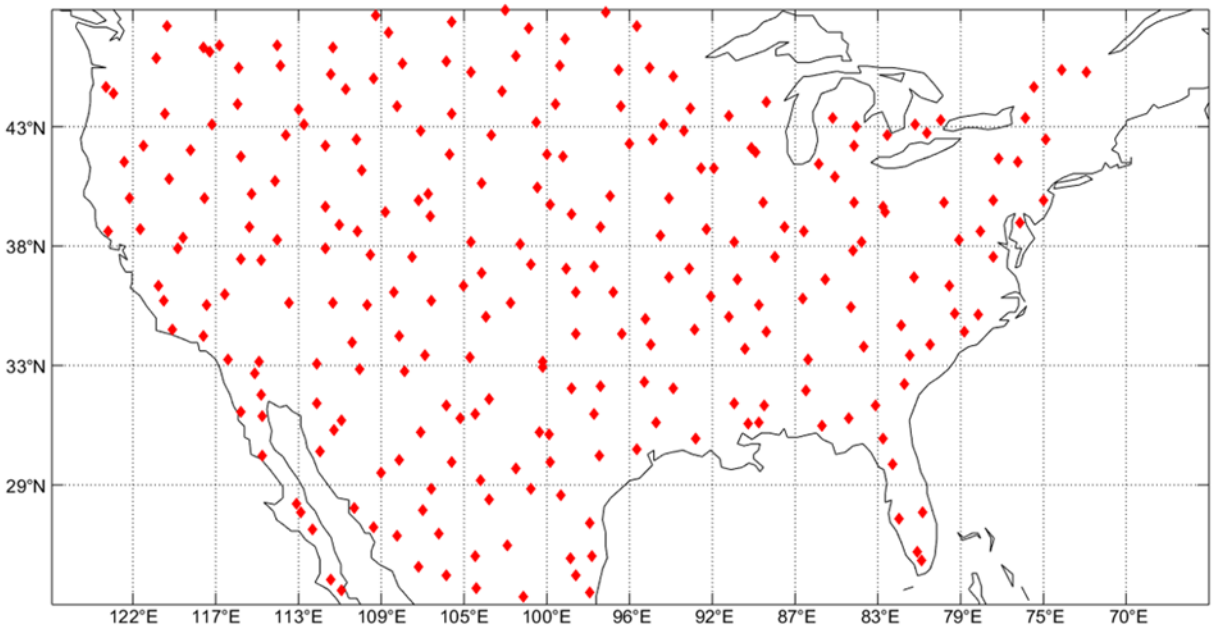


Figure S4. Distribution of 260 representative 9-km pixels used to seek for the optimal method flow for the precipitation-incorporated, 12-hour successive soil moisture simulations.

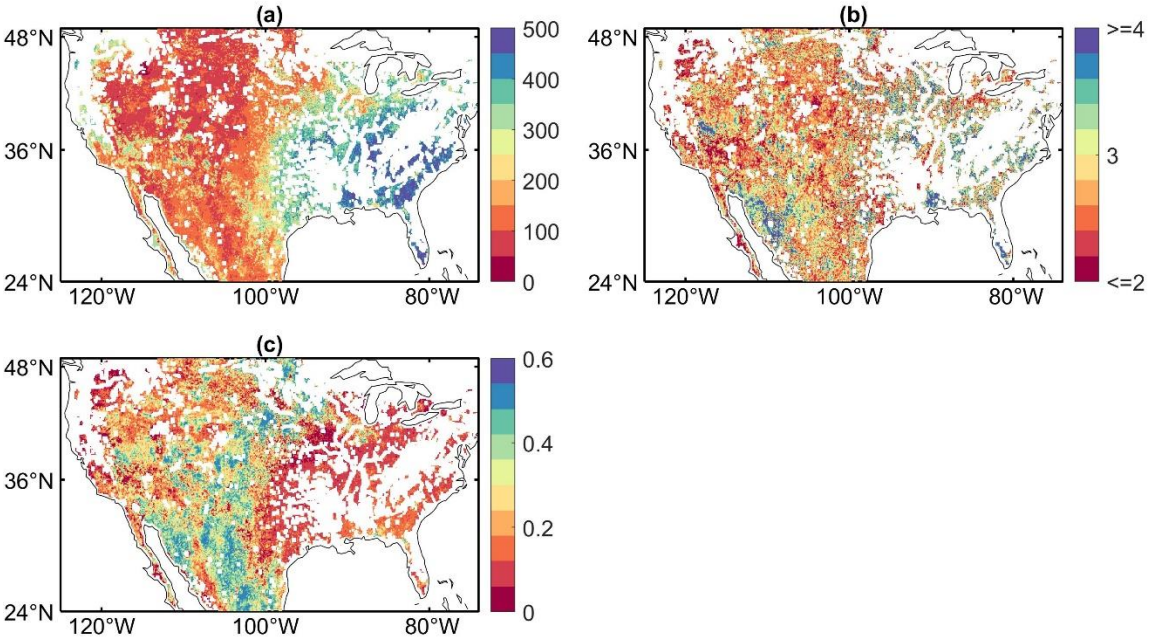


Figure S5. Spatial distributions of the optimal parameters (a) ΔZ (mm) (b) α, and (c) β over the CONUS.

