Remote Sensing of Environment 301 (2024) 113944

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A cross-resolution transfer learning approach for soil moisture retrieval from Sentinel-1 using limited training samples

Liujun Zhu [a,b], Junjie Dai [a,c], Yi Liu [c,d,*], Shanshui Yuan [a,d], Tianling Qin [e], Jeffrey P. Walker [b]

[a] Yangtze Institute for Conservation and Development, Hohai University, Nanjing 210098, China
[b] Department of Civil Engineering, Monash University, Clayton, Vic 3800, Australia
[c] College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China
[d] Key Laboratory of Hydrologic-Cycle and Hydrodynamic-System of Ministry of Water Resources, Hohai University, Nanjing 210098, China
[e] State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

ARTICLE INFO

Editor: Jing M. Chen

Keywords:
Soil moisture
Sentinel-1
Multiscale transfer learning
Deep learning

ABSTRACT

Synthetic Aperture Radar (SAR) data is increasingly popular as a data source for global near-surface soil moisture mapping, but large-scale applications are still challenging due to the complex scattering process and the cumbersome data preprocessing. The emergence of deep learning methods has allowed advances in the remote sensing of large-scale surface parameters, but its application in SAR soil moisture retrieval has suffered from the availability of ground soil moisture measurements. Accordingly, this study proposed a cross-resolution transfer learning framework, with the assumption that sophisticated models for different spatial resolutions share a similar model architecture and trainable parameters. A robust high-resolution model can thus be trained with fewer samples by using coarse models. Accordingly, 25 deep learning models were pre-trained taking ~387,000 Soil Moisture Active Passive (SMAP) Level-3 9 km enhanced passive soil moisture measurements as the truth, with an average validation RMSE of 0.03 m^3/m^3. They were then transferred to finer grids of 0.1–1 km using a small number of in-situ samples. A total of ~190,000 daily soil moisture measurements from the international soil moisture network (ISMN) were used to evaluate the proposed framework in three scenarios. The results show that 1) 5000–6000 random samples are sufficient to achieve a target RMSE of 0.06 m^3/m^3; 2) training samples from a short period (2 or 4 months for Sentinel-1) of 2021 resulted in an overall RMSE of ~0.068 m^3/m^3 in an independent period of 2016–2020; 3) the transfer learning also improved the retrieval accuracy (10–30% in relative) over areas without ground samples used for training but failed to yield an acceptable accuracy over mountainous areas. The promising results from this study confirmed the effectiveness of using "pre-trained models + scenario specific models" for regional to global soil moisture retrieval from Sentinel-1.

1. Introduction

Near surface soil moisture (top 5 cm) from remote sensing is critical for advances to be made in applications such as weather prediction, water resource management, soil pollution control, and smart agriculture (Peng et al., 2020). Arrival of the modern-era Synthetic Aperture Radar (SAR) missions (active microwave) has seen an increased utilization for operational soil moisture monitoring, due to the improved temporal resolution and introduction of open-access data policies. While substantial efforts have been made to develop operational algorithms (Balenzano et al., 2021; Bauer-Marschallinger et al., 2018; Zhu et al., 2022; Zhu et al., 2023), active microwave has not yet gained the same level of acceptance as passive microwave because of the cumbersome data preprocessing, complex forward scattering modeling and retrieval process implementation, and limitations in achieved accuracy. A target unbiased root mean square error (RMSE) of 0.06 m^3/m^3 has been defined for radar-based soil moisture mission products (Kim et al., 2014b; Lal et al., 2023), being less restrictive than the 0.04 m^3/m^3 science requirement of passive missions (Entekhabi et al., 2010; Kerr et al., 2010).

Soil moisture retrieval from SAR data generally involves inversion of scattering models, including those for bare soil (Fung et al., 1992; Oh et al., 2002) and vegetated areas (Bracaglia et al., 1995). Though the earth surface has been parameterized using a limited number of soil surface and vegetation parameters, the number of independent SAR observations are less than that of the surface parameters to be

* Corresponding author at: College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China.
E-mail address: liuyihhdx@126.com (Y. Liu).

https://doi.org/10.1016/j.rse.2023.113944
Received 4 July 2023; Received in revised form 22 November 2023; Accepted 28 November 2023
Available online 7 December 2023

determined, resulting in an ill-posed problem (Zhu et al., 2019b). Successful inversion of these models therefore requires a considerable amount of site specific a priori information. Consequently, studies to date have mostly focused on a small area (e.g., Yadav et al., 2020; Zhu et al., 2019b), with their applications to other areas being questionable.

