Journal of Hydrology 580 (2020) 124351

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental U.S

Qiangqiang Yuan[a,b], Hongzhang Xu[a,b], Tongwen Li[c], Huanfeng Shen[b,c,d,*], Liangpei Zhang[b,e]

[a] School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, Hubei, China
[b] The Collaborative Innovation Center for Geospatial Technology, Wuhan 430079, Hubei, China
[c] School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, Hubei, China
[d] The Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, Wuhan 430079, Hubei, China
[e] The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, Hubei, China

ARTICLE INFO

This manuscript was handled by Emmanouil Anagnostou, Editor-in-Chief, with the assistance of Xinyi Shen, Associate Editor

Keywords:
Soil moisture
Data fusion
GRNN
Microwave remote sensing
SMAP
Triple collocation

ABSTRACT

This study attempted to develop a point-surface collaborative inversion (PSCI) method for the estimation of regional surface soil moisture (SSM) using a generalized regression neural network (GRNN) trained on sparse ground-based measurements. Specifically, GRNN was employed to establish a nonlinear relationship between ground-based measurements from sparse network stations (SNSs) and passive microwave observations from the Soil Moisture Active Passive (SMAP) satellite in the continental U.S. for April 2015 to March 2018. More importantly, the extended triple collocation (ETC) technique was applied to address the scale mismatch issue resulting from the small spatial support of ground-based measurements, whereby individual SNSs' reliability at the SMAP coarse footprint scale could be determined before fed into GRNN. The 10-fold cross-validation results showed that the GRNN model trained on reliable SNSs obtained a satisfactory performance—the cross-validated R and unbiased RMSE values were 0.88 and 0.050 cm^3 cm^{-3}, respectively, which outperformed the back-propagation neural network (BPNN) and the other GRNN model trained on all SNSs. Furthermore, temporal and spatial comparisons between the PSCI-based SSM retrievals and other SSM datasets indicated the former agreed the best with ground measurements both in time and space, suggesting the proposed PSCI method had shown great potential in estimating reliable regional SSM climate records.

1. Introduction

Soil moisture is a key climate variable in the global water cycle budget and influences the partitioning of water and energy fluxes between the terrestrial surface and the atmosphere (Entekhabi et al., 1996). An improved observational understanding of soil moisture will therefore contribute to better monitoring of natural hazards, such as drought (Trenberth and Guillemot, 1995) and flooding (Viterbo and Betts, 1999), and give us a better idea of the plant growth (Williams and Albertson, 2004), crop yield (Rosenzweig et al., 2002), future climate change (Seneviratne et al., 2010), etc.

Microwave remote sensing provides a unique opportunity to estimate regional/global-scale surface soil moisture (SSM hereinafter) with satisfactory temporal coverage, thanks to the all-weather and all-time sensing ability of microwave electromagnetic waves (Njoku and Entekhabi, 1996; Wigneron et al., 1998). In particular, various microwave sensors onboard satellites, including both passive (e.g., the Advanced Microwave Scanning Radiometer - Earth observing system (AMSR-E) onboard the Aqua satellite, and its successor, AMSR-2 onboard the Global Change Observation Mission for Water (GCOM-W) satellite) and active (e.g., the Advanced SCATterometer (ASCAT) onboard the series of Meteorological Operational (MetOp) satellites), have been intensively exploited to estimate SSM in the past four decades. Nevertheless, the first dedicated SSM mission—Soil Moisture and Ocean Salinity (SMOS)—was not launched until 2009 by the European Space Agency (ESA), then followed the Soil Moisture Active Passive (SMAP), the second and the most recent SSM-dedicated mission launched by the National Aeronautics and Space Administration (NASA) in 2015 (Entekhabi et al., 2010a; Kerr et al., 2010). By carrying the optimal L-band instruments for sensing the moisture content for the top ~5 cm

* Corresponding author at: School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, Hubei, China.
E-mail addresses: qqyuan@sgg.whu.edu.cn (Q. Yuan), zbh531@whu.edu.cn (H. Xu), litw@whu.edu.cn (T. Li), shenhf@whu.edu.cn (H. Shen), zlp62@whu.edu.cn (L. Zhang).

https://doi.org/10.1016/j.jhydrol.2019.124351
Received 14 May 2019; Received in revised form 20 September 2019; Accepted 10 November 2019
Available online 12 November 2019
0022-1694/

depth of soil (i.e., SSM), these two missions are devoted to producing long-term and global SSM products with the accuracy requirement of 0.04 cm^3 cm^{-3} volumetric soil moisture unbiased root-mean-square error (ubRMSE) (Entekhabi et al., 2010a; Kerr et al., 2016).

Traditionally, the methodology used to produce the official SMOS and SMAP SSM products is through the inversion of radiative transfer models (RTMs) (Kerr et al., 2012; O'Neill et al., 2015). The RTM describes the thermal emission from the terrestrial surface, linking surface soil moisture to satellite brightness temperature observations. However, this emission process is affected by a large number of interacting factors (e.g., soil texture, surface roughness, topography, temperature, and vegetation coverage). The nonlinearity of this process also poses technical difficulties in quantifying the complex physical relationships. Hence, in order to minimize the uncertainties and enhance the retrieval accuracy from remote sensing data, artificial neural networks (ANNs) have been widely employed for implementing the soil moisture retrieval algorithms over the past two decades (Aires et al., 2005; Chai et al., 2009; Kolassa et al., 2013, 2016, 2018; Rodriguez-Fernandez et al., 2015; Santi et al., 2016). ANNs present several advantages over RTM inversions, including the potential in mapping any kind of input–output relationship without an explicit parameterization of physical relationships, the ability to handle numerous data from nonlinear systems, and the flexibility in combining information from different sources.

In general, ANN-based algorithms for the retrieval of large-scale SSM are implemented through using global land surface model (LSM) simulations as references to train ANNs, given LSMs are able to provide soil moisture estimates at the desired temporal and spatial resolution (Bierkens et al., 2015). Several studies have utilized this approach to retrieve SSM at global scale from either passive or active instruments or a combination of the two (Aires et al., 2005; Jiménez et al., 2013; Kolassa et al., 2013). Recently, the global retrieval of SSM for both SMOS and SMAP using ANNs are accomplished by Rodriguez-Fernandez et al. (2015) and Kolassa et al. (2018), taking the well-known European Centre for Medium-Range Weather Forecasts (ECMWF) model and the Goddard Earth Observing System version 5 (GEOS-5) model as references to train the algorithms, respectively. In addition, there are also studies focusing on using ANNs to test the a priori merging of active and passive instruments, such as ASCAT and AMSR-E (Kolassa et al., 2016, 2017). It was found that such synergy strategy could give much better results than the a posteriori merging of the individual retrievals from each sensor.

However, the problem to train an ANN with LSM simulations is that every LSM is affected by significant uncertainties due to the errors of model parameter estimates, since soil processes are extremely difficult to parameterize (Vereecken et al., 2016). Hence, SSM retrievals using LSM simulations as references are prone to suffering from uncertainties too and should be treated with caution. In contrast, high-accuracy ground-based soil moisture can be measured directly using in-situ devices inserted into the soil at various depths. Although a great number of large-scale ($> 100^2$ km^2) soil moisture networks (Vinnikov et al., 1999) have been set up over the globe in the last few decades, they are primarily served for the calibration/validation of satellite SSM products as the site distribution is too sparse to produce soil moisture maps at regional to global scales. Then arises the question: is there any possibility to use ground-based measurements instead of LSM simulations as references to train an ANN for the large-scale SSM retrieval?

To our knowledge, few studies have investigated the potential of training an ANN with ground-based measurements for the regional/global retrieval of SSM. For instance, (Rodríguez-Fernández et al., 2017) preliminarily used ground-based measurements from three sparse soil moisture networks to train an ANN for the continental-scale retrieval of SSM from SMOS observations in North America. In addition, (Xu et al., 2018) proposed a multi-source data fusion method based on the generalized regression neural network. They showed that it is effective to use this approach for the quality improvement of SMAP retrievals by training the SMAP Level 3 radiometer SSM product with ground-based measurements from five sparse networks in the western continental U.S. However, these studies didn't try to address the spatial-scale mismatch issue, the biggest problem to train a neural network with ground-based measurements, resulting from the spatial representation difference between the point support (~1 m^2) of in-situ probes and large footprint (tens of km^2) of microwave remote sensing sensors. Certainly, this can add huge uncertainty to the training process of neural networks that takes satellite observations as input and ground-based measurements as references.

Solving the (spatial-)scale mismatch issue between in-situ and remotely-sensed data remains a major challenge. In response to this problem, previous studies (Crow et al., 2012; Famiglietti et al., 2008) have been earlier characterized the behavior of SSM variability across spatial scales and analyzed the upscaling errors from the sparse ground-based measurements for the validation of the coarse-resolution satellite-based SSM products. The triple collocation (TC) technique (Stoffelen, 1998) was found more appropriate as an upscaling approach capable of compensating for the impact of random measurement error on ground-based measurements, after comparing it with other sophisticated SSM upscaling strategies, such as the time stability approach and the block kriging algorithm (Crow et al., 2012). Furthermore, Chen et al. (2017) first applied the extended version of TC (ETC) method (McColl et al., 2014) to assess the spatial representativeness of the sparse ground-based measurements within the satellite footprint for the validation of the SMAP Level-2 radiometer SSM product.

Another point worth noting is that in most of the above publications that use ANNs to retrieve large-scale SSM, the back-propagation neural network (BPNN) configuration is proposed. However, in spite of the popular utilization of BPNN as shown by these studies, it has several limitations such as slow convergence and easily being trapped by a local minimum instead of finding the global minimum error surface (Yu, 1992). In addition, BPNN needs too much human manipulation as to the determination of the number of hidden layers and neurons. An alternative to BPNN that avoids above weakness is the generalized regression neural network (GRNN). GRNN is basically an associative memory feed-forward type of ANN (Al-Mahasneh et al., 2018), which possesses many advantages: the network is able to learn from the training data by "one-pass" and does not require an iterative procedure; there is only one free parameter (called the "spread" parameter) in the network; GRNN estimation is always able to converge to a global solution. Numerous studies have demonstrated the superiority of GRNN over BPNN (Cigizoglu and Alp, 2006; del Rosario Martinez-Blanco et al., 2016; Konate et al., 2015; Li et al., 2017b; Xu et al., 2018). This makes GRNN one of the most popular networks in environmental monitoring from remote sensing data, such as evapotranspiration modeling (KIŞI, 2006), fine particulate matter ($PM_{2.5}$) estimation (Li et al., 2017b), forest cover estimation (Boyd et al., 2002), solar radiation derivation (Şenkal, 2010), SSM retrieval (Ozerdem et al., 2017; Xu et al., 2018), etc.

Inspired by these facts above, in this study we develop a novel "point-surface collaborative inversion (PSCI)" method for estimating regional SSM from SMAP brightness temperature (TB) observations and other auxiliary data, using ground-based measurements from sparse network stations (SNSs) as references to train GRNN in the continental U.S. for April 2015 to March 2018. To address the scale mismatch issue, the ETC technique is applied to select reliable SNSs before the GRNN training, thereby the established TB–SSM relationship can become more representative. The best trained GRNN model is then implemented through the proposed PSCI process to produce regional SSM maps.

The rest of this paper is organized as follows. Section 2 summarized the characteristics of CONUS, and datasets used for the development of the proposed PSCI method as well as for the construction of the ETC triplet. Section 3 describes the ETC method for the selection of reliable SNSs and the PSCI procedure for SSM estimation based on GRNN. Section 4 demonstrates the results of the selected reliable SNSs using

ETC, the performance of the GRNN model validated through the 10-fold cross-validation technique, and the spatial and temporal evaluation results of the PSCI-based SSM retrievals. Section 5 provides a summary and puts forward an outlook for future developments.

2. Study area and data

2.1. Study area

The study area of our research is the Continental U.S. (CONUS). The CONUS, which borders both the North Atlantic and North Pacific Oceans and is bordered by Canada and Mexico, has a varied topography. The eastern regions consist of hills and low mountains while the central interior is a vast plain (called the Great Plains region), and the west has high rugged mountain ranges. There are three major mountain ranges in CONUS that run north to south: the Appalachian Mountains located in the east, the Rocky Mountains in the central west, and the Sierra Nevada & Cascade Mountains in the west. Interspersed throughout are the Great Lakes in the northeast, the Mississippi River in the Midwest, the Great Salt Lake in the west, and the Grand Canyon in the southwest. Like its topography, the climate of CONUS also varies depending on location. It is considered mostly temperate but is tropical in the southeastern regions (e.g., Florida), semiarid in the plains west of the Mississippi River and arid in the Great Basin of the southwest. The Pacific Northwest is one of the wettest parts of CONUS and is densely forested. The Rocky Mountains, Sierra Nevada & Cascades have typical highland climates and are also heavily forested.