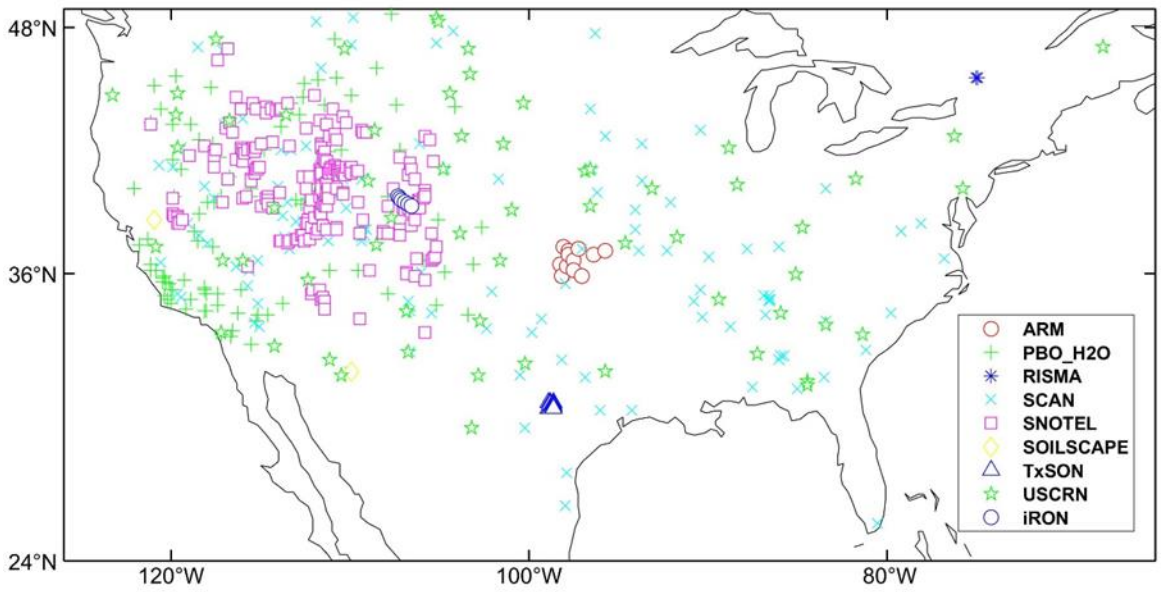


Figure S6. Spatial distributions of 526 ground stations used in validation.

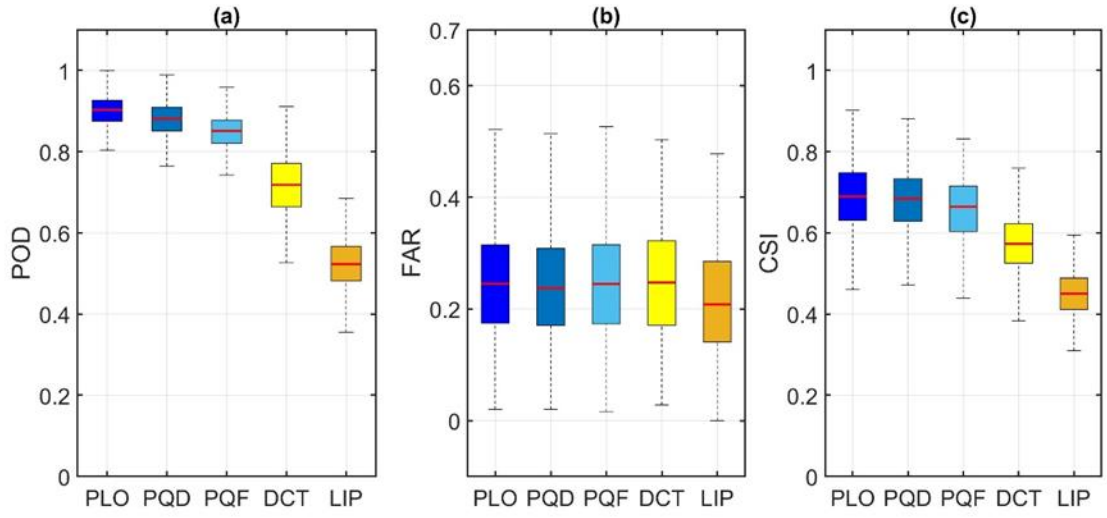


Figure S7. Boxplots of the categorical performance metrics of (a) POD, (b) FAR, and (c) CSI for five different SMAP-based gap-filled soil moisture datasets in capturing soil moisture peaks caused by all the rainfall events with the rates over 0.5 mm/day. Those soil moisture peaks observed by in-situ soil moisture measurements are used as the benchmarks.

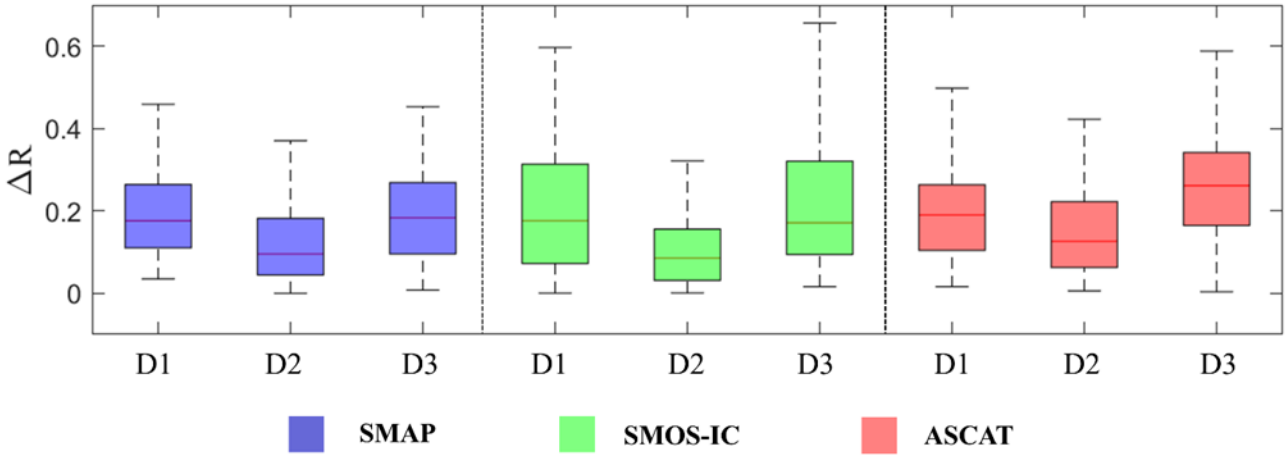


Figure S8. Boxplot of ΔR, the absolute values of differences between in-situ conR and the three types of R: in-situ TCR (D1=|in-situ TCR – in-situ conR|), conR (D2 =|conR- in-situ conR|) and TCR (D3 =|TCR- in-situ conR|).