The change detection methods (Balenzano et al., 2011; Palmisano et al., 2020; Wagner et al., 1999; Zhu et al., 2022) and time series methods (Kim et al., 2014a; Kim et al., 2012; Zhu et al., 2019a, 2019b) assume time-invariant surface roughness and vegetation parameters over a short time window (a few weeks) and thus decouple the scattering contribution of soil moisture from that of other surface parameters, allowing near operational soil moisture retrieval at continental or global scales. However, in many cases they suffer from the validity of the time-invariant roughness and vegetation assumption (Zhu et al., 2019c) and also require a cumbersome pre-processing of incidence angle normalization (Bauer-Marschallinger et al., 2018). These cannot be addressed shortly because of the principles behind the time series methods as well as the scattering model-based methods. The new generalization of data-driven methods, e.g., the so-called deep learning approaches, provide a promising alternative. Importantly, they do not require complex physical modeling and can make full use of the multimodal data, with the most powerful features (explainable variables) being learned automatically.

Many studies have applied various deep neural network approaches to the Soil Moisture Active and Passive (SMAP) and Soil Moisture and Ocean Salinity (SMOS) missions for making enhanced soil moisture retrievals, including the low resolution SMAP and SMOS surface soil moisture products (Gao et al., 2022; Kolassa et al., 2018; Rodriguez-Fernandez et al., 2015), a prolongation of the existing SMAP products (Fang et al., 2019; Fang et al., 2017), downscaled SMAP or SMOS products (Xu et al., 2022; Zhao et al., 2022) and soil moisture at deeper layers (Karthikeyan and Mishra, 2021; Yinglan et al., 2022; Yu et al., 2021). The SMAP, SMOS and/or climate model reanalysis (e.g., ERA5) soil moisture products have been commonly used as the truth for training, with the in-situ records being used for validation at resolutions ranging from 1 km to 10s km. To achieve improved performance at a higher resolution, in-situ records were used together with the coarse truth in a multi-scale training scheme (Liu et al., 2022).

Data-driven methods have also been increasingly used for high resolution soil moisture retrieval from SAR data. Some studies trained neural networks as approximations of the scattering models to avoid the complex inversion process (Dong et al., 2023). Other studies trained various machine learning models for SAR soil moisture retrieval from field to continental (Chaudhary et al., 2022; Chen et al., 2021) and global (Batchu et al., 2023; Celik et al., 2022) scales based on the International Soil Moisture Network (ISMN, Dorigo et al., 2021). These approaches have generally achieved higher accuracy than the aforementioned methods based on physical models or time series. However, the relatively higher accuracy was achieved only on scenarios having a sufficient amount of training samples, e.g., more than half of the samples was used in training. The performance likely to be poor in areas where model inputs have different distributions or only limited training samples are available. While the ground soil moisture measurements of the ISMN archive have enabled the training of deep and wide neural networks, further improvements can be challenging due to the high cost of collecting more samples and the fact that existing stations are sited unevenly in space and time.

A natural idea to address this limitation is to include a priori knowledge that relates to the soil moisture retrieval at high spatial resolutions. The *a prior* knowledge can be the aforementioned scattering models, process-based models (Li et al., 2021), and/or the empirical knowledge learned from a similar task. The latter can be categorized as transfer learning that reuses a pre-trained model of task A as the starting point for developing a model to undertake a similar task B. The success of transfer learning relies on the effectiveness of the pre-trained model A and the similarity between task A and B (Tan et al., 2018). In view of soil moisture retrieval at high resolution (task B), task A can be soil moisture at low resolutions and can also be soil texture because they are tightly related in the land surface process and affect the remote sensing data collectively. Deep transfer learning has demonstrated performance for applications with sparse training samples, inheriting the advantages of deep learning in feature presentation and transfer learning in knowledge transferring across tasks. A considerable number of deep transfer learning approaches have been proposed recently for advances in many applications (Zhuang et al., 2020). Since a sophisticated deep neural network can contain millions to billions of trainable parameters, it is impossible to start from scratch for each application, and so a machine learning model built on a well pre-trained model can be more flexible, being another advantage of deep transfer learning. For example, the Chat Pre-trained Transformer (ChatGPT) was finetuned from the GPT-4 that has one trillion parameters (Katz et al., 2023).