There are several reasons why we chose CONUS as the study area. To begin with, large-scale and long-term soil moisture networks, such as the Soil Climate Analysis Network (SCAN) and the U.S. Climate Reference Network (USCRN), are spread across the whole CONUS (Fig. 1). Hence, enough ground samples that overlap with satellite-based data sets can be obtained to perform the subsequent collaborative inversion process. In addition, all soil moisture sites are evenly distributed in this region, though a relatively higher density appears in the rugged west. Moreover, operational soil moisture networks using new ground measurement technologies (namely the COsmic-ray Soil Moisture Observing System (COSMOS) network and the PBO H2O network) that can provide larger spatial support of soil moisture measurements than typical point measurements derived from in-situ devices, are also deployed in this area. Such measurements are also supposed to weaken the scale mismatch issue resulting from the small spatial support of sparse ground-based measurements to a certain degree.

2.2. SMAP brightness temperature (TB) observations

Launched on 31 January 2015, the SMAP mission is operated by NASA and is equipped with both a radiometer (passive, 1.41 GHz) and a high-resolution radar (active, 1.26 GHz), dedicated to measuring SSM at L-band, with a fixed incidence angle of 40°, a revisit time of ~3 days at the equator, and 6:00 a.m. (descending)/6:00p.m. (ascending) local equatorial overpass times (Entekhabi et al., 2010a). Although the SMAP radar has failed since 7 July 2015 (Chan et al., 2016), the SMAP radiometer remains to operate as planned. This L-band radiometer observes the earth on four different channels: horizontal (H) and vertical (V) polarization as well as the 3rd and 4th Stokes parameter. The first two channels are the primary science channels that are directly sensitive to SSM, while the 3rd and 4th Stokes parameters are used to help detect the radio frequency interference (Piepmeier et al., 2014).

The SMAP Level-3 global daily brightness temperature (TB) product (SPL3TB, version 5) (O'Neill et al., 2018), which is resampled to a global cylindrical 36-km Equal-Area Scalable Earth grid version 2 (EASEv2) (Brodzik et al., 2012), was used in this study for the estimation of SSM. The observed TB has been adjusted for the presence of water bodies and corrected for atmospheric and reflected Sky radiation contributions from SMAP Level-1C TB observations (De Lannoy et al., 2015). We utilized TB observations at both H and V polarizations only from the 6:00 a.m. descending overpass, for the reason that vertical profiles of soil and canopy temperatures are expected to be more uniform during early morning overpasses than during other times of the day (Basharinov and Shutko, 1975). Note the SPL3TB product is actually provided within the SMAP Level 3 radiometer SSM product introduced below.

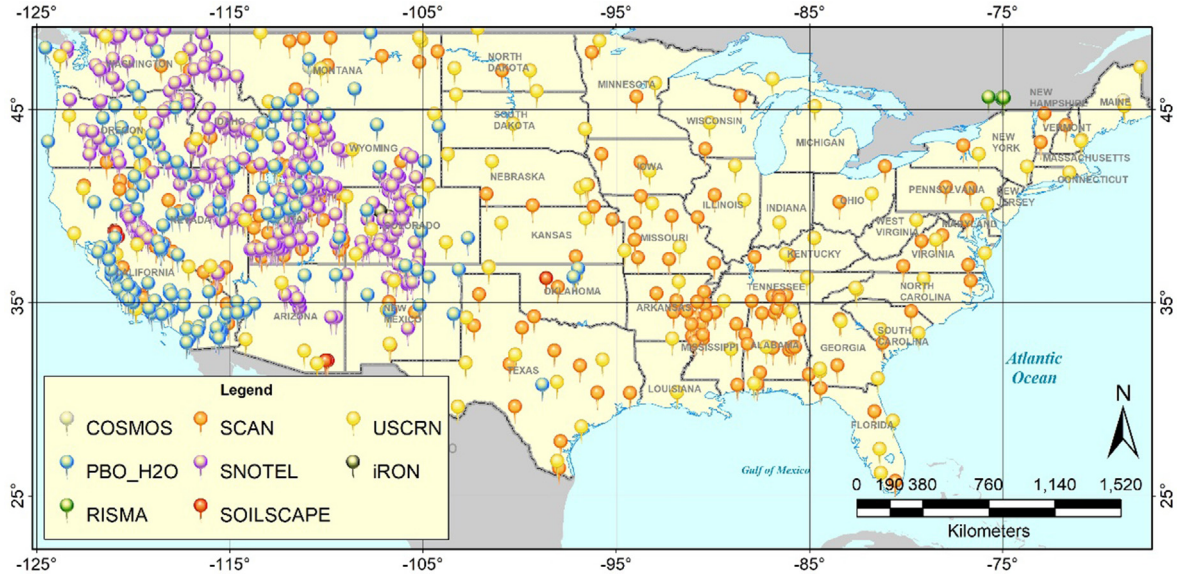

Fig. 1. Study area of CONUS and distribution of eight ground-based soil moisture networks: the COsmic-ray Soil Moisture Observing System (COSMOS) network, the interactive Roaring fork Observation Network (iRON), the PBO H2O network, the Real-time In-situ Soil Monitoring for Agriculture (RISMA) network, the Soil Climate Analysis Network (SCAN), the SNOwpack TELemetry (SNOTEL) network, the Soil moisture Sensing Controller and oPtimal Estimator (SoilSCAPE) network, the U.S. Climate Reference Network (USCRN).

2.3. Surface soil moisture (SSM) datasets

2.3.1. Satellite-based SSM retrievals from SMAP

The SMAP Level-3 radiometer global daily 36-km EASEv2-grid soil moisture (SPL3SMP, version 5), which is a composite of Level-2 half-orbit soil moisture over one day, is available since 31 March 2015 (O'Neill et al., 2018). In fact, the retrieval of SSM from SMAP TB observations occurs in the Level-2 processing, which currently uses the V-pol Single Channel Algorithm (SCA-V) as the baseline retrieval algorithm (Chan et al., 2016; Zeng et al., 2016), developed on the basis of the well-known physical tau-omega model (O'Neill et al., 2015; Wigneron et al., 1995).

The SPL3SMP product was downloaded for the 31 March 2015 to 30 March 2018 period through the NASA National Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC): https://nsidc.org/data/SPL3SMP. We only made use of SSM retrievals from the 6:00 a.m. descending overpass for the same reason as aforementioned SPL3TB. In addition, only data pixels that are "recommended for retrieval" based on the SMAP quality flag (O'Neill et al., 2015) were used.

2.3.2. Ground-based SSM measurements from sparse network sites (SNSs)

The ground-based SSM measurements used as references for the subsequent collaborative inversion process are from eight soil moisture networks distributed across CONUS: the COSMOS network (Zreda et al., 2008, 2012), the interactive Roaring fork Observation Network (iRON) (Osenga et al., 2019), the PBO H2O network (Larson et al., 2008b), the Real-time In-situ Soil Monitoring for Agriculture (RISMA) network (Ojo et al., 2015), the SCAN network (Schaefer et al., 2007), the SNOwpack TELemetry (SNOTEL) network (Leavesley et al., 2008), the Soil moisture Sensing Controller and oPtimal Estimator (SoilSCAPE) network (Moghaddam et al., 2010, 2016), and the USCRN network (Bell et al., 2013). All these networks are made up of Sparse Network Sites (SNSs) (defined as the site that typically provides just one single point measurement within a satellite footprint (Chan et al., 2016; Chen et al., 2017; Colliander et al., 2017)), thus resulting in the scale mismatch issue. It is worth noticing that measurements from COSMOS and PBO H2O are derived based on new SSM measuring technologies, which can significantly extend the spatial support of point-scale in-situ observations. The COSMOS network utilizes the so-called "cosmic-ray moisture probe", a stationary instrument that measures cosmic-ray neutrons in the air, whose intensity is inversely related to SSM, to integrates and produces area-average SSM over a footprint with a horizontal radius of ~500 m (Desilets et al., 2010). On the other hand, the PBO H2O network employs the existing geodetic Global Position System (GPS) equipment, which has been found capable to derive roughly ~1000 m^2 SSM, based on its nearly linear relationship with respect to the phase offset caused by the received multipath signals reflected from the land surface (Larson et al., 2010, 2008a). Nevertheless, the footprint made by such technologies is insignificant compared to that sensed through microwave remote sensing (tens of km^2), thus the scale mismatch issue being unavoidable and demanding to be accounted for.

A total of 921 SNSs' SSM data from aforementioned eight networks for the 31 March 2015 to 30 March 2018 period were downloaded through the International Soil Moisture Network (ISMN) website portal: https://ismn.geo.tuwien.ac.at/en/, an well-known international cooperation in the soil moisture community aiming to establish and maintain a global in-situ soil moisture database (Dorigo et al., 2011). Fig. 1 shows the spatial distribution of the SNSs in the study area and Table 1 summaries the characteristics of each network.

Only the shallowest measurements representing the moisture content of the topsoil (i.e., SSM) were considered, trying to be consistent with the sensing depth (~5 cm) of the L-band satellite. It is worth noticing that sampling depth can be different among various in-situ sensors. For example, COSMOS network measures SSM within the 0–8 cm topsoil layer, which is several centimeters deeper than the typical penetration depth of the L-band satellite data (0–5 cm), while PBO H2O network measures the identical depth of SSM to that of the L-band satellite data. Most of the other networks measure SSM at ~5 cm. Additionally, SSM measurements that are not with "good" quality based on the ISMN quality flag (Dorigo et al., 2013) were masked out. The time series of ground-based SSM measurements from each SNS are recorded at one-hour intervals, except for the PBO H2O network, which provides only one SSM estimate for each day (recorded at 12:00 UTC). For the hourly SSM time series, only measurements that are closest to the SMAP 6:00 a.m. local equatorial overpass time using a 3-hour window for each day was taken to generate a daily time record. A least of 30 daily measurements during the study time period was required for each site.

2.3.3. Model-based SSM simulations from ERA-Interim

ERA-Interim is a model-based dataset released by ECMWF, showing the results of a global climate reanalysis from 1979 to date and continuing to be updated in near-real-time (Dee et al., 2011). It uses the Tiled ECMWF Scheme for Surface Exchange over Land (H-TESSEL) land surface model (Viterbo et al., 1999; Viterbo and Beljaars, 1995) to solve for a variety of parameters including a four-layer soil profile (0–7, 7–28, 28–100, and 100–289 cm), provided every 6 h (0:00, 6:00, 12:00, and 18:00 UTC) (Berrisford et al., 2011). The data are archived and can be freely accessed at the ECMWF website portal: https://apps.ecmwf.int/datasets/data/interim-full-daily.

Here, only the upper layer (0–7 cm) soil moisture, provided on a regular 0.25° spatial resolution, was used. Simulations at each point of time (0:00, 6:00, 12:00, and 18:00 UTC) were aggregated to a daily average. They were further regridded to 36-km EASEv2 grids to be consistent with SMAP data. In addition, we applied quality control to the model-based SSM simulations by using the upper layer (0–7 cm) soil temperature from ERA-Interim to identify times and grids where the soil temperature is below 1 °C. These data were deemed not feasible to generate SSM and were masked out (Kolassa et al., 2018).

2.4. Auxiliary data

In addition to TB observations, we take advantage of another two datasets as auxiliary input to our model; they are also required as auxiliary data in the SMAP soil moisture retrieval algorithm, which include: 1) surface soil temperatures (Ts hereinafter) from the NASA's GEOS-5 land modeling system, utilized to determine the soil surface emissivity from TB observations, and 2) vegetation water contents (VWC hereinafter) based on an empirical relationship with respect to the normalized difference vegetation index from MODIS, used to correct the vegetation effects. For a detailed description of the two datasets, readers can refer to the "SMAP Level 2 & 3 Soil Moisture (Passive) Algorithm Theoretical Basis Document" (O'Neill et al., 2015).

Like the SPL3TB, Ts and VWC are also provided within the SPL3SMP product and have already been posted on the same 36-km EASEv2 grid as SMAP SSM. Note that we excluded Ts values that are below 1 °C and VWC values that are higher than 5 kg/m^2, because the soil moisture retrieval is not feasible under such conditions (O'Neill et al., 2015).

3. Methodology

3.1. Extended triple collocation (ETC) technique for the selection of reliable SNSs

In order to address the scale mismatch issue resulting from the small spatial support of sparse ground-based measurements which can add huge uncertainty to the subsequent collaborative inversion process, we turned to the triple collocation (TC) technique (Stoffelen, 1998).

TC is an analyzing tool for estimating the unknown random error variances of three collocated datasets of the same geophysical variable, without treating any one dataset as perfectly observed "truth" (Gruber et al., 2016). The basic assumptions of TC are: 1) the triplet each are

Table 1
Characteristics of the eight ground-based soil moisture networks: the COsmic-ray Soil Moisture Observing System (COSMOS) network, the interactive Roaring fork Observation Network (iRON), the Real-time In-situ Soil Monitoring for Agriculture (RISMA) network, the Soil Climate Analysis Network (SCAN), the SNOwpack TELemetry (SNOTEL) network, the Soil moisture Sensing Controller and oPtimal Estimator (SoilSCAPE) network, the U.S. Climate Reference Network (USCRN), and the PBO H2O network.