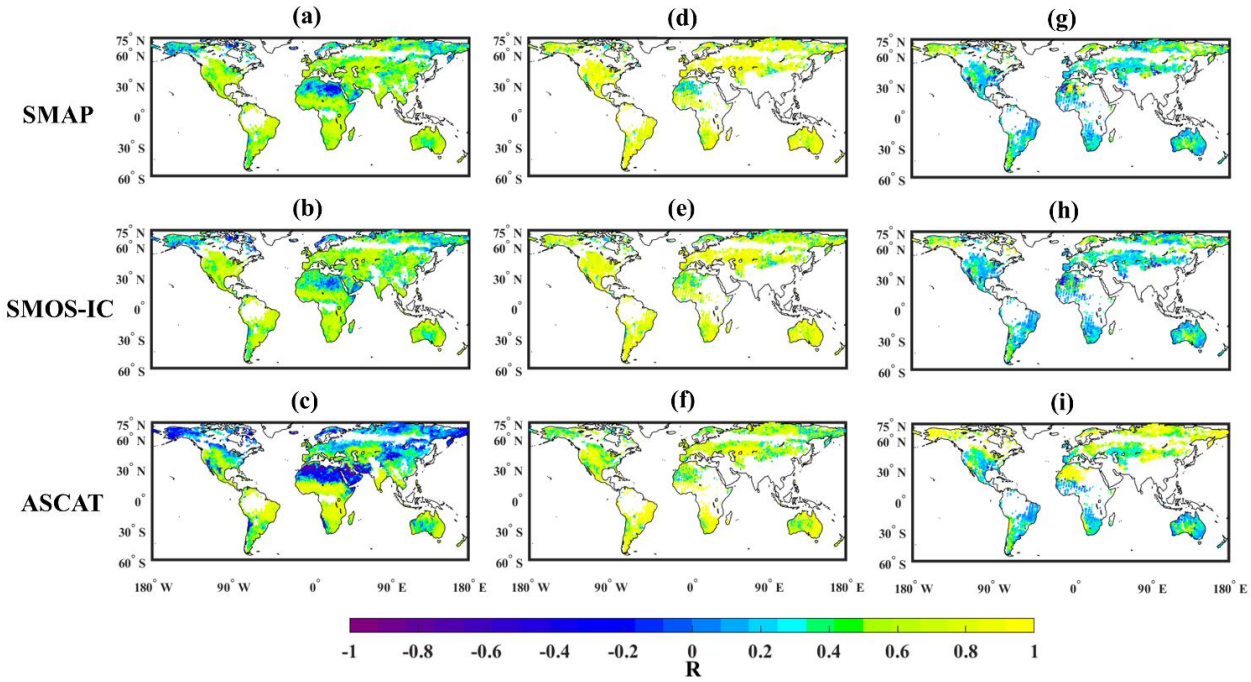


Figure S9. Global distribution of conventional R for SMAP, SMOS-IC, and ASCAT using MERRA2 as reference (a-c), and the TC-based R using ERA-Interim as the component of triplets (d-f). The second column "minus" the first column (g-i).

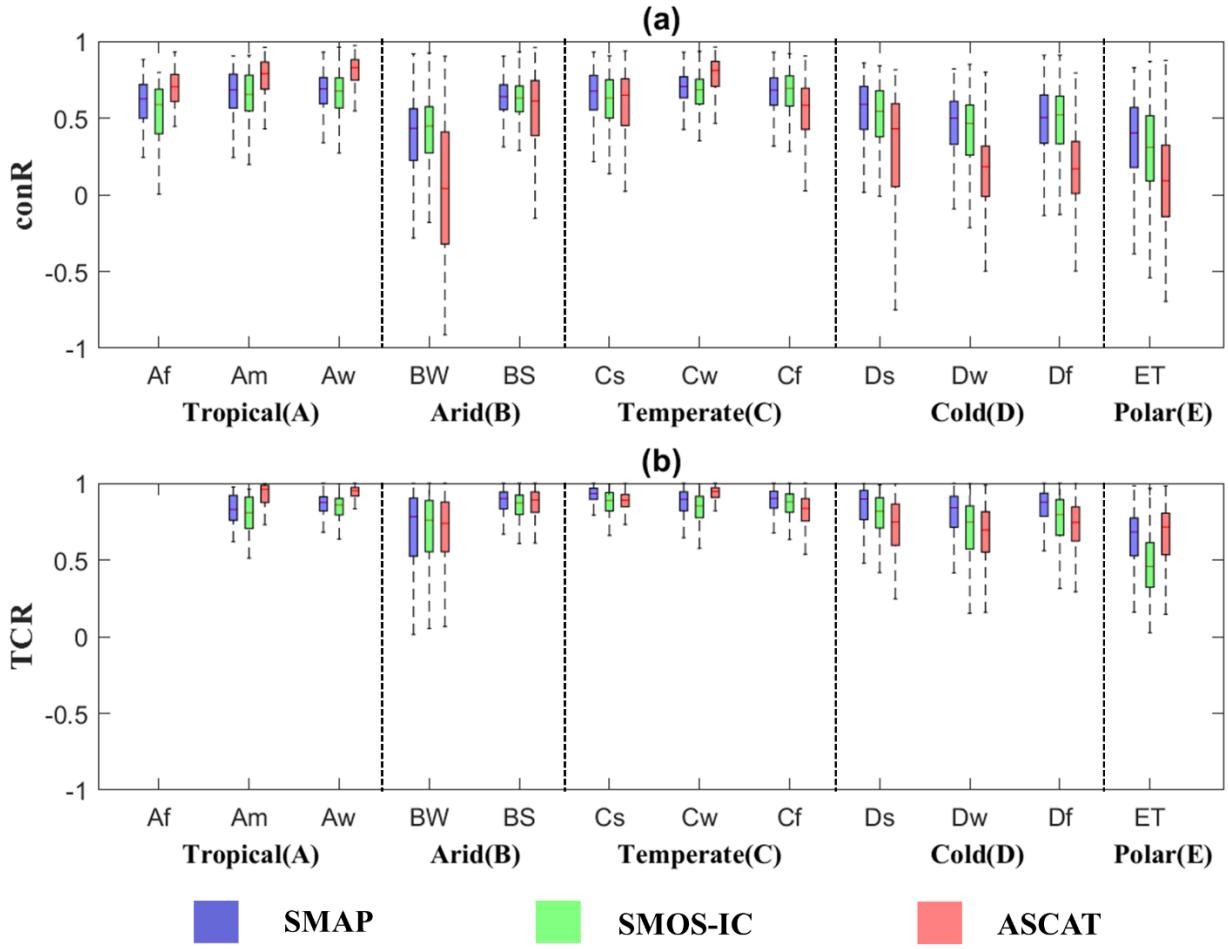


Figure S10. Boxplots for variations in the conR (a) against MERRA2 and TCR (b) under different climate zones. Tropical (Af: rainforest, Am: tropical monsoon, Aw: tropical savannah). Arid (BW: arid desert; BS: arid steppe). Temperate (Cs: temperate with dry summer; Cw: temperate with dry winter; Cf: temperate without dry season). cold and polar (Ds: cold with dry summer; Dw: cold with dry winter; Df: cold without dry season; ET: polar tundra).