A cross-resolution transfer learning framework was developed in this study for daily averaged soil moisture retrieval from Sentinel-1 data, focusing on areas and/or periods with sparse ground in-situ measurements. Since the SMAP and SMOS can provide soil moisture data with an accuracy of ~0.04 m^3/m^3 (Entekhabi et al., 2010; Kerr et al., 2010), reliable pre-trained models can be trained for global soil moisture retrieval from Sentinel-1 at a coarse grid, e.g., the 9 km Equal-Area Scalable Earth Grids 2.0 (EASE-2.0). The pre-trained models can then be assumed to share a similar model architecture and trainable parameters with models at higher resolutions as the scattering processes at different scales are similar. A robust high-resolution model can thus be trained with limited in-situ measurements by reusing the pre-trained coarse model. The proposed cross-resolution transfer learning concept was designed for both descending (6 am) and ascending (6 pm) data of Sentinel-1. However, the target soil moisture of the trained models was the daily averaged values, because use of descending or ascending soil moisture products alone would have a much lower temporal resolution with a reduced spatial coverage, and having two separate soil moisture products would be less valuable. Around 190,000 daily averaged soil moisture values from 1220 ISMN stations over 2016–2021 was used to validate the proposed concept. Following the mode of "pre-trained models + scenario specific models", three evaluation/application scenarios were involved taking the pre-trained models as the start points, attaining empirical answers to the following questions: 1) How many in-situ measurements are required for a retrieval root mean square error (RMSE) of 0.06 m^3/m^3 at scales from 0.1 to 1 km? 2) Are the archived in-situ soil moisture sufficient for periods without samples? and 3) Can reliable retrieval be made for areas without samples?

2. Data and preprocessing

The proposed deep transfer learning framework contains knowledge transferring from a coarse grid of 9 km to the target resolution of 0.1–1 km, and thus the spatial size of the model inputs (x) and the truth (y) were both pre-processed to 9 km, 1 km, 0.5 km and 0.1 km. Additional to the main data source of Sentinel-1, other related remote sensing data, weather variables and auxiliary variables were also included because deep learning methods are flexible enough to combine these multimodal data and have an enhanced capability in feature extraction and learning. The extra input data were selected based on two criteria: 1) they should be relevant to the surface scattering process or the soil moisture in view of land surface process; and 2) they should be open-access and thus can be used for operational soil moisture retrieval. A rigorous feature selection was not made in this study because the so-called "optimal" input combination may be different for different scales and models (Saarela and Jauhiainen, 2021).

2.1. Ground soil moisture

The soil moisture data from 2016 to 2021 available on the ISMN archive were used for training and validation at a resolution of 0.1 km,

0.5 km and 1 km. All the high-quality recordings of near surface soil moisture (≤ 5 cm) with a flag of good ("G") were considered from a total of 1220 stations, located in 1030 EASE-2.0 9 km grid cells. Since the measuring depths of soil moisture, the number of sensors at each measuring depth, and the recording rate varied across stations, the measurements of each station collected from the top 5 cm, on the same calendar day, and in the same grid cell, were averaged as the daily near surface soil moisture. A summary of networks, grid cells and the available samples is provided in Table 1 and Fig. 1, with the landcover and climate type being from Buchhorn et al. (2020) and Peel et al. (2007) respectively.