Name	# of stations	Available time	Depth (cm)	Temporal resolution	Sensor
COSMOS	3	2008/04–	0–8	Hourly	Cosmic-ray Probe
iRON	7	2012/08–2017/01	5	Hourly	EC5, EC5 I, EC5 II
RISMA	6	2013/04–2017/02	5	Hourly	Hydraprobe II Sdi-12
SCAN	179	1996/01–	5.08	Hourly	Hydra Probe Digital SDI-12 (2.5 V), Hydra Probe Analog (2.5 V)
SNOTEL	376	1980/10–	5	Hourly	Hydra Probe Analog (5.0 V), Hydra Probe Analog (2.5 V), Hydra Probe Digital SDI-12 (2.5 V)
SoilSCAPE	99	2011/08–2017/03	5	Hourly	EC5
USCRN	113	2000/11–	5	Hourly	Stevens Hydra Probe II SDI-12
PBO H2O	138	2004/09–2017/12	0–5	Daily	GPS

linearly related to the (unknown) truth, 2) the error statistics are stable and do not change over time, 3) the errors of the triplet are mutually independent of each other, and 4) are independent of the true value as well. According to the first assumption, the collocated datasets X_i can be related to the true state T via an affine error model:

$$X_i = \beta_i + \alpha_i T + \varepsilon_i \tag{1}$$

where β and α refer to the additive and multiplicative bias constants, respectively; ε is the mean-zero additive error, and the subscript i stands for the three collocated datasets of the same geophysical variable.

The objective of TC is to find a solution that individually estimates the variance of each ε_i in Eq. (1) based on the above assumptions. To this end, it is required to choose one dataset from the triplet as a reference and rescale the other two into the same reference data space. Apparently, this will result in a dependency of the error variance estimates on the climatology of the scaling reference (Draper et al., 2013). In order to address this issue, McColl et al. (2014) proposed the Extended Triple Collocation (ETC) method, which is based on exactly the same assumptions as the original TC and derives an additional performance metric, the correlation coefficient of X_i with respect to T as formulated below (Lei et al., 2015):

$$R_{ETC}(T, X_i) = \text{sign}(\pm)\sqrt{\frac{\text{Cov}(X_i, X_j)\text{Cov}(X_i, X_k)}{\text{Cov}(X_i, X_i)\text{Cov}(X_j, X_k)}} \tag{2}$$

where the sign of R_{ETC} is corrected up by assuming to be always positive.

In this study, we applied the ETC method to determine the reliability of individual SNS observations in terms of their spatial representativeness over the SPL3SMP footprint scale (i.e., 36 km EASEv2 grid) following the idea of Chen et al. (2017). First, we constructed the SSM triplet from satellite-based retrievals (i.e., SMAP), ground-based measurements (i.e., SNS), and model-based simulations (i.e., ERA-Interim), provided that they belong to different types of measuring systems and none of them are dependent on one another. Then, the $R_{ETC}(T, SNS)$ can be derived based on Eq. (2), which indicates the relation between the sparse in-situ measurements and the unknown truth within the 36-km satellite footprint scale—i.e., it determines the fraction of footprint-scale SSM dynamics captured by point-scale observations at individual SNSs. An experimental threshold of 0.7 for $R_{ETC}(T, SNS)$ was set, above which the SNSs could be deemed as "reliable". This threshold was chosen after testing a series of thresholds ranging from 0.4 to 0.9 with an interval of 0.1, which will be discussed in the Discussion section (see Section 4.4). Besides, a minimum threshold of 100 triplets was set to avoid sample impoverishment.

In addition, there also exists the representation difference problem of penetration depths among different SSM datasets (see Section 2.3) and among various in-situ sensors (see Table 1). Draper et al. (2013) had demonstrated that this would have only a marginal impact on the ETC-derived correlation coefficient.

3.2. Point-surface collaborative inversion (PSCI) method for SSM estimation

3.2.1. Generalized regression neural network (GRNN) algorithm

The GRNN is a one-pass learning algorithm first introduced by Specht (1991). The network features fast learning that does not require an iterative procedure. It is a powerful tool for regression, approximation, fitting, and prediction problems.

Assume a vector random variable, $\boldsymbol{x}$, and a scalar random variable, y, are input and output variables, respectively, and can be expressed as follows:

$$\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_p}) \tag{3}$$

$$y = (y_1, y_2, \cdots, y_n)^T \tag{4}$$

where p is the dimension of $\boldsymbol{x}$ and n is the number of observations. Assuming $\boldsymbol{X}^i$ and Y^i ($i = 1, 2, \ldots, n$) are sample values of $\boldsymbol{x}$ and y, respectively, the regression of y on an input value, $\boldsymbol{X}$, can then be estimated by GRNN as below:

$$\hat{Y}(\boldsymbol{X}) = \frac{\sum_{i=1}^{n} Y^i K(\boldsymbol{X}, \boldsymbol{X}^i)}{\sum_{i=1}^{n} K(\boldsymbol{X}, \boldsymbol{X}^i)} \tag{5}$$

$$K(\boldsymbol{X}, \boldsymbol{X}^i) = \exp(-d_i^2/2\sigma^2),\ d_i^2 = (\boldsymbol{X} - \boldsymbol{X}^i)^T(\boldsymbol{X} - \boldsymbol{X}^i) \tag{6}$$

where $K(\cdot,\cdot)$ stands for the Gaussian kernel and d_i is the Euclidean distance between $\boldsymbol{X}$ and $\boldsymbol{X}^i$. Last, σ is called the "spread" parameter, which is the only unknown parameter in the network and needs optimization (see Section 4.2.1). This free parameter affects the level of fitness in GRNN architecture. The larger the spread, the smoother the function approximation.

The architecture of GRNN presented in Fig. 2 contains four basic layers: an input layer, a pattern layer, a summation layer, and an output layer. The input layer consists of p neurons, where p is the dimension of the vector variable, $\boldsymbol{x}$ (here $p = 7$ in our study). It is completely connected to the patter layer, and provides the input measurement, $\boldsymbol{X}$, directly to the pattern units. The pattern layer has n nodes, where n is the number of training samples, and each pattern unit is assigned with a sample vector, $\boldsymbol{X}^i$, in the training data. In each node, the Euclidean distance between the input vector, $\boldsymbol{X}$, and the assigned vector, $\boldsymbol{X}^i$, is first calculated and then fed into the Gaussian kernel based on Eq. (6). The pattern unit outputs are sent to the summation units. The summation layer has two types of processing units: one is called the "Numerator" part, the other is the "Denominator" part. The input to the Numerator unit is the sum of the pattern layer outputs, each weighted by an observed output scalar, Y^i, corresponding to $\boldsymbol{X}^i$ in the training samples, while the input to the Denominator unit is the sum of the pattern unit outputs. At last, the output layer has only one neuron, which receives the two outputs from the summation units and divides the "Numerator" part by the "Denominator" part (Eq. (5)) to produce an estimate for y given $\boldsymbol{X}$. It can be seen from above that GRNN memorizes every unique

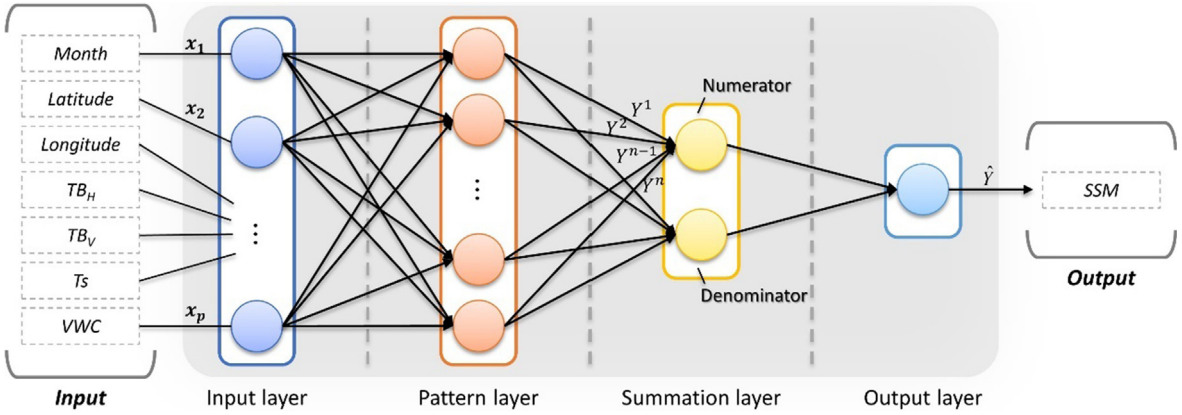

Fig. 2. Schematic of the generalized regression neural network (GRNN) used for SSM estimation.

pattern in the training samples. This is why it is a single-pass learning network and does not require any back-propagation algorithm.

In this study, the input signals include seven parameters: SPL3TB observations at both H and V polarizations (TB_H, TB_V), surface soil temperatures (Ts) from GEOS-5, vegetation water contents (VWC) based on MODIS, and coincident time (month) and location (latitude and longitude) of each SNS. The output signal is SSM. In fact, the essential function of GRNN is to determine a non-linear relationship between SSM and the primary input—TB observations (i.e., TB_H and TB_V), as TB is directly related to SSM. Here, Ts and VWC data are also taken as auxiliary input fed into GRNN, in order to add contributive information to the TB–SSM relationship. Besides, coincident month, latitude and longitude are used to account for the seasonal variability and spatial heterogeneity of SSM. Such practice has been commonly carried out in several previous studies (Li et al., 2017b; Wu et al., 2012; Yao and Lu, 2014).

3.2.2. Procedures of the proposed PSCI method for SSM estimation

The entire workflow of the methodology in this study is depicted in Fig. 3. To begin with, ETC method was applied to the triplet constructed from satellite-, ground- and model-based SSM to identify reliable SNSs. Then, TB, Ts, and VWC datasets were spatially and temporally matched using ground-based measurements only from reliable SNSs as references, to generate an input–output training data set. To be specific, the site-specific SSM data were collocated with the 36-km EASEv2 grid covering the site. If one grid corresponded more than one sites, then measurements from these sites were averaged. Afterward, the generated input–output training data set was fed into the GRNN by taking the reliable SNSs' SSM as references to train the network for the 31 March 2015 to 30 March 2018 period. Once trained, the model's performance was evaluated using cross-validation technique to identify the best one by tuning the spread parameter of GRNN to an optimum value. Finally, the best trained GRNN model was applied to the input datasets for the whole study area and the whole period to produce maps of PSCI-based SSM retrievals.

3.3. Model evaluation

The "10-fold cross-validation (CV)" technique (Rodriguez et al., 2010) is chosen to test the model's predictive power. This technique is basically a resampling procedure often used to evaluate machine learning models on a limited data sample (Li et al., 2017a,b; Ma et al., 2014). To begin with, all data samples (i.e., SSM measurements from reliable SNSs in this study) are randomly shuffled and averagely divided into ten groups. Next, each unique group is taken as a hold-out or test data set, while the remaining nine groups are taken as a training data set to fit the model. Then, repeat fitting a model on the training set and evaluating it on the test set for every group. At last, evaluation scores from all ten rounds will be summarized and averaged to show how accurately the predictive model performs. The results could also tell whether the model is over-fitted—i.e., the model performs better on the training set than on the test set.

Four classic statistical metrics in the soil moisture community were adopted to give a quantitative assessment of the model performance, i.e., the Pearson correlation coefficient (R, unitless), the root-mean-square error (RMSE, $cm^3\ cm^{-3}$), the bias ($cm^3\ cm^{-3}$), and unbiased RMSE (ubRMSE, $cm^3\ cm^{-3}$). Detailed descriptions of these indicators are provided in (Entekhabi et al., 2010b).

4. Results and discussion

4.1. Selection of reliable SNSs

The number of reliable and unreliable SNSs for each soil moisture network at the threshold of 0.7 for the ETC-derived correlation coefficient is summarized in Table 2. In general, just 372 are considered reliable, accounting for ~40% of 921 SNSs in total. To be specific, both COSMOS and iRON remain no site, which is more surprising for COSMOS since it measures larger spatial support of SSM than typical point-scale in-situ observations. Perhaps this is because there exist quite a few COSMOS measurements whose qualities are not "good" based on the ISMN quality flags, thus being masked out in our research. That is to say, not enough COSMOS measurements overlap with other SSM datasets at the threshold of 100 triplets for the ETC implementation. For iRON, this phenomenon is also blamed to the poor qualities of its measurements, as well as its relatively short data available period (August 2012 to January 2017, see Table 1) with respect the study period (March 2015 to March 2018). In contrast, PBO H2O, the other network apart from COSMOS providing large spatial support ground-based measurements, remains more than half of its sites, which demonstrated the promising utility of applying these new ground measurement technologies in the soil moisture community. The RISMA network kept the biggest proportion (~83%) of SNSs—only one site considered unreliable, whereas the SNOTEL network left the largest number of reliable SNSs, though only accounting for ~28% of its sites.