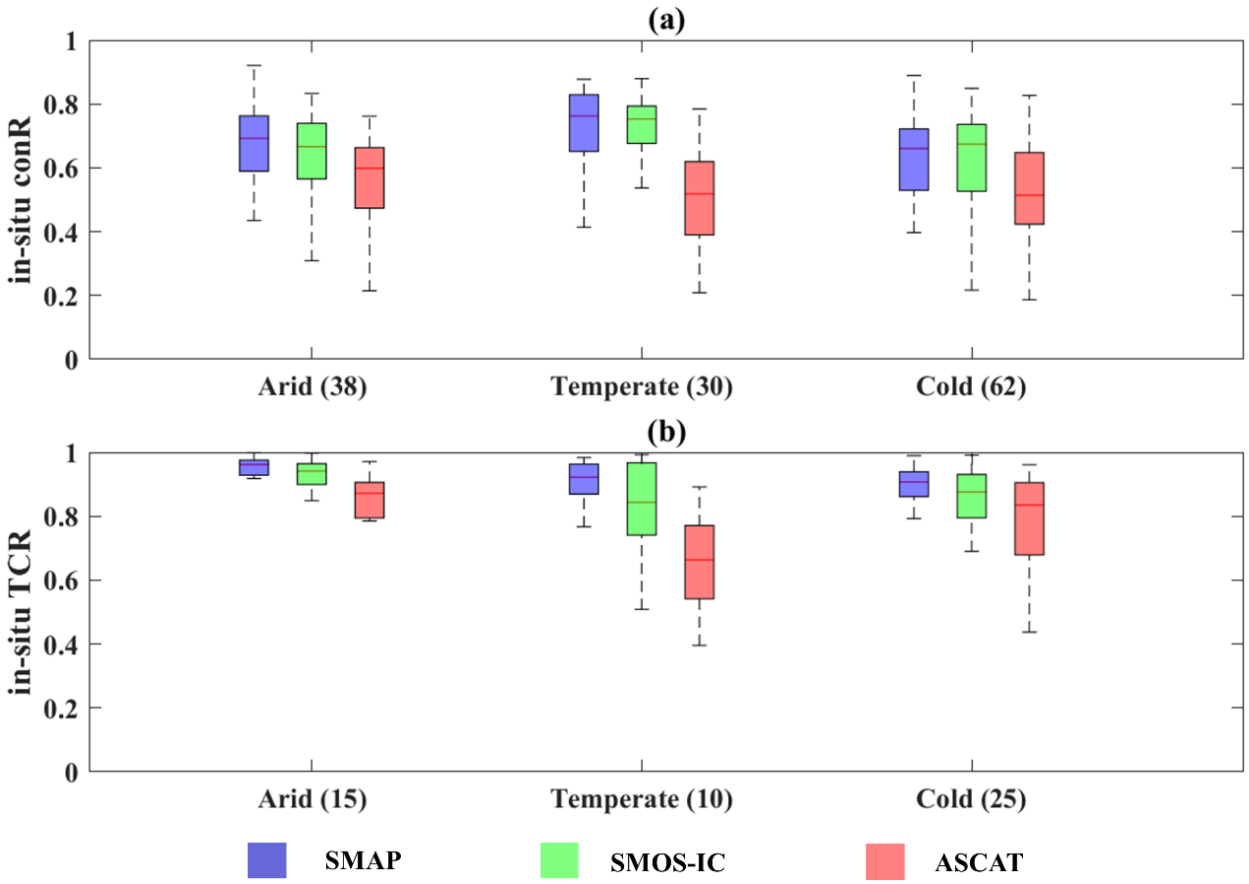


Figure S11. Boxplots for variations in the in-situ conR (a) and in-situ TCR (b) under different climate zones. The value behind the CZ class is the number of pixels included.