2.2. Remote sensing data

All the pre-processed Sentinel-1 Interferometric Wide (IW) Ground Range Detected (GRD) acquisitions over the 1030 9 km grid cells were used in this study. The Google Earth Engine (GEE) Sentinel-1 GRD archive was used because it is ready for end users with the standard preprocessing steps (e.g., radiometric calibration) already applied at the backend. In this study, the original GRD VV/VH backscatter coefficients (sigma null) with a pixel spacing of 10 m were first re-projected to the EASE-2.0 9 km grid. The Sentinel-1 inputs of the 9 km coarse model were then obtained using a weight averaging resample method. For a 9 km grid cell, the weight of an overlapped Sentinel-1 pixel was calculated as the fraction of the 9 km pixel area covered by the Sentinel-1 pixel and all the overlapped Sentinel-1 pixels averaged as the output based on their weights. Similarly, the Sentinel-1 inputs for 0.1 km, 0.5 km and 1 km were obtained based on the corresponding EASE-2.0 sub-grids using the same method.

Since the radiometric terrain flattening was not made by the GEE Sentinel-1 GRD data, three terrain features, i.e., elevation, slope and aspect, extracted from the Shuttle Radar Topography Mission Digital Elevation Data Version 3 (SRTM DEM V3) were used as part of the model inputs, assuming that the effect of terrain on radar observations can be learned by the learning system. Notably, the slope and aspect were calculated at the original resolution of 30 m and then resampled to the target grids using the aforementioned weight averaging method. Moreover, three soil texture parameters were extracted from the global gridded soil information (SoilGrids 250 m, Poggio et al., 2021) to account for their effect on backscatter, being sand ratio, clay ratio and bulk density. The SoilGrids is a system for global digital soil mapping, built on over 230,000 soil profile observations and a few environmental covariates. Similarly, the original soil textures at a spatial resolution of 250 m were reprojected and resampled to the four EASE-2.0 grids as an auxiliary input of the retrieval models.

The SMAP Level-3 9 km enhanced passive soil moisture product Version 5 (O'Neill et al., 2021) was used for training of the coarse models, considering its relatively higher resolution with respect to other passive products. The day specific bias of measurements from descending (local solar time of 6 am) and ascending (local solar time of 6 pm) passes were removed taking the descending data as the benchmark. The average of the measurements from descending and ascending passes was then calculated as the truth of the coarse model at 9 km. Though the SMAP and Sentinel-1 data are available globally and a coarse model can be trained using all historical data, only the 1030 9 km grid cells containing ground stations were used due to the limited storage and computation resources, resulting in 383,596 Sentinel-1 image patches with a spatial size of 9 km × 9 km.

The time series Normalized Difference Vegetation Index (NDVI) produced from the Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) was utilized to partly represent the effect of vegetation. The two MODIS 16-day composite products (MOD13Q1 and MYD13Q1, V6) with a spatial resolution of 250 m were combined to provide a temporal resolution of 8 days, starting from the first day of

Table 1
The number of samples for the four grid resolutions from the 26 soil moisture networks.