The distribution of the selected reliable SNSs based on the ETC-derived correlation coefficient in excess of 0.7 from each soil moisture network is presented in Fig. 4. It appears from the figure that the number of reliable SNSs is able to cover most of CONUS with many clustering in the western part. However, there are few or nearly no reliable SNSs in the northeastern states of CONUS as well as in the southeast coast, which may be due to the existence of the Appalachian Mountains along with the small amount of SNSs in the east (see Fig. 1). In addition, it makes sense that the sites neighboring the sea or lakes are mostly screened, since the retrieval of satellite-based SSM may be suspectable or unsuccessful due to the presence of significant water

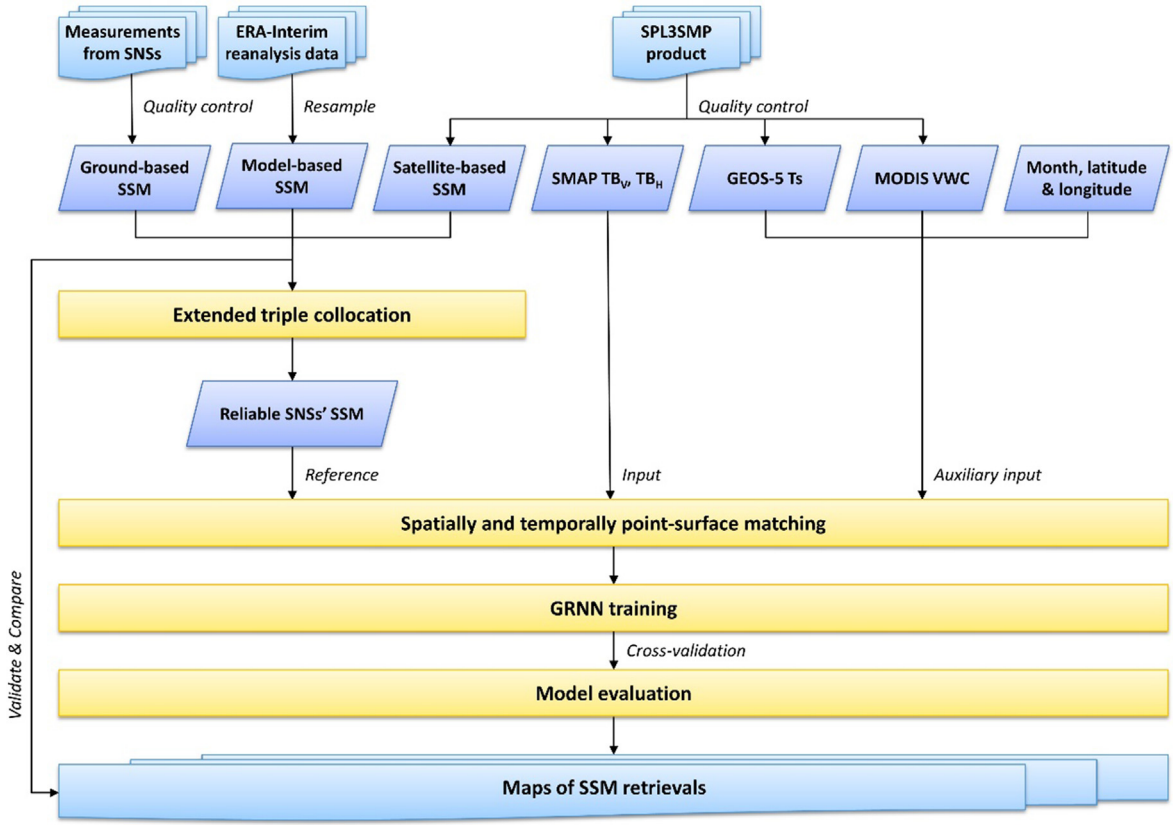

Fig. 3. Entire workflow of proposed point-surface collaborative inversion (PSCI) method for SSM estimation.

Table 2
Summary of the reliability of each network.

Networks	Reliable	Unreliable
COSMOS	0	3
iRON	0	7
PBO H2O	73	65
RISMA	5	1
SCAN	86	93
SNOTEL	104	272
SoilSCAPE	56	43
USCRN	48	65

bodies within the satellite sensing footprint.

4.2. Assessment of the GRNN model

4.2.1. Model settings

As mentioned in Section 3.2.1, the "spread" parameter is the only unknown parameter in the GRNN architecture; therefore its optimum value requires to be identified prior to the PSCI process for SSM estimation. This parameter affects the level of fitness in GRNN architecture. Large values of the spread parameter tend to force the estimation to be smooth, while lower values provide a closer approximation to the sample values (Specht, 1991). In general, the spread parameter must be greater than 0 and can usually range from 0.01 to 1 with good results (del Rosario Martinez-Blanco et al., 2016).

Here, the 10-fold CV technique is utilized to select an appropriate value for the spread parameter. A series of the spread parameter values ranging from 0.001 to 1 with the interval of 0.001 were tested to give the cross-validated performance of the GRNN model. Additionally, in order to demonstrate the effect of using ETC to address the scale mismatch issue, two GRNN models were chosen for comparison. One is the GRNN trained on only reliable SNSs, the other is the one trained on all SNSs in the study area (i.e., without ETC implementation). The variations of CV-derived ubRMSE statistics against the spread parameter of the two models are depicted in Fig. 5. It can be seen that the minimum ubRMSE for these two GRNN models both appear at the spread value of 0.011. Besides, it is apparent that the performance of the GRNN trained on reliable SNSs is much better than the one trained on all SNSs, which suggests the added value of the ETC implementation for the selection of reliable SNSs. To sum up, the spread parameter is set as 0.011 for each model.

4.2.2. Algorithm for comparison

In order to give a comprehensive analysis of the GRNN model used in this study, the popular back-propagation neural network (BPNN) as mentioned in the Introduction Section was also chosen for comparison. Here, a three-layer, fully-connected feedforward architecture was constructed, with one single hidden layer of the tangent-sigmoid transfer function, a classical back-propagation training algorithm (Rumelhart et al., 1995) and a Levenberg–Marquardt approach (Levenberg, 1944; Marquardt, 1963). Input and output parameters of BPNN are identical with those in GRNN employed in this study. Likewise, two BPNN models that are trained on reliable and all SNSs, respectively, were taken into consideration. According to previous works (Gardner and Dorling, 1998; Li et al., 2017b; Reich et al., 1999; Xu et al., 2018), the optimum number of neurons in the hidden layer can be determined based on the range from $2\sqrt{n+\mu}$ to $2n+1$, where n and μ are the number of neurons in the input and output layer. Therefore, by testing a series of hidden layer neurons ranging from 6 to 16, 15 neurons (which performed the best using 10-fold CV, not shown) were chosen for the BPNN model trained on reliable SNSs, while for the one trained on all SNSs, 16 neurons were selected.

4.2.3. Overall performance of models

Both GRNN and BPNN models were implemented through the same

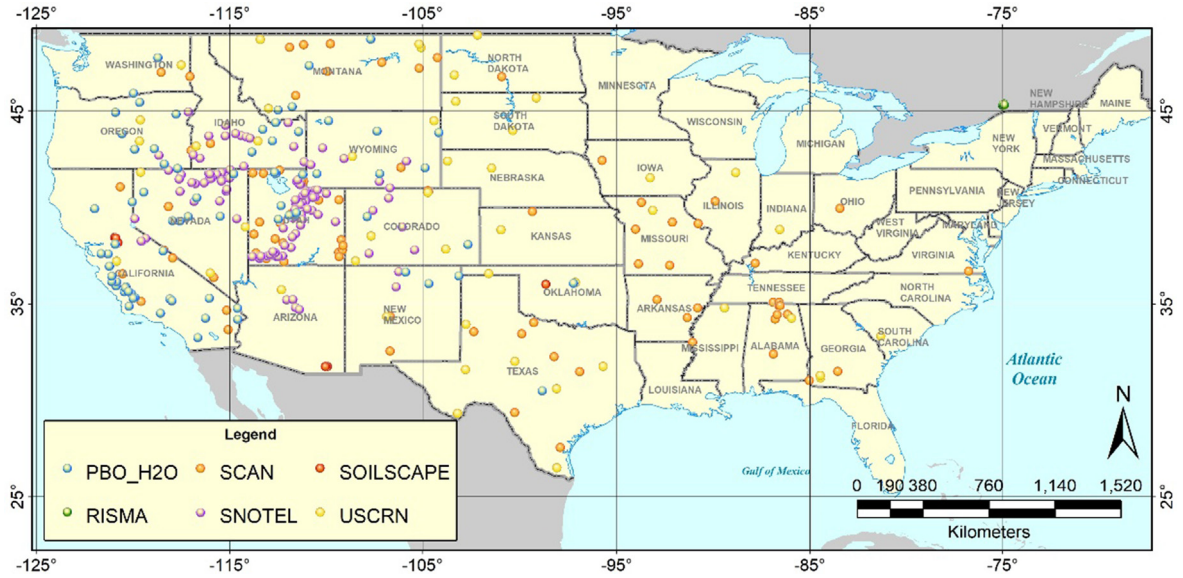

Fig. 4. The distribution of reliable sparse network stations (SNSs) from six soil moisture networks.

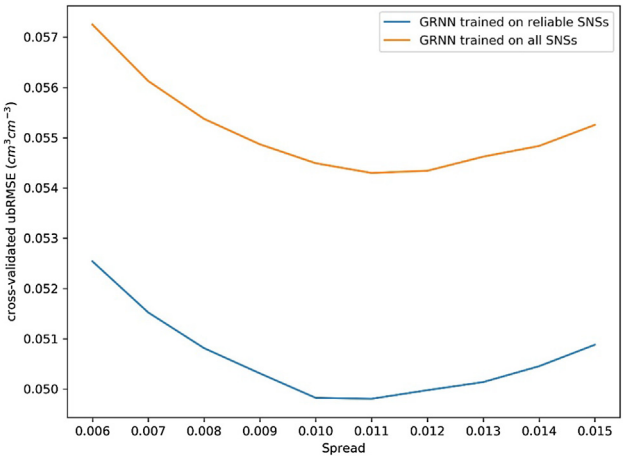

Fig. 5. The effect of spread parameter on GRNN performance trained on reliable (in blue) and all SNSs (in orange) based on 10-fold cross-validation technique.

evaluation process using the 10-fold CV, and their obtained performances are summarized in Table 3. Note in the table the model fitting results are from the best fitting model over the 10 rounds of the 10-fold CV, while the cross-validation results are the 10-round combined. From the table, it shows that GRNN obtains much better performance than BPNN, with cross-validated R and ubRMSE values improved by 0.13 (0.08 $cm^3\ cm^{-3}$) and 0.021 (0.013 $cm^3\ cm^{-3}$) for the model trained on all (reliable) SNSs, respectively. When comparing between the results using all and only reliable SNSs' measurements, the models trained on reliable SNSs significantly outperform those trained on all SNSs, indicating the added value of the ETC implementation for the selection of reliable SNSs. Additionally, all CV-derived biases are close to zero, demonstrating the unbiased estimation of the models. Further, Fig. 6 shows the scatter plots of model fitting and cross-validation results for the GRNN model trained on reliable SNSs. From model fitting to cross-validation, R score drops by 0.09 (from 0.97 to 0.88), which demonstrates that slight over-fitting is presented in the proposed model. However, with the highest cross-validated R (0.88) and lowest cross-validated ubRMSE (0.050 $cm^3\ cm^{-3}$), the GRNN model trained on reliable SNSs outperforms the other models. Therefore, despite its slight over-fitting, the GRNN model trained on reliable SNSs (with the spread parameter of 0.011) is considered as the best-trained model and is selected for the subsequent process.

4.2.4. *The predictive power of GRNN over each SNS/network*

In order to give a comprehensive assessment of the predictive power of the GRNN model, the evaluation metrics between SSM estimates using the best trained GRNN model (hereinafter "PSCI-SSM" for short) and ground-based measurements were calculated over each SNS, including both reliable and unreliable sites. For comparison, SPL3SMP is also chosen to compute its metrics against ground-based measurements over each SNS. The spatial distribution of the calculated R and ubRMSE for individual SNSs are depicted in Fig. 7 and Fig. 8, respectively. In both figures, the redder the points, the better the evaluation metrics.

Generally, the number of red points for both R and ubRMSE against ground-based measurements has increased a lot for PSCI-SSM with regard to SPL3SMP. According to the statistics, the number of sites with an R value greater than 0.70 is increased from 372 to 484, whereas the number of sites with ubRMSE score smaller than 0.04 $cm^3\ cm^{-3}$ is more than doubled from 141 to 346 out of the total 921 SNSs. From a spatial point of view, for SPL3SMP, high R values in excess of 0.70 are mainly clustered in the northwestern part of CONUS and sparsely scattered in

Table 3
The 10-fold cross-validation performances of GRNN models compared to BPNN.