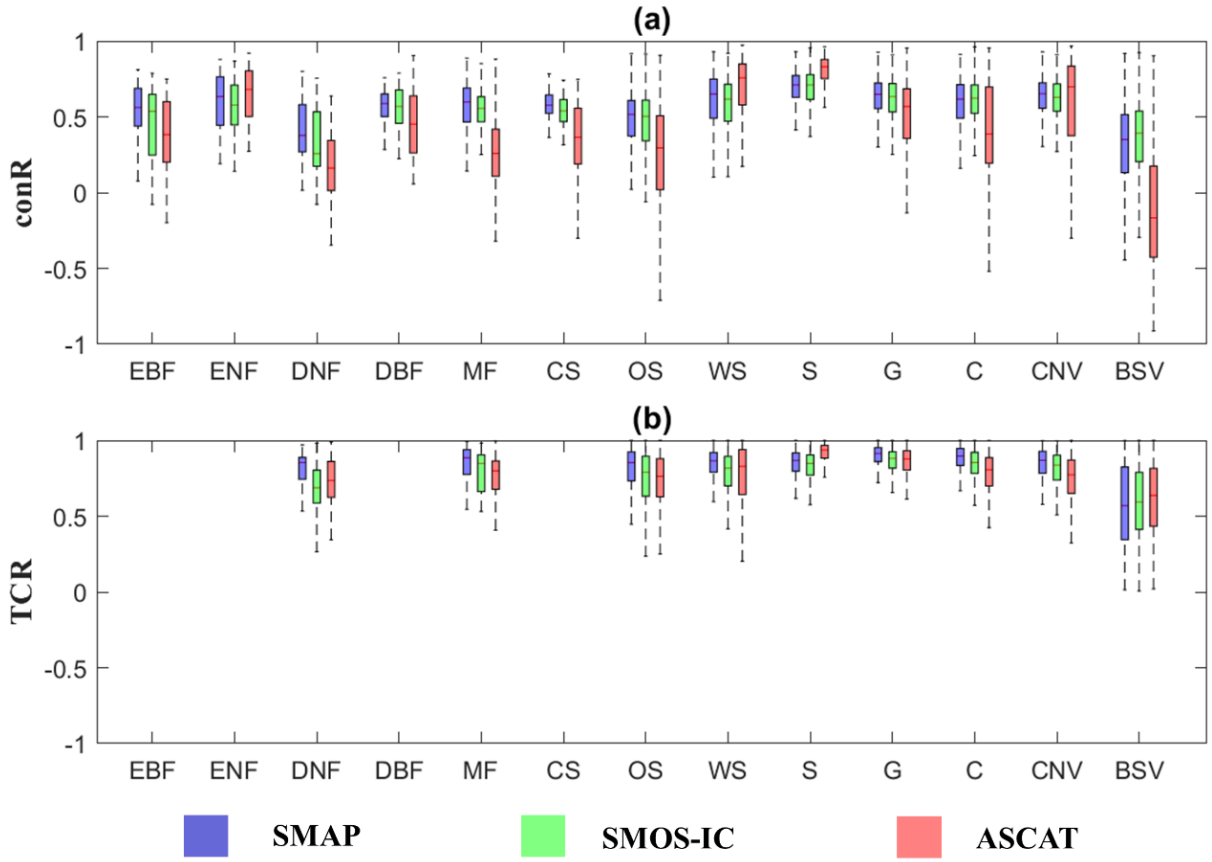


Figure S12. Boxplots for variations in the conventional R (a) against ERA-interim and TC-based R (b) under different land cover conditions. EBF (evergreen broadleaf forests); ENF (evergreen needleleaf forests); DNF (deciduous needleleaf forests); DBF (deciduous broadleaf forests); MF (mixed forests); CS (closed shrublands); OS (open shrublands) (OS); WS (woody savannas); S (savannas); G (grassland); C (croplands); CNV (cropland and natural vegetation mosaics); BSV (barren and sparsely vegetated areas).

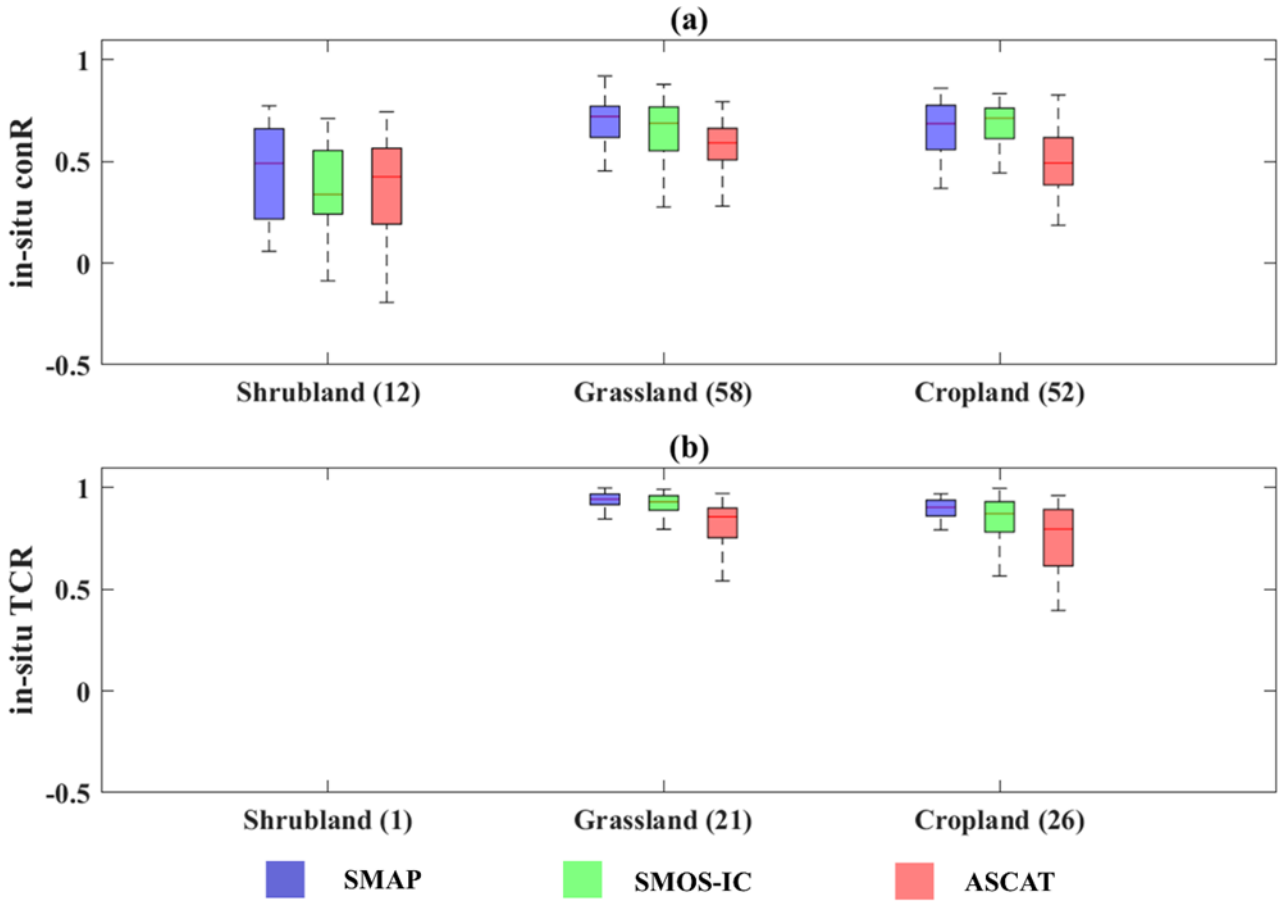


Figure S13. Boxplots for variations in the in-situ conR (a) and in-situ TCR (b) under different land cover classes. The value behind the LC class is the number of pixels included.