Network	Number of stations	Number of samples				Land cover*	Köppen climate classification**
		0.1 km	0.5 km	1 km	9 km		
AMMA-CATCH	7	639	543	543	1615	3, 11	Aw, BS
ARM (Galle et al., 2018)	17	4055	3979	3979	3873	2, 3	Cf
BIEBRZA_S-1 (Cook, 2016)	18	1429	745	419	1712	2, 3	Df
FLUXNET-AMERIFLUX	5	875	869	869	2200	2, 3, 11	Cs
FR_Aqui	4	3032	3032	3032	253	3, 10, 11	Cf
HOAL (Blöschl et al., 2016)	31	3055	1345	791	1848	3	Df
HOBE (Jensen and Refsgaard, 2018)	29	7637	6661	6418	30,683	3, 10, 11	Df
HYDROL-NET_PERUGIA (Flammini et al., 2018)	2	56	56	56	709	3, 11	Cs
iRON (Osenga et al., 2021)	10	1264	1162	1162	2150	2, 3, 11	Df
Tibet-Obs (Su et al., 2011)	41	2207	2172	2172	7804	2	Dw, BW, Ds, E
OzNet (Smith et al., 2012)	19	1559	1559	1559	2629	2, 3	BS, Cf
PBO_H2O (Larson et al., 2008)	142	6414	6414	6414	52,712	1–3, 5, 11	Af, BW, BS, Cs, Cw, Cf, Ds, Df
REMEDHUS (González-Zamora et al., 2019)	20	8039	8039	8039	14,347	1–3, 11	BS
RISMA (Ojo et al., 2015)	22	1830	1830	1830	2912	2, 3	Df
RSMN	19	11,929	11,929	11,929	16,649	3, 4, 10, 11	Cf, Df
Ru_CFR	2	67	67	67	380	10	Df
SCAN (Schaefer et al., 2007)	187	33,542	33,527	33,527	44,109	1–5, 10, 11	Af, Aw, BW, BS, Cs, Cf, Ds, Df
SD_DEM (Ardö, 2012)	1	109	109	109	0	3	BW
SMN-SDR (Zhao et al., 2020)	32	1622	1622	1580	9813	2, 3	Dw
SMOSMANIA (Calvet et al., 2016)	22	12,921	12,921	12,921	17,749	1, 3, 10, 11	Cs, Cf, Df
SNOTEL (Service, 2022)	377	66,492	66,492	66,492	121,936	1–5, 7, 10, 11	BW, BS, Cs, Cf, Ds, Df, E
SOILSCAPE (Moghaddam et al., 2016)	95	397	163	149	6992	1–3, 11	BS, Cs, Cf
TAHMO	4	424	424	424	590	1, 4, 10	Aw
TERENO (Bogena et al., 2018)	5	2631	2631	2631	5101	1, 3	Cf
USCRN (Bell et al., 2013)	107	20,180	20,159	20,159	23,565	1–3, 5, 8, 10, 11	BW, BS, Cs, Cf, Ds, Dw, Df
WSMN (Petropoulos and McCalmont, 2017)	2	26	26	26	2853	2, 3	Cf
Total	1220	192,431	188,476	187,297	383,596		

* 1–11 refers to shrubs, herbaceous, cultivated vegetation, built up, bare, snow, water, wetland, moss and lichen, closed forest, open forest.

** The first letter of A–E are tropical, arid, temperate, continental and polar respectively, while the second letter of f, m, w, W, s, S represent rainforest/no dry season, Monsoon, dry winter, arid desert, dry summer, semi-arid.

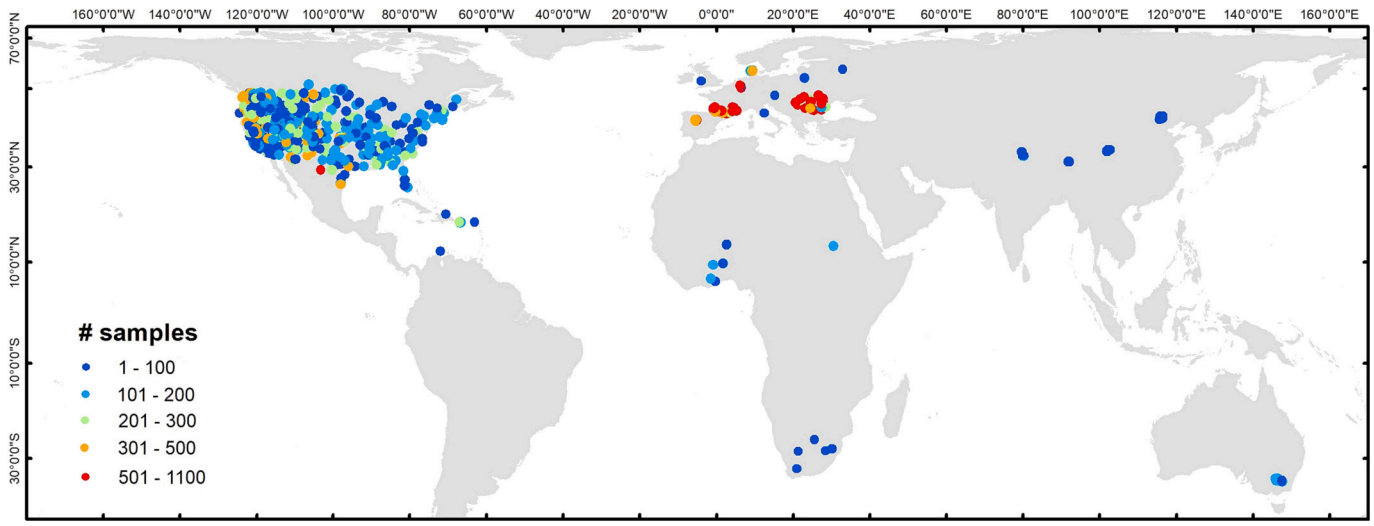

Fig. 1. The distribution of 1 km grid cells and the number of available samples for each grid cell.

every year (Didan, 2015). They were then reprojected and resampled to 0.1 km, 0.5 km, 1 km and 9 km using the aforementioned weight averaging method.