Model	Used SNSs	Model fitting (N = 88059)				Cross-validation (N = 97843)			
		R	RMSE	bias	ubRMSE	R	RMSE	bias	ubRMSE
BPNN	All	0.75	0.074	0.000	0.074	0.74	0.075	0.000	0.075
	Reliable	0.82	0.060	0.000	0.060	0.80	0.063	0.000	0.063
GRNN	All	0.97	0.028	0.000	0.028	0.87	0.054	0.000	0.054
	Reliable	0.97	0.024	0.000	0.024	0.88	0.050	0.000	0.050

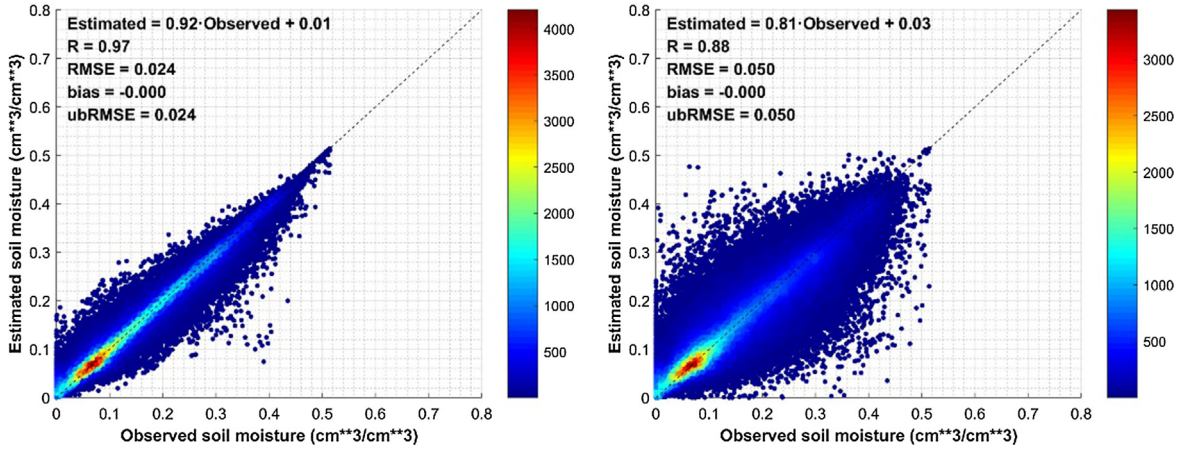

Fig. 6. Scatter plots of (a) model fitting and (b) cross-validation results for the GRNN trained on reliable SNSs. The dashed line is the 1:1 line.

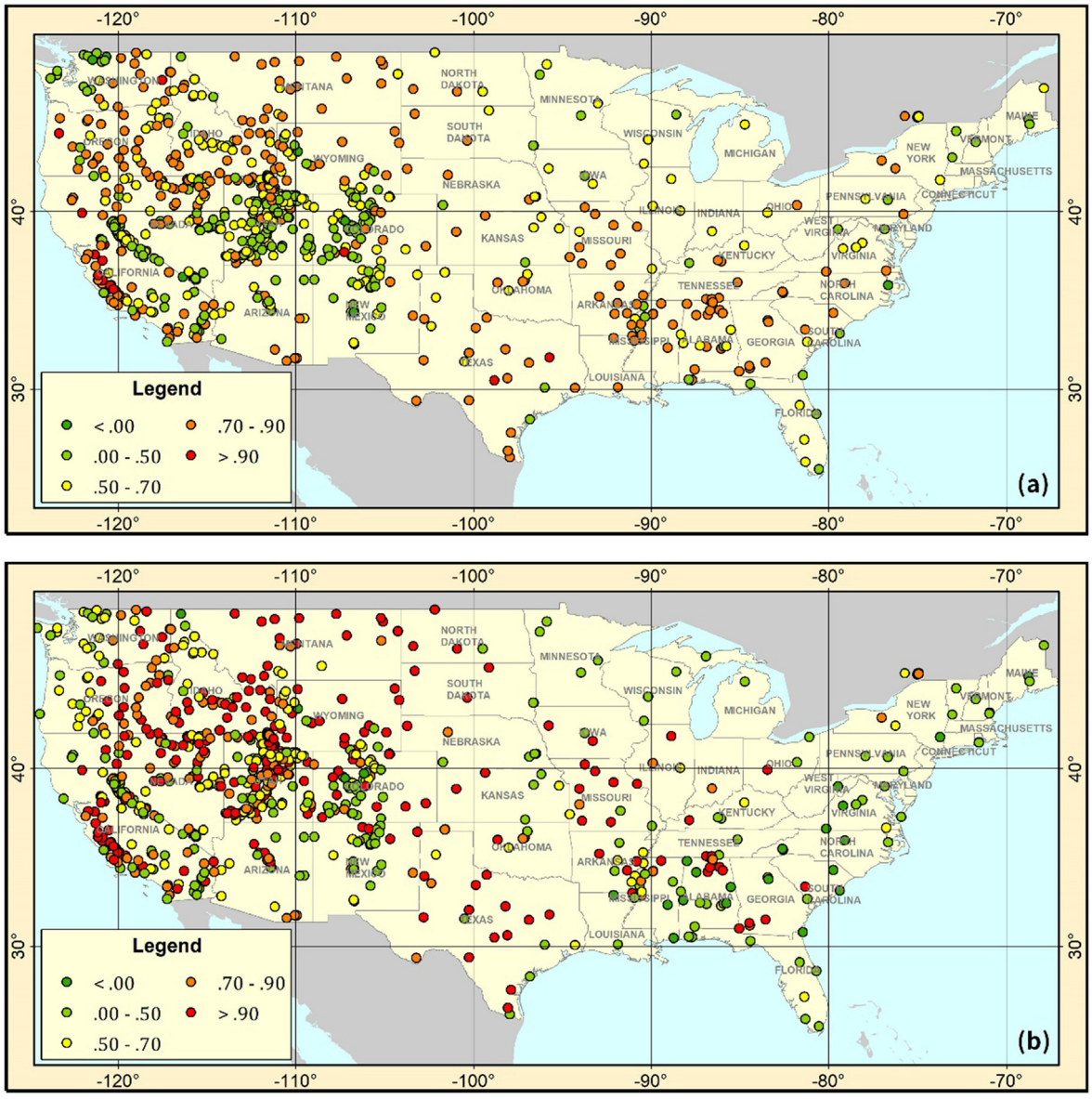

Fig. 7. Spatial distribution of R (unitless) of (a) SPL3SMP and (b) PSCI-SSM against ground-based measurements from all SNSs, respectively.

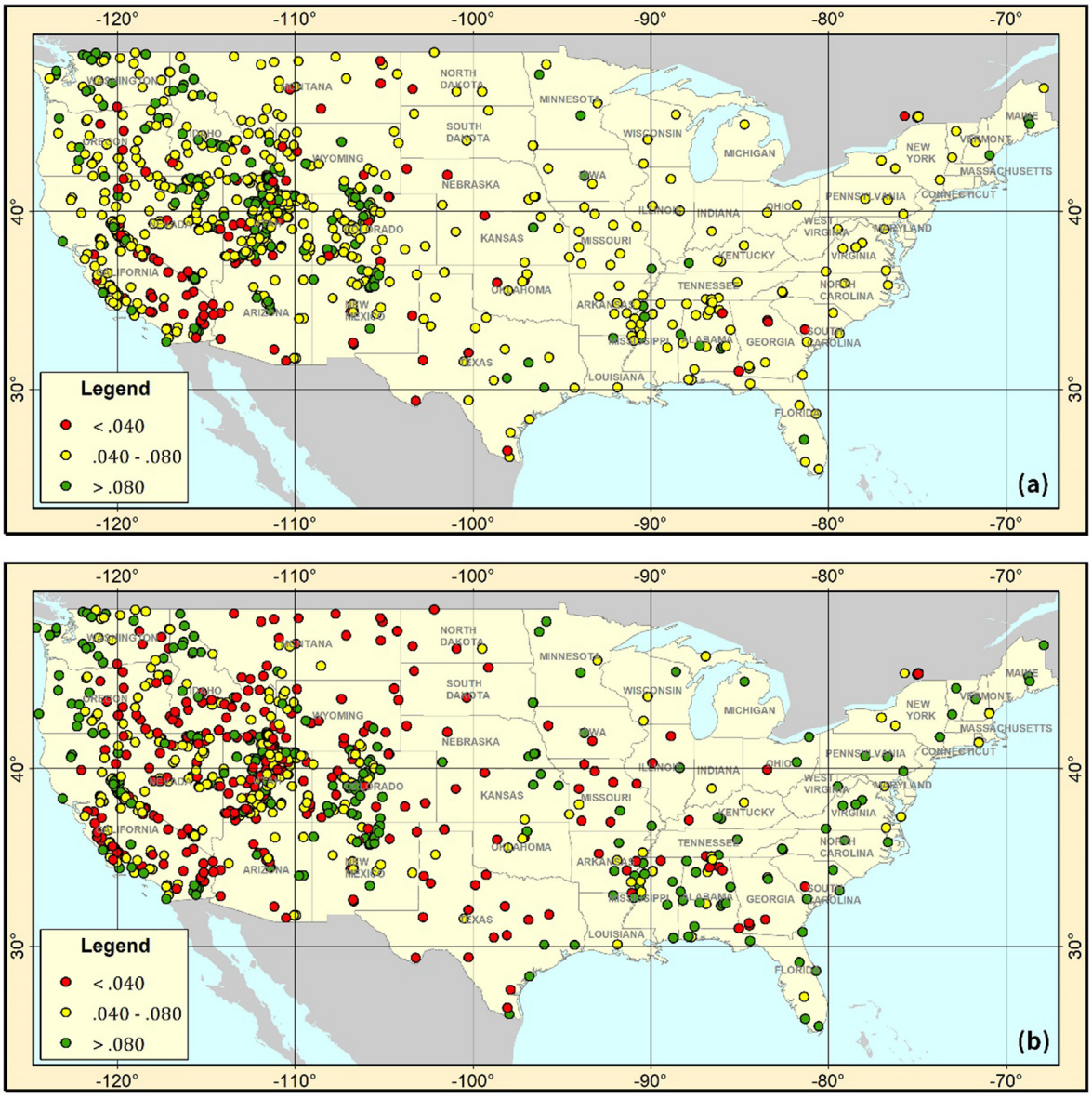

Fig. 8. Spatial distribution of ubRMSE ($cm^3\ cm^{-3}$) of (a) SPL3SMP and (b) PSCI-SSM against ground-based measurements from all SNSs, respectively.

Table 4
Average statistics of the evaluation of PSCI-SSM and SPL3SMP against ground-based measurements over each network.

Network	SPL3SMP				Proposed PSCI			
	R	RMSE	bias	ubRMSE	R	RMSE	bias	ubRMSE
COSMOS	0.49	0.104	0.048	0.069	0.75	0.082	−0.029	0.065
iRON	0.50	0.098	−0.038	0.057	0.46	0.091	−0.042	0.063
PBO H2O	0.70	0.068	−0.007	0.051	0.73	0.058	0.009	0.048
RISMA	0.75	0.075	0.042	0.048	0.84	0.055	0.017	0.039
SCAN	0.64	0.091	0.032	0.056	0.66	0.074	0.011	0.056
SNOTEL	0.52	0.115	−0.019	0.077	0.69	0.095	−0.021	0.067
SoilSCAPE	0.77	0.087	0.022	0.048	0.82	0.064	−0.017	0.039
USCRN	0.68	0.106	0.044	0.052	0.63	0.087	0.022	0.056

the mid to mid-east, while for PSCI-SSM, there are more red points gathering in the mid to western regions, and high R values over 0.90 are further improved. Additionally, the majority of ubRMSE values for SPL3SMP is at the intermediate level (0.040 to 0.080 $cm^3\ cm^{-3}$), while for PSCI-SSM, a great number of ubRMSE values less than 0.040 $cm^3\ cm^{-3}$ appear. The above consequences indicate a generally satisfactory predictive power of the GRNN model. However, it is also worth noticing that in the northeastern states and along the southeast coast, both R and ubRMSE against ground-based measurements obtain little improvement for PSCI-SSM with respect to SPL3SMP, which is supposed to have something to do with the lack of reliable SNSs in these regions (see Fig. 4). This suggests a further refinement of our approach regarding a

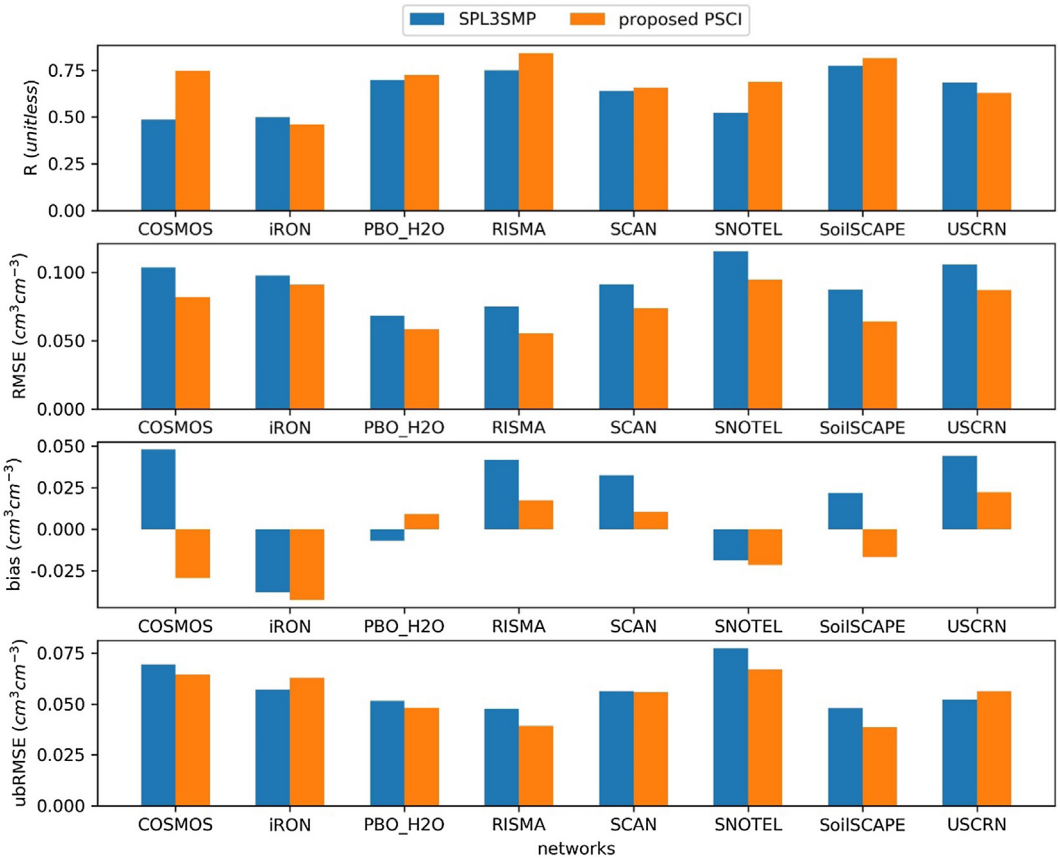

Fig. 9. Bar charts showing the average performance of PSCI-SSM and SPL3SMP against ground-based measurements over each network.