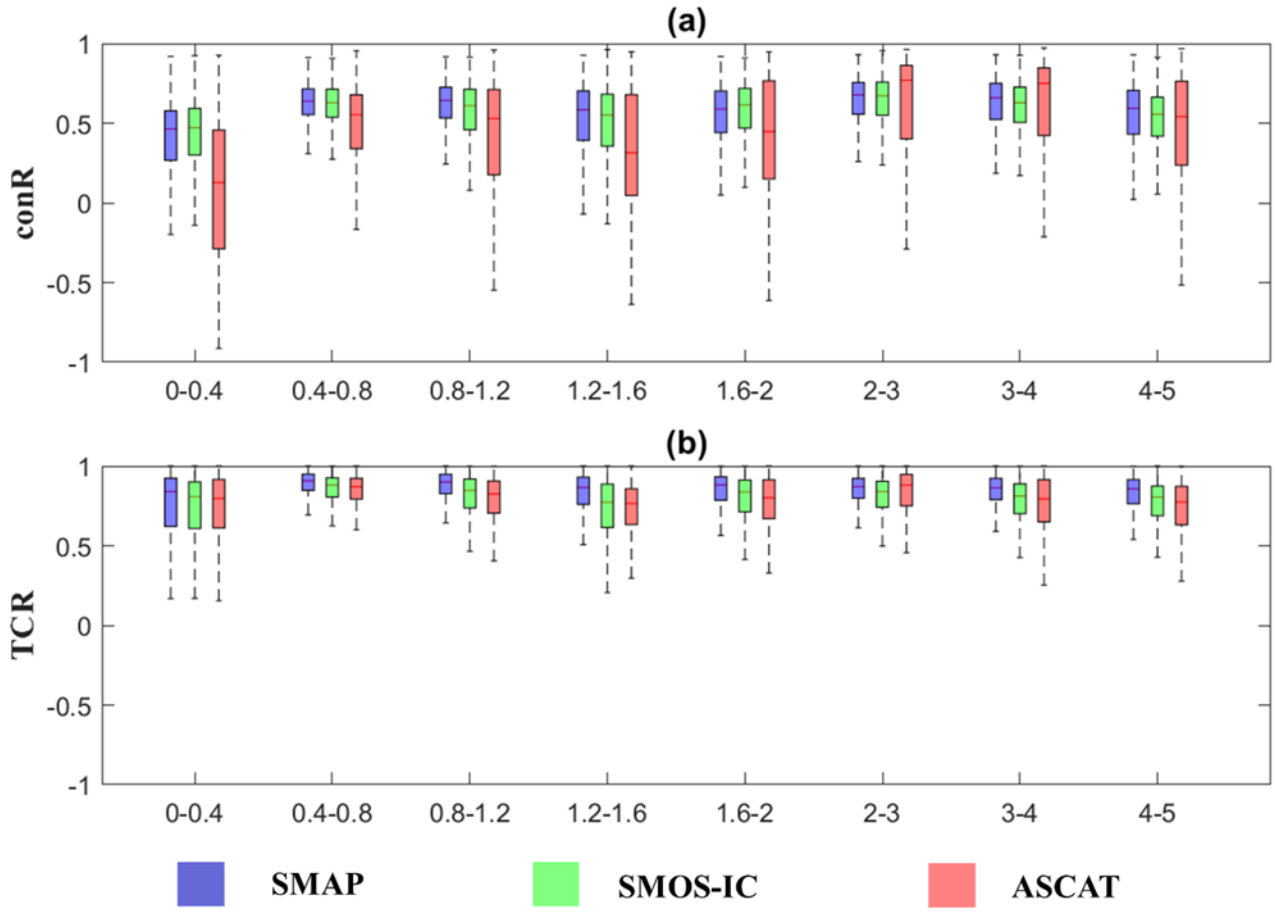


Figure S14. Boxplots for variations in the conventional R (a) against MERRA and TC-based R (b) under different mean VWC ranges in kg/m^2.

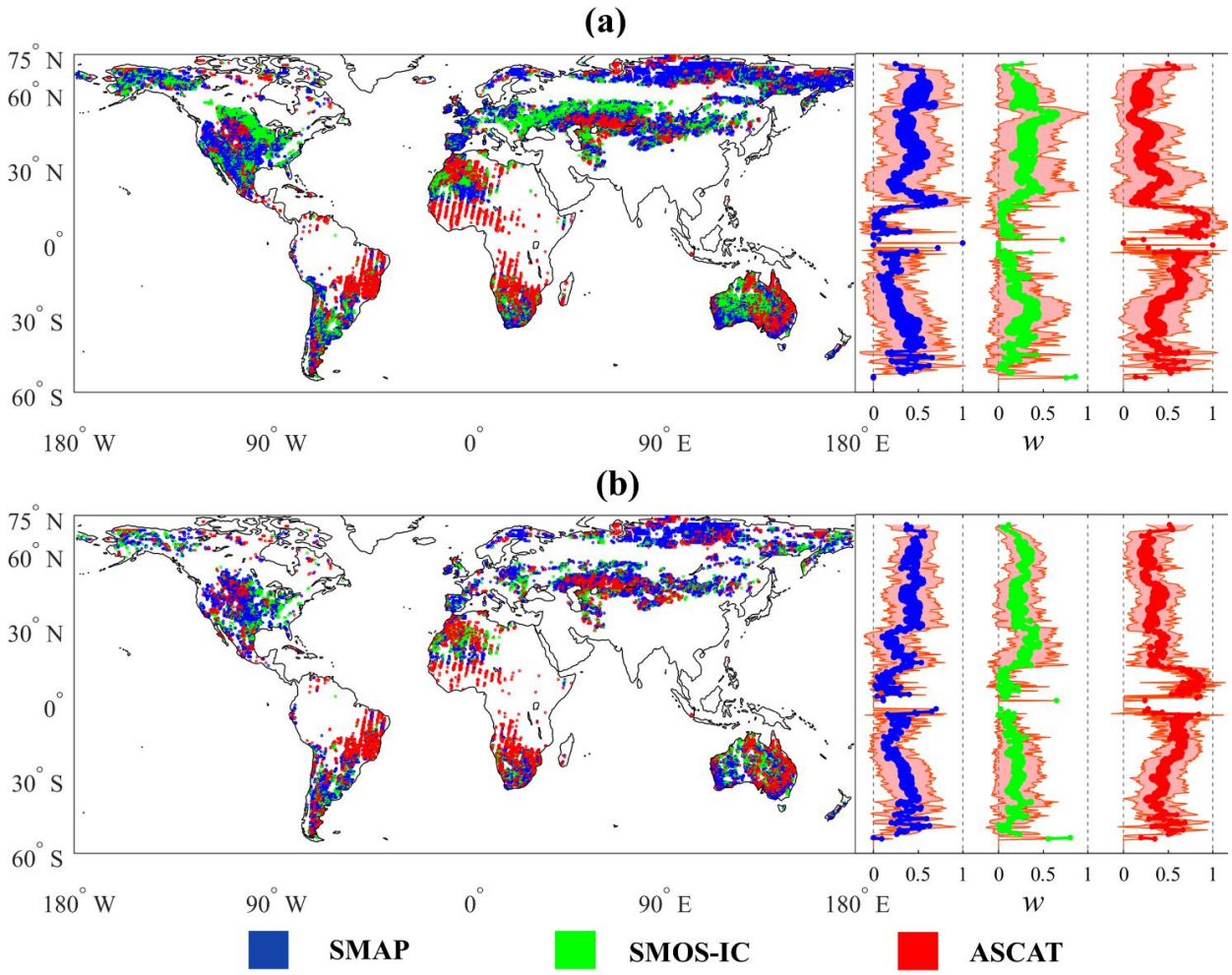


Figure S15. Global maps indicating areas exhibiting the highest optimal weight (w) for SMAP, SMOS-IC and ASCAT based on (a) w_{con} and (b) w_{TC}. The line plots in the right-hand panel present variations of zonal mean optimal weights with ±1 standard deviation.