2.3. Reanalysis weather data

The fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric ReAnalysis land data (ERA5-land, Muñoz-Sabater et al., 2021) were used to provide valuable information about climate types and drivers of soil moisture. The ERA5-land products contain 50 hourly variables about, lake, snow, soil water, and vegetation status along with air temperature, radiation, evaporation, runoff, wind speed, surface pressure and precipitation at a spatial resolution of ~11 km. In general, including more variables is expected to result in a more powerful deep learning model, at the expense of a more complex structure requiring more training samples for the pre-trained and high-resolution models. Accordingly, only the total precipitation and air temperature at 2 m were selected, because they are the main indicators of climate types (Kottek et al., 2006) and have high correlation with the surface soil moisture (Feng and Liu, 2015). To be consistent with the 8-day composite NDVI, an 8-day moving average of the hourly ERA5-land data was made using a stride of 8 days, resulting in 8-day averaged temperature and precipitation. They were then reprojected and resampled to the target grids.

2.4. Preparation of samples

Each sample (the input of the proposed models) contains a set of time-invariant variables including the terrain features, soil texture and spatial location. The variables representing the sample locations are the same for the four target grids of 0.1, 0.5, 1 and 9 km. This enabled the model to learn special specific patterns but ignore the location difference within a 9 km grid. Since the Sentinel-1 data was treated as the main data source, the other input variables were prepared according to the calendar dates of Sentinel-1 data. Given a 9 km Sentinel-1 input of VV, VH and Local Incidence Angle (LIA) collected on Day of Year (DoY) t, one-year time series (8 days × 46 acquisitions) of NDVI, temperature and precipitation preceding DoY t were extracted. Moreover, the DoY of the Sentinel-1 observation was also included in the inputs to allow the model to differentiate patterns with similar dynamics but different seasons. For the training and validation of the coarse models, the truth at DoY t was interpolated from the pre-processed SMAP 9 km series using a spline function of first order, while the truth for the high-resolution models was the daily averaged in-situ soil moisture of DoY t. The spline interpolation was implemented using the 'sline' method of the Python Pandas module. Accordingly, the 383,596 Sentinel-1 images led to 383,596 samples for the coarse models, while the number of samples for high-resolution models was reduced to ~190,000 (Table 1) as many ISMN stations only covered a part of 2016–2021.

All variables were normalized to a range 0–1 using the min-max normalization method, with the maximum and minimum values of each variable being listed in Table 2. The ranges listed in Table 2 were expected to cover all the potential scenarios in the global mapping. Specifically, the DoY, terrain aspect and column number were first normalized from -π to π and then converted to variables ranging from 0 to 1 using cosine and sine functions, allowing the model to understand cyclic variables. For example, an aspect of 359° is closer to 0° than 90°, which cannot be made using a linear encoding of 0 to 359°. Consequently, each sample included 15 snapshot variables and 3 time series variables with a length of 46 (Table 2), with a total of 153 variables. The available samples for the 4 involved grids were provided in Table 1. The small difference of available samples among the 0.1, 0.5 and 1 km sets

Table 2
The variables contained in a sample (x) of the proposed model and the maximum and minimum values of each variable used in data normalization.