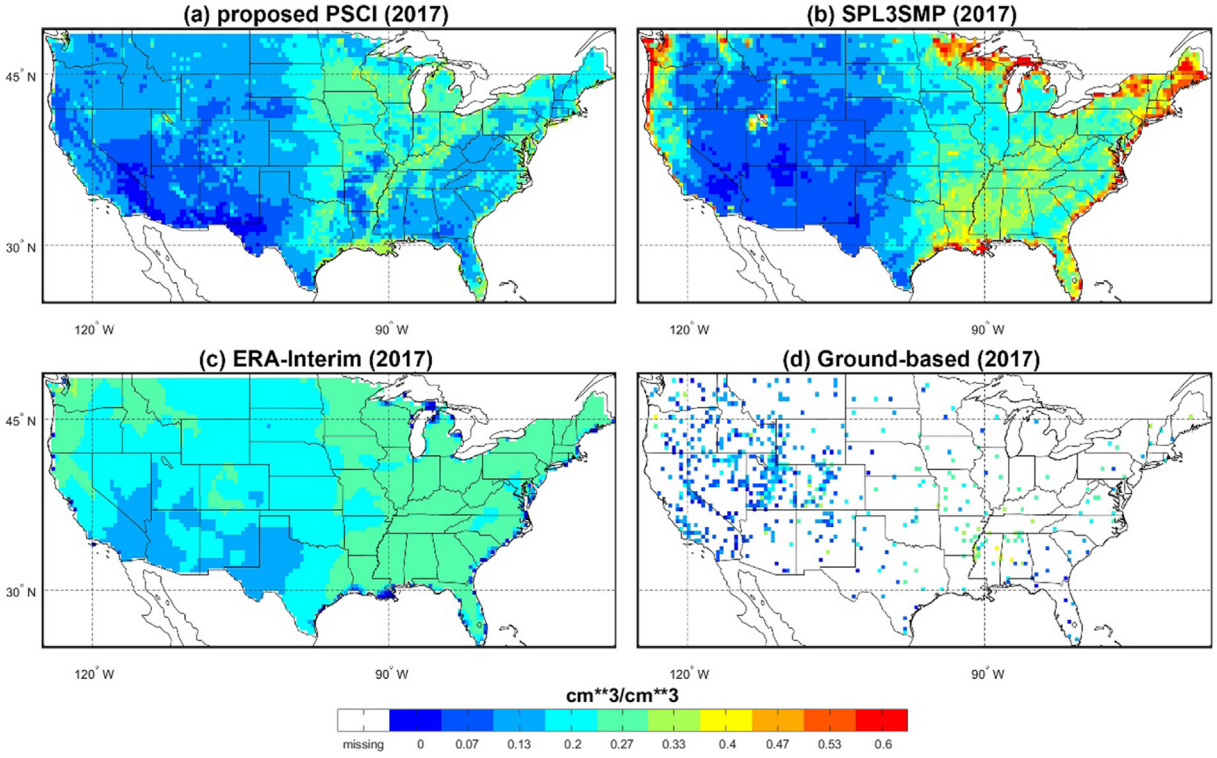

Fig. 10. Annual mean SSM map of the year 2017 for (a) the proposed PSCI-based SSM retrievals (PSCI-SSM), (b) the official SMAP Level 3 radiometer SSM product (SPL3SMP), (c) ERA-Interim model simulations, and (d) ground-based measurements from all SNSs.

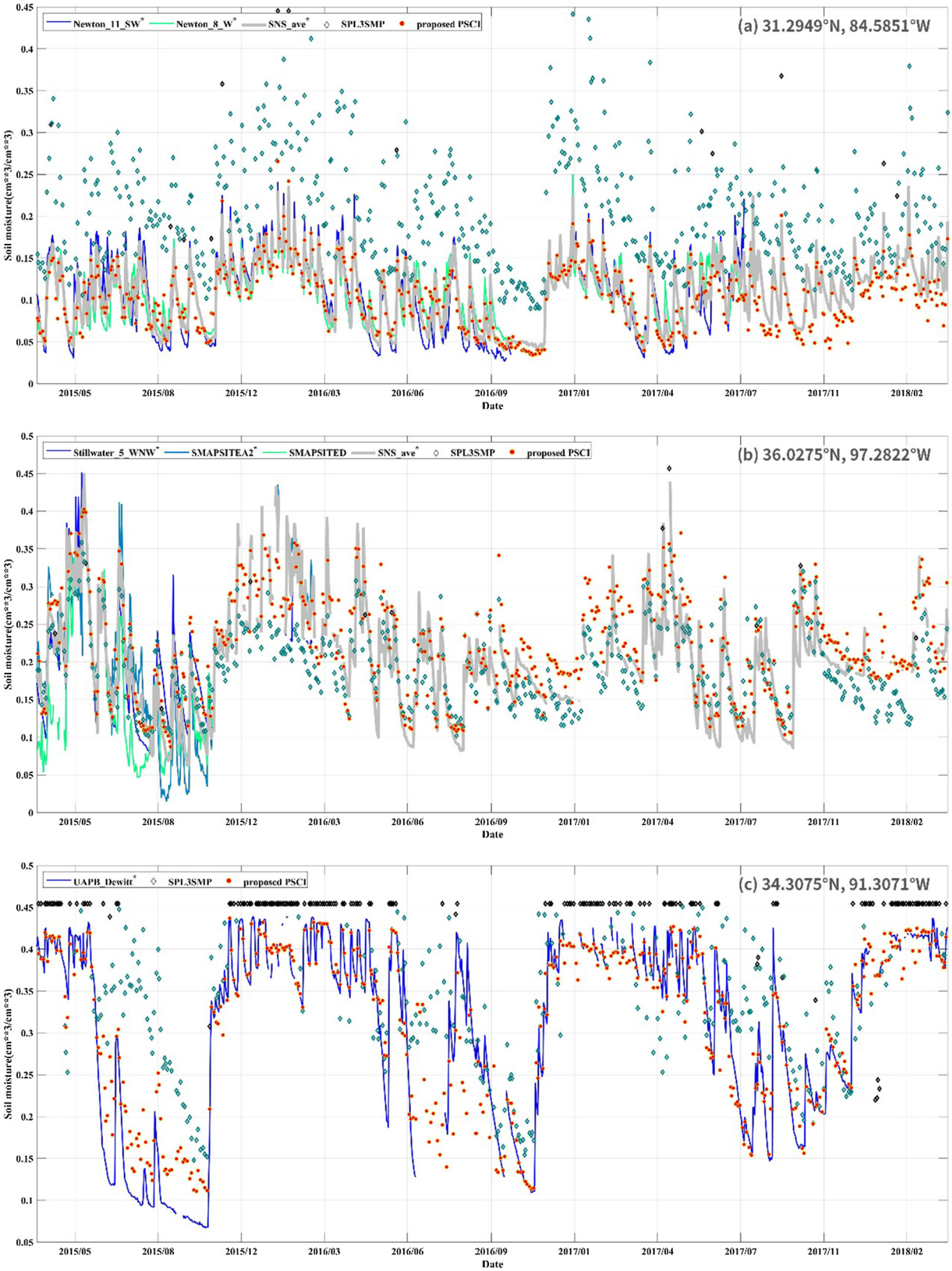

Fig. 11. The coincident SPL3SMP and PSCI-SSM time series compared with ground-based SSM over four example of 36-km EASEv2 grids, covering (a) two reliable SNSs (Newton_11_SW, Newton_8_W), (b) two reliable SNSs (Stillwatere_5_WNW, SMAPSITEA2), and one unreliable SNS (SMAPSITED), (c) one reliable SNS (UAPB_Dewitt), and (d) two reliable SNSs (Hodges and Gadsden_19_N), respectively. The reliable SNSs are marked with an asterisk. Note "SNS-ave" denotes averaged measurements from only reliable SNSs within the grid. The dark cyan hollow diamonds indicate "recommended" retrievals from SPL3SMP while the black ones refer to "non-recommended" values.

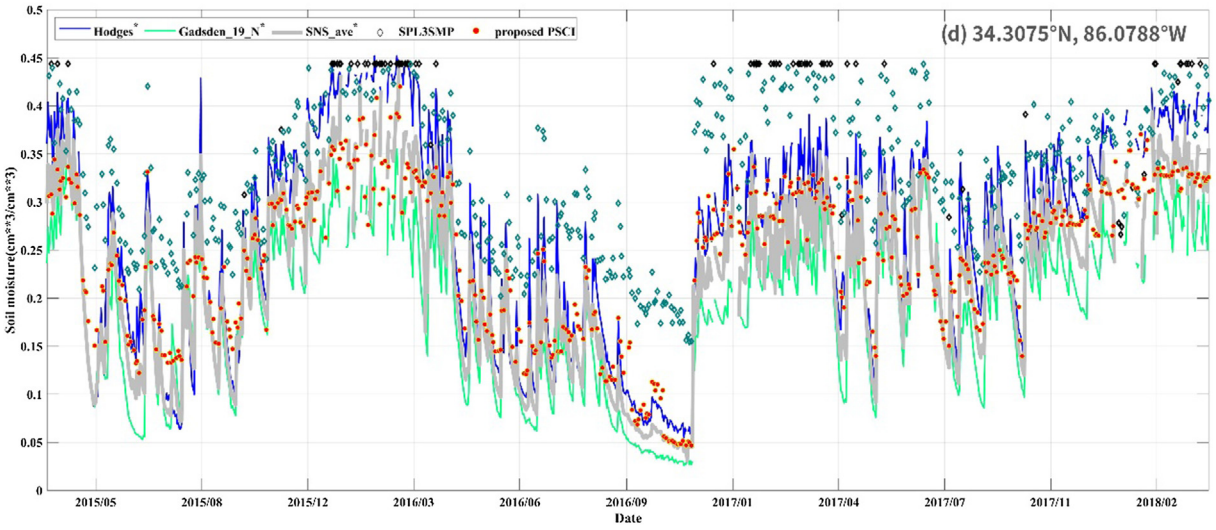

Fig. 11. (*continued*)

Table 5
Average statistics of the evaluation of SSM retrievals using GRNN models trained on ERA-Interim simulations ("ERA-PSCI-SSM") and reliable SNSs ("SNS-PSCI-SSM") against ground-based measurements over each network.

Network	ERA-PSCI-SSM				SNS-PSCI-SSM			
	R	RMSE	bias	ubRMSE	R	RMSE	bias	ubRMSE
COSMOS	0.38	0.109	−0.004	0.069	0.75	0.082	−0.029	0.065
iRON	0.43	0.085	0.038	0.058	0.46	0.091	−0.042	0.063
PBO H2O	0.66	0.109	0.087	0.058	0.73	0.058	0.009	0.048
RISMA	0.70	0.071	0.016	0.055	0.84	0.055	0.017	0.039
SCAN	0.55	0.112	0.063	0.059	0.66	0.074	0.011	0.056
SNOTEL	0.57	0.118	0.062	0.074	0.69	0.095	−0.021	0.067
SoilSCAPE	0.75	0.119	0.091	0.047	0.82	0.064	−0.017	0.039
USCRN	0.57	0.117	0.072	0.057	0.63	0.087	0.022	0.056

flexible selection of reliable SNSs in the regions with a limited quantity of ground stations.

Further, the quantitative evaluation metrics over each SNS were averaged on the basis of individual soil moisture networks for PSCI-SSM and SPL3SMP, respectively, in order to get a general idea of the predictive power of GRNN over each network (Table 4 and Fig. 9). Note that the mean values were only computed when a site has a p-value < 0.05. It can be seen from Table 4 and Fig. 9 that PSCI-SSM outperforms the SPL3SMP over most of the soil moisture networks. Specifically, PSCI-SSM obtains higher R values than SPL3SMP over six soil moisture networks out of the total eight networks, except for iRON and USCRN. However, the R differences over iRON and USCRN are very small—only 0.04 and 0.05 for each, respectively. The most significant difference for R is observed over COSMOS where the mean R value has increased by 0.26 for PSCI-SSM with regard to SPL3SMP. As for the mean RMSE, all networks have experienced a decline from SPL3SMP to PSCI-SSM, with the biggest decrease obtained over SoilSCAPE from 0.087 $cm^3\ cm^{-3}$ to 0.064 $cm^3\ cm^{-3}$. Likewise, ubRMSE scores of PSCI-SSM outperforms SPL3SMP over six networks. All these results suggest that the GRNN model can generally predict accurate SSM estimates.