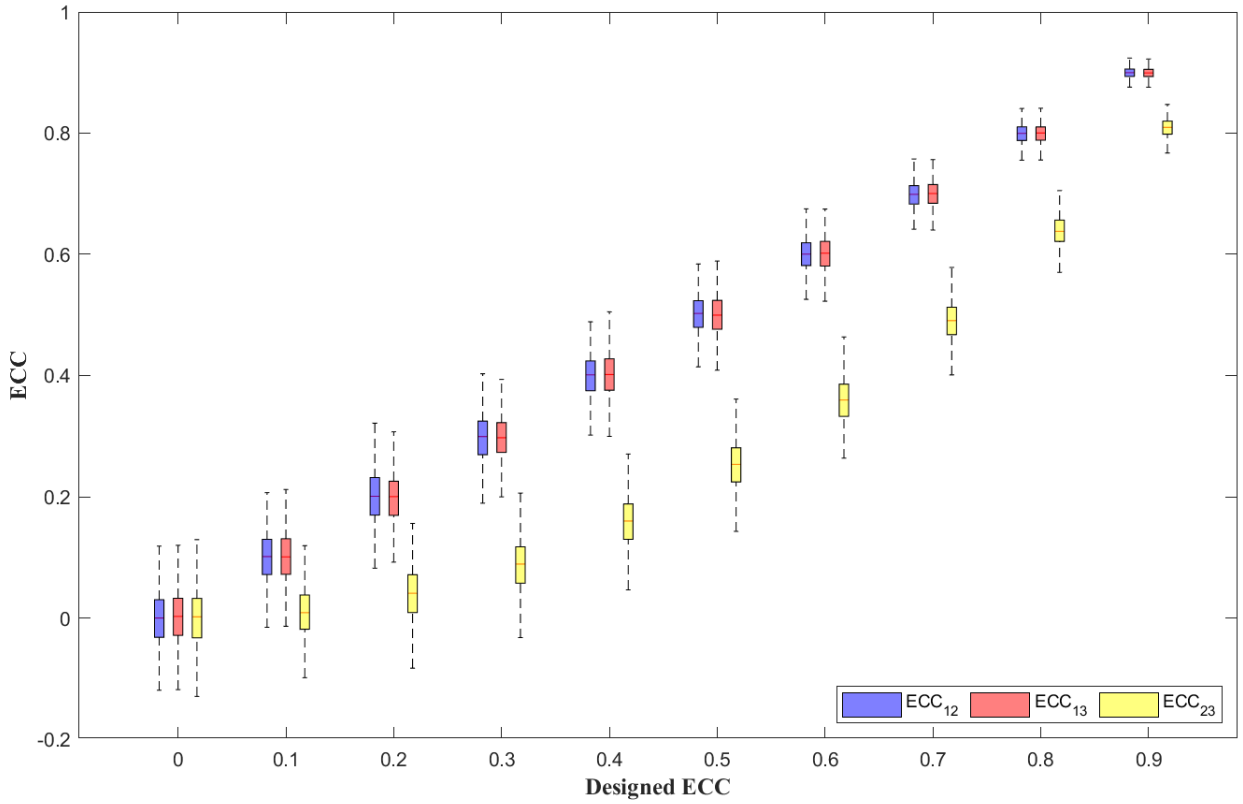


Figure S16. Boxplot of the error cross-correlations (ECC) used in the synthetic experiments, the subscript 1,2,3 represents the respective observation product. ECC_{12} and ECC_{13} are the controlled ECC increasing with a step of 0.1 and ECC_{23} is the dependent ECC.

Supplementary Tables

Table S1. Detailed information of all in-situ stations investigated in this study.

Index	Station Name	Network	ID	Latitude	Longitude	Available Source	Agreement
1	Anchor River Divide	SNOTEL	1062	59.86	-151.32	NWCC/ISMN	/
2	Aniak	SNOTEL	2065	61.58	-159.58	NWCC/ISMN	Partially Agreed
3	Atigun Pass	SNOTEL	957	68.13	-149.48	NWCC/ISMN	Partially Agreed
4	Canyon Lake	SCAN	1232	59.42	-161.16	NWCC/ISMN	/
5	Checkers Creek	SCAN	2213	65.40	-164.71	NWCC/ISMN	/
6	Chisana	SNOTEL	1093	62.07	-142.05	NWCC	/
7	Coldfoot	SNOTEL	958	67.25	-150.18	NWCC/ISMN	Agreed
8	Eagle Summit	SNOTEL	960	65.49	-145.41	NWCC/ISMN	Agreed
9	Exit Glacier	SNOTEL	1092	60.19	-149.62	NWCC/ISMN	/
10	Fielding Lake	SNOTEL	1268	63.20	-145.63	NWCC	/
11	Galena AK	SNOTEL	429	64.70	-156.72	NWCC	/
12	Gobblers Knob	SNOTEL	962	66.75	-150.67	NWCC/ISMN	Agreed
13	Granite Crk	SNOTEL	963	63.94	-145.40	NWCC/ISMN	Agreed
14	Gulkana River	SNOTEL	2222	62.41	-145.38	NWCC/ISMN	Agreed
15	Imnaviat Creek	SNOTEL	968	68.62	-149.30	NWCC/ISMN	Agreed
16	Innoko Camp	SCAN	2211	63.64	-158.03	NWCC/ISMN	Disagreed
17	Jack Wade Jct	SNOTEL	1275	64.15	-141.33	NWCC	/

18	Kanaryagak Camp	SCAN	2208	61.36	-165.12	NWCC/ISMN	/
19	Kanuti Lake	SCAN	2212	66.18	-151.74	NWCC/ISMN	/
20	Kelly Station	SNOTEL	1175	67.93	-162.28	NWCC/ISMN	/
21	Kenai Moose Pens	SNOTEL	966	60.73	-150.48	NWCC/ISMN	/
22	Little Chena Ridge	SNOTEL	947	65.12	-146.73	NWCC/ISMN	Agreed
23	Lower Mulchatna	SNOTEL	1233	59.82	-156.99	NWCC/ISMN	/
24	May Creek	SNOTEL	1096	61.35	-142.71	NWCC	/
25	McGrath	SNOTEL	785	62.95	-155.61	NWCC	/
26	Mcneil River SGS	SNOTEL	1191	59.08	-154.28	NWCC/ISMN	/
27	Monahan Flat	SNOTEL	1094	63.31	-147.65	NWCC/ISMN	Agreed
28	Monument Creek	SNOTEL	949	65.18	-145.87	NWCC/ISMN	Agreed
29	Moore Creek Bridge	SNOTEL	1176	59.59	-135.21	NWCC/ISMN	Agreed
30	Moose Inc	SCAN	2062	59.68	-151.37	NWCC	/
31	Mt. Ryan	SNOTEL	948	65.25	-146.15	NWCC/ISMN	Partially Agreed
32	Munson Ridge	SNOTEL	950	64.85	-146.21	NWCC/ISMN	Agreed
33	Naknek River	SCAN	2209	58.67	-156.57	NWCC/ISMN	/
34	Nenana	SNOTEL	2081	64.69	-149.91	NWCC/ISMN	Agreed
35	Nuka Glacier	SNOTEL	1037	59.70	-150.71	NWCC/ISMN	/
36	Port Graham	SNOTEL	987	59.35	-151.85	NWCC/ISMN	/
37	Prudhoe Bay	SNOTEL	1177	70.27	-148.57	NWCC/ISMN	/
38	Rocky Point	SNOTEL	973	64.53	-163.42	NWCC/ISMN	/
39	Schor Garden	SCAN	2063	59.68	-151.38	NWCC	/
40	Spring Creek	SNOTEL	2044	61.65	-149.13	NWCC/ISMN	Agreed
41	Summit Creek	SNOTEL	955	60.62	-149.53	NWCC/ISMN	Agreed
42	Susitna Valley High	SNOTEL	967	62.13	-150.04	NWCC/ISMN	Agreed
43	Telaquana Lake	SNOTEL	1266	60.98	-153.92	NWCC	/
44	Tok	SNOTEL	2080	63.35	-142.98	NWCC/ISMN	Disagreed
45	Tokositna Valley	SNOTEL	1089	62.63	-150.78	NWCC/ISMN	Partially Agreed
46	Unalakleet	SCAN	2221	63.91	-160.75	NWCC/ISMN	/
47	Upper Nome Creek	SNOTEL	1090	65.37	-146.59	NWCC/ISMN	Agreed
48	Upper Tsaina River	SNOTEL	1055	61.19	-145.65	NWCC/ISMN	Agreed
49	Weary Lake	SCAN	1234	59.13	-159.09	NWCC/ISMN	/
50	Hozatka Lake	SNOTEL	2210	65.20	-156.64	ISMN	/
51	Ikalukrok Creek	SCAN	/	68.08	-163.00	ISMN	/
52	Point_Mackenzie	SNOTEL	1002	61.39	-150.03	ISMN	/

Table S2. Annual R values between soil moisture retrievals from various dielectric models and in-situ measurements and the SMAP vertically polarized brightness temperature.

Dielectric Models		Mineral Soil Based Models					Organic Soil Based Models				Ground Measurements
Year	N	Wang 1980	Dobson 1985	Mironov 2009	Mironov 2013	Park 2017	Bircher 2016	Mironov 2019	Park 2019	Park 2021	in-situ
2015	9	-0.715	-0.691	-0.716	-0.711	-0.700	-0.715	**-0.780**	-0.711	-0.709	-0.591
2016	8	-0.714	-0.689	-0.714	-0.711	-0.701	-0.713	**-0.779**	-0.711	-0.704	-0.502
2017	9	-0.786	-0.763	-0.786	-0.783	-0.774	-0.786	**-0.841**	-0.784	-0.785	-0.438
2018	7	-0.768	-0.744	-0.768	-0.766	-0.756	-0.769	**-0.830**	-0.768	-0.763	-0.782
2019	3	-0.591	-0.553	-0.590	-0.588	-0.573	-0.590	**-0.695**	-0.586	-0.583	-0.373
2020	4	-0.741	-0.722	-0.741	-0.733	-0.725	-0.740	**-0.775**	-0.734	-0.725	-0.472
2021	4	-0.746	-0.721	-0.745	-0.740	-0.730	-0.747	**-0.800**	-0.745	-0.737	-0.744
Mean	/	-0.723	-0.698	-0.723	-0.719	-0.709	-0.723	**-0.786**	-0.720	-0.715	-0.557

where N represents the number of in-situ stations used for the last-column R values, and where the column of the number tagged by bold font represents the dielectric model with the highest negative correlations with the SMAP vertically polarized brightness temperature.

Table S3. Summary of simulated soil moisture products**.**

Abbreviation	Loss Regression	Observations Used During Simulation		
		Initial	Dry-down	All the observations
PLO	'LOWESS'	Yes	No	No
PLD	'LOWESS'	Yes	Yes	No
PQD	Quantile	Yes	Yes	No
PQF	Quantile	Yes	Yes	Yes

Table S4. Statistic metrics of different gap-filled products by comparing against in-situ measurements

Validation Type*/Product	Metric	PLO	PQD	PQF	DCT	LIP
Overall Validation	ubRMSE	0.08	0.07	0.06	0.06	0.06
	R	0.46	0.59	0.63	0.65	0.65
SMAP-non-synchronous Validation	ubRMSE	0.09	0.07	0.06	0.06	0.06
	R	0.46	0.59	0.63	0.67	0.66
SMAP-synchronous Validation	ubRMSE	0.08	0.07	0.06	0.06	0.06
	R	0.45	0.59	0.64	0.64	0.64

**Overall, SMAP-non-synchronous, and SMAP-synchronous validation metrics are computed using all the paired data, all the paired simulations non-synchronous with the SMAP observations, and all the paired simulation synchronous with the SMAP observations across the five gap-filled datasets, respectively.*