4.2.5. Evaluation of the PSCI-based SSM retrievals

Based on the above results, the best trained GRNN model is applied to the SPL3TB observations and other auxiliary input for the whole study area and the whole study period, thereby maps of SSM retrievals based on the proposed PSCI method are generated. Fig. 10 displays the annual mean SSM map of the year 2017 for PSCI-SSM, in comparison with the aforementioned SSM triplet—the SPL3SMP product, ERA-Interim simulations, and ground-based measurements. Note the ground-based SSM measurements are from all SNSs and used as the benchmark (Fig. 10(d)).

In general, all three surface products (i.e., PSCI-SSM, SPL3SMP, and ERA-Interim) show low SSM in the west, with an increase toward the east, which agrees with the topography across CONUS as described in Section 2.1–the west is more arid than the east. Among them is PSCI-SSM that reveals the most similar spatial pattern with ground-based SSM than both SPL3SMP and ERA-Interim. Besides, areas neighboring the sea or lakes generally have high SSM. However, SPL3SMP shows extremely high SSM at the regions adjacent to water bodies, indicating these retrievals are questionable or of bad quality due to the presence of water bodies within the 36-km satellite footprint. Additionally, ERA-Interim displays a too smooth spatial pattern in general, which is difficult to capture the spatial variations of SSM. The two issues of SPL3SMP and ERA-Interim are overcome in PSCI-SSM, suggesting our proposed PSCI method performs comparatively well in space.

Furthermore, in order to deepen our understanding of the temporal variations of our PSCI-SSM product, coincident time series from all SSM datasets except for ERA-Interim (for brevity) were inter-compared for each common 36-km EASEv2 grid. Four examples of such grids are illustrated in Fig. 11. Note that the "non-recommended" retrievals in the SPL3SMP product are also displayed for a more comprehensive comparison.

From the figure, a generally better agreement in temporal variations can be seen between PSCI-SSM and site-specific SSM time series than SPL3SMP. The absolute values of PSCI-SSM are much closer to site-specific time series than SPL3SMP. On one hand, "unrecommended" retrievals in SPL3SMP time series (e.g. Fig. 11(c) and (d)) are profoundly corrected in PSCI-SSM. On the other hand, the "recommended"

Fig. 12. Bar charts showing the average performance of ERA-PSCI-SSM and SNS-PSCI-SSM against ground-based measurements over each network.

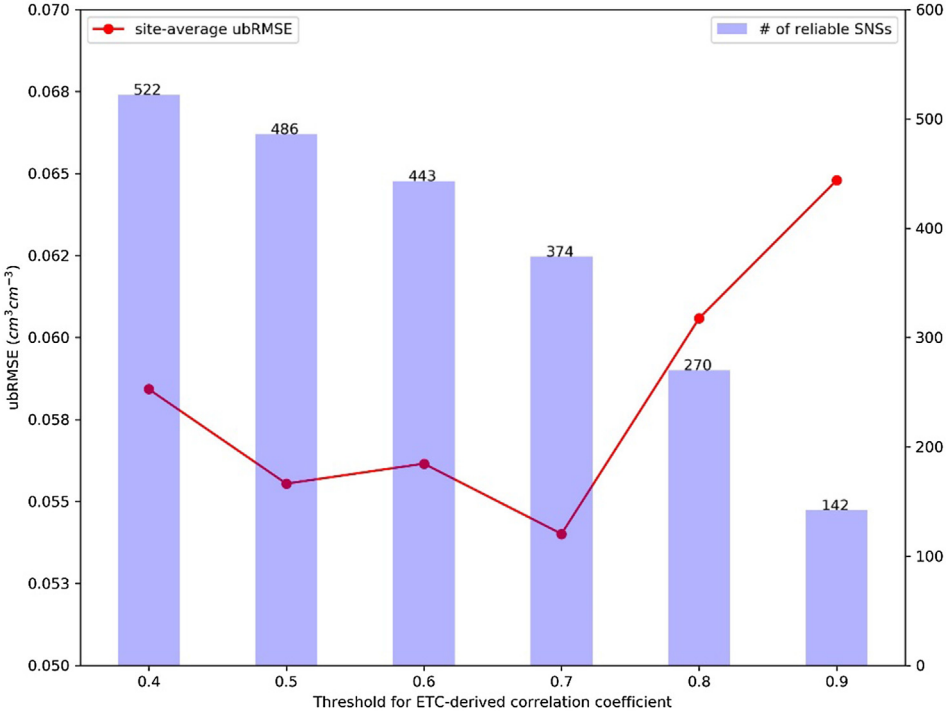

Fig. 13. The effect of the thresholds for extended triple collocation (ETC)-derived correlation coefficients on the number of reliable SNSs and the site-average ubRMSE of PSCI-SSM against ground-based measurements.

retrievals in SPL3SMP are also modified in PSCI-SSM time series to more approximate the SSM values of the site-specific time series. Moreover, sharp rises in SSM time series usually following precipitation or irrigation events are precisely captured by PSCI-SSM for most situations. More importantly, while over- and under-estimation do exist in PSCI-SSM time series, the biases are small compared to SPL3SMP. All these findings above suggest that our proposed PSCI approach can produce commendably accurate SSM retrievals whose spatial and temporal patterns are more consistent with ground-based measurements than the official SMAP Level-3 radiometer SSM product.

5. Discussion

As mentioned in Section 1, the inspiration for this work is to replace the widely use of uncertain LSM simulations as training references with high-accuracy ground-based measurements. Therefore, it is necessary to compare these two different datasets by taking them as references to train the networks, respectively. Here, the GRNN model trained on model SSM simulations from regridded (36-km EASEv2) ERA-Interim was used to run the PSCI process for the same study area and the same study period as previous sections. A spread parameter of 0.01 was set, which gave the best model performance. This model was then compared to the previous GRNN model trained on reliable SNSs. To be fair, only model simulations that correspond with 36-km EASEv2 grids covering the reliable SNSs were taken into consideration. A ground-based comparison between SSM estimates using GRNN trained on ERA-Interim model simulations (hereinafter "ERA-PSCI-SSM" for short) and those using GRNN trained on reliable SNSs' measurements (hereinafter "SNS-PSCI-SSM" for short, which is identical to the "PSCI-SSM" in previous sections) was drawn to evaluate the models' predictive power over each network (Table 5 and Fig. 12). Note that the averaged metrics were only computed when a site has a p-value < 0.05.

From Table 5 and Fig. 12, mean R scores against ground-based measurements for SNS-PSCI-SSM are better than ERA-PSCI-SSM over all eight networks. The biggest improvement of R is obtained in COSMOS, increasing from 0.38 for ERA-PSCI-SSM to 0.75 for SNS-PSCI-SSM. In terms of the mean RSME, all networks except for the iRON network experience a decline from ERA-PSCI-SSM to SNS-PSCI-SSM. A similar result can be seen for the mean ubRMSE values. Besides, the mean biases for SNS-PSCI-SSM are much more satisfactory over most networks as compared to ERA-PSCI-SSM. These results suggest that it is more appropriate to use reliable ground-based measurements instead of model simulations to train the networks.

Another important issue to be discussed is the choice of the thresholds for the ETC-derived correlation coefficient between in-situ measurements and the unknown truth within the 36-km EASEv2 grid (i.e., $R_{ETC}(T, SNS)$ in Eq. (2)), which is supposed to affect the selection of reliable SNSs and the subsequent PSCI process. Here, a series of thresholds ranging from 0.4 to 0.9 with the interval of 0.1 were tested. For each threshold, the selected reliable SNSs' measurements were fed into GRNN and ran the PSCI process to generate PSCI-SSM. Then, ubRMSE scores between respective PSCI-SSM and ground-based measurements were calculated over each SNS and further averaged to obtain one site-average ubRMSE for brevity. The number of reliable SNSs and site-average ubRMSE scores based on each threshold are illustrated in Fig. 13. It is apparent that with the increase of the thresholds, the number of reliable SNSs decreases and the higher the threshold, the greater the decline, whereas the site-average ubRMSE presents a general trend of decreasing first and then increasing. These findings indicate that selecting reliable SNSs using ETC does help address the scale mismatch issue, but higher thresholds do not always lead to better performance of the generated PSCI-SSM. This makes sense since a small amount of reliable SNSs' measurements resulting from high thresholds may not establish a representative TB–SSM relationship. The minimum site-average ubRMSE is achieved at the threshold of 0.7, which explains why this threshold was used in previous sections.

6. Conclusions

This paper proposed the so-called "PSCI" method for the estimation of high-quality regional SSM in CONUS for the April 2015 to March 2018 period. The main ideology behind this method is using a GRNN to establish a nonlinear relationship between SMAP TB observations (and auxiliary data) and reliable SNSs' SSM measurements. The reliable SNSs are determined based on the ETC-derived correlation coefficient in excess of 0.7, in order to address the scale mismatch issue resulting from small spatial support of ground-based measurements. The model evaluation results based on the 10-fold CV technique showed that GRNN gave much better performance compared to BPNN and that the GRNN model trained on reliable SNSs outperformed the other GRNN model trained on all SNSs, suggesting the added value of the ETC technique. The obtained quantitative indicators for the best trained GRNN model were fairly good—the cross-validated R and ubRMSE values were 0.88 and 0.050 $cm^3\ cm^{-3}$, respectively, demonstrating our methodology was able to accurately describe the TB–SSM relationship. The further temporal and spatial analysis of the retrieved PSCI-SSM product in comparison with SPL3SMP and ERA-Interim revealed our PSCI-SSM held the most stable and reasonable estimation and could well capture the spatio-temporal variations of SSM. In addition, the ground-based comparisons between PSCI-SSM and SPL3SMP/ERA-PSCI-SSM over each SNS/network showed PSCI-SSM was more consistent with ground-based measurements.

To conclude, the PSCI method proposed in this study has shown great potential in estimating reliable regional SSM records from satellite observations, using the ETC technique and GRNN model trained on reliable ground-based measurements. The ideology of our work may extend to SSM retrievals from other satellite missions such as ASCAT and SMOS and apply in other geographical regions in the world. Future work will focus on a more appropriate selection of reliable SNSs, including a variable threshold for the ETC-derived correlation coefficient according to the number of SNSs in different regions, other indicators to identify individual SNSs' reliability, and new technologies, such as the "inverse footprint" method (Orlowsky and Seneviratne, 2014), for determining the representativeness of in-situ sites. The incorporation of more advanced deep learning techniques may also be taken into consideration to replace GRNN.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 41922008). The authors would like to thank the NASA NSIDC DAAC for making the SMAP data publicly available and the ISMN for providing the in-situ data.

References

Aires, F., Prigent, C., Rossow, W., 2005. Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: 2. Global statistical relationships. J. Geophys. Res. Atmos. 110.

Al-Mahasneh, A.J., Anavatti, S.G., Garratt, M.A., 2018. Review of applications of generalized regression neural networks in identification and control of dynamic systems. arXiv preprint arXiv:1805.11236.

Basharinov, A.Y., Shutko, A., 1975. Simulation studies of the SHF radiation characteristics of soils under moist conditions. *Simulation studies of the SHF radiation characteristics of soils under moist conditions NASA Transl. into ENGLISH of"" Modelnyye Issledovaniya SVCh Radiatsionnykh Kharacteristik Pochvo-Gruntov v Usloviyakh Uvlazhneniya"(unpublished report) Moscow, Acad. of Sci. of the USSR*. Inst. Radio Eng. Electron. 1975, 1–81.

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. J. Hydrometeorol. 14, 977–988.
Berrisford, P., Dee, D., Poli, P., Brugge, R., Fielding, K., Fuentes, M., Kallberg, P., Kobayashi, S., Uppala, S., Simmons, A., 2011. The ERA-Interim archive, version 2.0.
Bierkens, M.F., Bell, V.A., Burek, P., Chaney, N., Condon, L.E., David, C.H., de Roo, A., Döll, P., Drost, N., Famiglietti, J.S., 2015. Hyper-resolution global hydrological modelling: what is next? “Everywhere and locally relevant”. Hydrol. Process. 29, 310–320.
Boyd, D., Foody, G., Ripple, W., 2002. Evaluation of approaches for forest cover estimation in the Pacific Northwest, USA, using remote sensing. Appl. Geogr. 22, 375–392.
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 Int. J. Geo-Inf. 1, 32–45.
Chai, S.-S., Walker, J.P., Makarynskyy, O., Kuhn, M., Veenendaal, B., West, G., 2009. Use of soil moisture variability in artificial neural network retrieval of soil moisture. Remote Sens. 2, 166–190.
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 Trans. Geosci. Remote Sens. 54, 4994–5007.
Chen, F., Crow, W.T., Colliander, A., Cosh, M.H., Jackson, T.J., Bindlish, R., Reichle, R.H., Chan, S.K., Bosch, D.D., Starks, P.J., 2017. Application of triple collocation in ground-based validation of Soil Moisture Active/Passive (SMAP) level 2 data products. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 489–502.
Cigizoglu, H.K., Alp, M., 2006. Generalized regression neural network in modelling river sediment yield. Adv. Eng. Softw. 37, 63–68.
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 Sens. Environ. 191, 215–231.
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. Rev. Geophys. 50.
De Lannoy, G.J., Reichle, R.H., Peng, J., Kerr, Y., Castro, R., Kim, E.J., Liu, Q., 2015. Converting between SMOS and SMAP level-1 brightness temperature observations over nonfrozen land. IEEE Geosci. Remote Sens. Lett. 12, 1908–1912.
Dee, D.P., Uppala, S., 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. Q. J. R. Meteorolog. Soc. 137, 553–597.
del Rosario Martinez-Blanco, M., Castañeda-Miranda, V.H., Ornelas-Vargas, G., Guerrero-Osuna, H.A., Solis-Sanchez, L.O., Castañeda-Miranda, R., Celaya-Padilla, J.M., Galvan-Tejada, C.E., Galvan-Tejada, J.I., Vega-Carrillo, H.R., 2016. Generalized Regression Neural Networks with Application in Neutron Spectrometry. InTech Croatia.
Desilets, D., Zreda, M., Ferré, T.P., 2010. Nature's neutron probe: land surface hydrology at an elusive scale with cosmic rays. Water Resour. Res. 46.
Dorigo, W., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., Oevelen, P.V., 2011. The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. Hydrol. Earth Syst. Sci. 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 J. 12.
Draper, C., Reichle, R., de Jeu, R., Naeimi, V., Parinussa, R., Wagner, W., 2013. Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. Remote Sens. Environ. 137, 288–298.
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. Proc. 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. J. Hydrometeorol. 11, 832–840.
Entekhabi, D., Rodriguez-Iturbe, I., Castelli, F., 1996. Mutual interaction of soil moisture state and atmospheric processes. J. Hydrol. 184, 3–17.
Famiglietti, J.S., Ryu, D., Berg, A.A., Rodell, M., Jackson, T.J., 2008. Field observations of soil moisture variability across scales. Water Resour. Res. 44.
Gardner, M.W., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. 32, 2627–2636.
Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., Wagner, W., 2016. Recent advances in (soil moisture) triple collocation analysis. Int. J. Appl. Earth Obs. Geoinf. 45, 200–211.
Jiménez, C., Clark, D.B., Kolassa, J., Aires, F., Prigent, C., 2013. A joint analysis of modeled soil moisture fields and satellite observations. J. Geophys. Res.: Atmos. 118, 6771–6782.
Kerr, Y.H., Al-Yaari, A., Rodriguez-Fernandez, N., Parrens, M., Molero, B., Leroux, D., Bircher, S., Mahmoodi, A., Mialon, A., Richaume, P., 2016. Overview of SMOS performance in terms of global soil moisture monitoring after six years in operation. Remote Sens. Environ. 180, 40–63.
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 Trans. Geosci. Remote Sens. 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. Proc. IEEE 98, 666–687.
Kişi, Ö., 2006. Generalized regression neural networks for evapotranspiration modelling. Hydrol. Sci. J. 51, 1092–1105.
Kolassa, J., Aires, F., Polcher, J., Prigent, C., Jimenez, C., Pereira, J.-M., 2013. Soil moisture retrieval from multi-instrument observations: information content analysis and retrieval methodology. J. Geophys. Res.: Atmos. 118, 4847–4859.
Kolassa, J., Gentine, P., Prigent, C., Aires, F., 2016. Soil moisture retrieval from AMSR-E and ASCAT microwave observation synergy. Part 1: Satellite data analysis. Remote Sens. Environ. 173, 1–14.
Kolassa, J., Reichle, R.H., Draper, C.S., 2017. Merging active and passive microwave observations in soil moisture data assimilation. Remote Sens. Environ. 191, 117–130.
Kolassa, J., Reichle, R.H., Liu, Q., Alemohammad, S.H., Gentine, P., Aida, K., Asanuma, J., Bircher, S., Caldwell, T., Colliander, A., Cosh, M., Collins, C.H., Jackson, T.J., Martinez-Fernandez, J., McNairn, H., Pacheco, A., Thibeault, M., Walker, J.P., 2018. Estimating surface soil moisture from SMAP observations using a Neural Network technique. Remote Sens. Environ. 204, 43–59.
Konate, A.A., Pan, H., Khan, N., Yang, J.H., 2015. Generalized regression and feed-forward back propagation neural networks in modelling porosity from geophysical well logs. J. Pet. Explor. Prod. Technol. 5, 157–166.
Larson, K.M., Braun, J.J., Small, E.E., Zavorotny, V.U., Gutmann, E.D., Bilich, A.L., 2010. GPS multipath and its relation to near-surface soil moisture content. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 3, 91–99.
Larson, K.M., Small, E.E., Gutmann, E., Bilich, A., Axelrad, P., Braun, J., 2008a. Using GPS multipath to measure soil moisture fluctuations: initial results. GPS Solutions 12, 173–177.
Larson, K.M., Small, E.E., Gutmann, E.D., Bilich, A.L., Braun, J.J., Zavorotny, V.U., 2008b. Use of GPS receivers as a soil moisture network for water cycle studies. Geophys. Res. Lett. 35.
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.
Lei, F., Crow, W., Shen, H., Parinussa, R., Holmes, T., 2015. The impact of local acquisition time on the accuracy of microwave surface soil moisture retrievals over the contiguous United States. Remote Sensing 7, 13448–13465.
Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. 2, 164–168.
Li, T., Shen, H., Yuan, Q., Zhang, X., Zhang, L., 2017a. Estimating ground-level PM2. 5 by fusing satellite and station observations: a geo-intelligent deep learning approach. Geophys. Res. Lett. 44.
Li, T., Shen, H., Zeng, C., Yuan, Q., Zhang, L., 2017b. Point-surface fusion of station measurements and satellite observations for mapping PM2. 5 distribution in China: methods and assessment. Atmos. Environ. 152, 477–489.
Ma, Z., Hu, X., Huang, L., Bi, J., Liu, Y., 2014. Estimating ground-level PM2. 5 in China using satellite remote sensing. Environ. Sci. Technol. 48, 7436–7444.
Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11, 431–441.
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. Geophys. Res. Lett. 41, 6229–6236.
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 J. Sel. Top. Appl. Earth Obs. Remote Sens. 3, 522–535.
Moghaddam, M., Silva, A., Clewley, D., Akbar, R., Hussaini, S., Whitcomb, J., Devarakonda, R., Shrestha, R., Cook, R., Prakash, G., 2016. Soil Moisture Profiles and Temperature Data from SoilSCAPE Sites, USA, ORNL DAAC, Oak Ridge, Tennessee, USA.
Njoku, E.G., Entekhabi, D., 1996. Passive microwave remote sensing of soil moisture. J. Hydrol. 184, 101–129.
O'Neill, P., Chan, S., Njoku, E., Jackson, T., Bindlish, R., 2018. SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 5. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado USA.
O'Neill, P., Chan, S., Njoku, E., Jackson, T., Bindlish, R., 2015. SMAP Algorithm Theoretical Basis Document: L2 & L3 Radiometer Soil Moisture (Passive) Products, Rev. In: B.
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 J. 14.
Orlowsky, B., Seneviratne, S.I., 2014. On the spatial representativeness of temporal dynamics at European weather stations. Int. J. Climatol. 34, 3154–3160.
Osenga, E., Arnott, J.C., Endsley, K.A., Katzenberger, J., 2019. Bioclimatic and soil moisture monitoring across elevation in a mountain watershed: opportunities for research and resource management. Water Resour. Res.
Ozerdem, M.S., Acar, E., Ekinci, R., 2017. Soil moisture estimation over vegetated agricultural areas: Tigris basin, turkey from radarsat-2 data by polarimetric decomposition models and a generalized regression neural network. Remote Sensing 9, 21.
Piepmeier, J.R., Johnson, J.T., Mohammed, P.N., Bradley, D., Ruf, C., Aksoy, M., Garcia, R., Hudson, D., Miles, L., Wong, M., 2014. Radio-frequency interference mitigation for the soil moisture active passive microwave radiometer. IEEE Trans. Geosci. Remote Sens. 52, 761–775.
Reich, S.L., Gomez, D., Dawidowski, L., 1999. Artificial neural network for the identification of unknown air pollution sources. Atmos. Environ. 33, 3045–3052.
Rodriguez-Fernandez, N.J., Aires, F., Richaume, P., Kerr, Y.H., Prigent, C., Kolassa, J., Cabot, F., Jimenez, C., Mahmoodi, A., Drusch, M., 2015. Soil moisture retrieval using neural networks: application to SMOS. IEEE Trans. Geosci. Remote Sens. 53, 5991–6007.
Rodríguez-Fernández, N.J., de Souza, V., Kerr, Y.H., Richaume, P., Al Bitar, A., 2017. Soil

moisture retrieval using SMOS brightness temperatures and a neural network trained on in situ measurements. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 1574–1577.
Rodriguez, J.D., Perez, A., Lozano, J.A., 2010. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans. Pattern Anal. Mach. Intell. 32, 569–575.
Rosenzweig, C., Tubiello, F.N., Goldberg, R., Mills, E., Bloomfield, J., 2002. Increased crop damage in the US from excess precipitation under climate change. Global Environ. Change 12, 197–202.
Rumelhart, D.E., Durbin, R., Golden, R., Chauvin, Y., 1995. Backpropagation: the basic theory. In: Backpropagation: Theory, Architectures and Applications, pp. 1–34.
Santi, E., Paloscia, S., Pettinato, S., Brocca, L., Ciabatta, L., 2016. Robust assessment of an operational algorithm for the retrieval of soil moisture from AMSR-E data in central Italy. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 2478–2492.
Schaefer, G.L., Cosh, M.H., Jackson, T.J., 2007. The USDA natural resources conservation service soil climate analysis network (SCAN). J. Atmos. Oceanic Technol. 24, 2073–2077.
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 Sci. Rev. 99, 125–161.
Şenkal, O., 2010. Modeling of solar radiation using remote sensing and artificial neural network in Turkey. Energy 35, 4795–4801.
Specht, D.F., 1991. A general regression neural network. IEEE Trans. Neural Networks 2, 568–576.
Stoffelen, A., 1998. Toward the true near-surface wind speed: error modeling and calibration using triple collocation. J. Geophys. Res. Oceans 103, 7755–7766.
Trenberth, K.E., Guillemot, C.J., 1995. Evaluation of the global atmospheric moisture budget as seen from analyses. J. Clim. 8, 2255–2272.
Vereecken, H., Schnepf, A., Hopmans, J.W., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M., Amelung, W., Aitkenhead, M., 2016. Modeling soil processes: review, key challenges, and new perspectives. Vadose Zone J. 15.
Vinnikov, K.Y., Robock, A., Qiu, S., Entin, J.K., 1999. Optimal design of surface networks for observation of soil moisture. J. Geophys. Res.: Atmos. 104, 19743–19749.
Viterbo, P., Beljaars, A., Mahfouf, J.F., Teixeira, J., 1999. The representation of soil moisture freezing and its impact on the stable boundary layer. Q. J. R. Meteorolog. Soc. 125, 2401–2426.
Viterbo, P., Beljaars, A.C., 1995. An improved land surface parameterization scheme in the ECMWF model and its validation. J. Clim. 8, 2716–2748.
Viterbo, P., Betts, A.K., 1999. Impact of the ECMWF reanalysis soil water on forecasts of the July 1993 Mississippi flood. J. Geophys. Res.: Atmos. 104, 19361–19366.
Wigneron, J.-P., Chanzy, A., Calvet, J.-C., Bruguier, N., 1995. A simple algorithm to retrieve soil moisture and vegetation biomass using passive microwave measurements over crop fields. Remote Sens. Environ. 51, 331–341.
Wigneron, J.-P., Schmugge, T., Chanzy, A., Calvet, J.-C., Kerr, Y., 1998. Use of passive microwave remote sensing to monitor soil moisture. Agronomie 18, 27–43.
Williams, C.A., Albertson, J.D., 2004. Soil moisture controls on canopy-scale water and carbon fluxes in an African savanna. Water Resour. Res. 40.
Wu, Y., Guo, J., Zhang, X., Tian, X., Zhang, J., Wang, Y., Duan, J., Li, X., 2012. Synergy of satellite and ground based observations in estimation of particulate matter in eastern China. Sci. Total Environ. 433, 20–30.
Xu, H., Yuan, Q., Li, T., Shen, H., Zhang, L., Jiang, H., 2018. Quality improvement of satellite soil moisture products by fusing with in-situ measurements and GNSS-R estimates in the western continental U.S. Remote Sensing 10.
Yao, L., Lu, N., 2014. Spatiotemporal distribution and short-term trends of particulate matter concentration over China, 2006–2010. Environ. Sci. Pollut. Res. 21, 9665–9675.
Yu, X.-H., 1992. Can backpropagation error surface not have local minima. IEEE Trans. Neural Networks 3, 1019–1021.
Zeng, J., Chen, K.-S., Bi, H., Chen, Q., 2016. A preliminary evaluation of the SMAP radiometer soil moisture product over United States and Europe using ground-based measurements. IEEE Trans. Geosci. Remote Sens. 54, 4929–4940.
Zreda, M., Desilets, D., Ferré, T., Scott, R.L., 2008. Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. Geophys. Res. Lett. 35.
Zreda, M., Shuttleworth, W., Zeng, X., Zweck, C., Desilets, D., Franz, T., Rosolem, R., 2012. COSMOS: the cosmic-ray soil moisture observing system. Hydrol. Earth Syst. Sci. 16, 4079–4099.