Type (Source)	Variable name	Minimum	Maximum	#Variables
SAR observations (Sentinel-1 A/B)	VV [dB]	−30	5	1
	VH [dB]	−35	0	1
	Incidence angle [°]	29.1	46	1
	Sentinel-1 DoY*	1	365	2
Soil textures (SoilGrids 250 m)	Sand [%]	0	100	1
	Clay [%]	0	100	1
	Bulk density [g/cm3]	1	1.8	1
Terrain features (SRTM DEM V3)	Elevation [m]	0	5500	1
	Slope [°]	0	40	1
	Aspect [°]*	0	360	2
Locations (EASE-2.0 9 km grids)	9 km grid row number	0	1624	1
	9 km grid column number*	0	3856	2
Vegetation descriptors (MOD13Q1 and MYD13Q1, V6)	NDVI	−0.2	1	46
Climate variables (EAR5-Land)	Temperature [°C]	−10	35	46
	Precipitation [m/d]	0	0.3	46

* Projected to the cosine and sine space.

was caused by the averaging of multiple stations within the grid.

3. Methods

3.1. Densely connected network for so il moisture (DenseSM)

The proposed DenseSM was built on the densely connected convolution networks (DenseNet, Huang et al., 2017b) and the temporal convolution neural network (Zhu et al., 2021). The DenseSM (Fig. 2) starts from three sequential temporal convolutional blocks to learn temporal dynamics of vegetation and weather information from the 1-year time series of precipitation, temperature and NDVI. Each convolutional block contains a one-dimensional temporal convolution layer (Conv1D), followed by an averaging pooling layer to reduce the dimension of the input feature space. The number of convolution filters in the three blocks was 32, 16 and 4, respectively, while the convolution kernel size and stride were fixed to 5 and 1 respectively. In contrast, the three pooling layers had a kernel size of 2, 2 and 4, respectively. Since the ceiling mode was used for pooling layers, the length of time series after the three pooling layers was 23 (46/2), 12 (23/2) and 3 (12/4) respectively. Accordingly, a total of 12 (4 filters $\times$ 3 in length) weather and NDVI features were learned by the three convolution blocks, being similar to the number of snapshot features. Complex convolution blocks with more output features can be more powerful to capture the critical temporal dynamics of precipitation, temperature and NDVI but they may result in imbalance between temporal and snapshot features. A batch normalization layer and a dropout layer with a fixed dropout rate of 0.5 (Srivastava et al., 2014) were included after each pooling layer to avoid overfitting. A Rectified Linear Unit (ReLU) activation function was also applied for non-linear transformation.

The precipitation, temperature and NDVI of the last day were then concatenated with the 12 temporal features and the 15 snapshot features as the inputs to dense blocks. Each dense block contained three fully connected layers, with additional operations of applying the batch normalization and the ReLU activation function. A shortcut was implemented to concatenate the input and output of each dense block, being the input of all the subsequent dense blocks. This allows feature reuse and substantially improved parameter efficiency (Huang et al., 2017b). The output of the dense blocks was fed into a fully connected layer of 32 neurons, followed by a fully connected layer of 1 neuron, being the final output (soil moisture) of the retrieval model. The number of neurons within each dense layer (w) and the number of dense blocks (d) are the two hyperparameters required in the DenseSM model, representing the width and depth of the model.

3.2. Ensemble transfer learning framework

The main assumption behind the proposed transfer learning method is that a deep neural network trained for the 9 km coarse grid can share a similar model architecture and trainable parameters with that for high-resolution soil moisture. A robust high-resolution model can thus be trained with limited samples by reusing the coarse models (Fig. 3 a). Accordingly, a coarse model should be trained first taking the 9 km SMAP enhanced L3 radiometer soil moisture as the truth. Since sufficient training samples (383,596 in this study) were available for 9 km resolution, a deeper and wider model with more trainable parameters can be more powerful, as suggested by many studies (Eldan and Shamir, 2016; Wu et al., 2019). However, retrieval performance at 9 km resolution is not the main target of this study, and a huge coarse model will not necessarily result in a better accuracy when transferred to high resolution, especially for the scenarios with sparse samples. Accordingly, the w and d were set to a relatively small number of ≤ 128 and ≤ 5 respectively. Searching for the optimal w and d using the commonly used cross-validation technique requires extra validation samples, being challenging for areas with sparse in-situ measurements. Moreover, sparseness of the measurements varied in different application scenarios, and thus a determined w or d can be questionable. Consequently, instead of searching for the optimal width (w), depth (d) and other hyperparameters, a few DenseSM models with different w and d were trained for 9 km. Specifically, w was set to five conventional values, being powers of 2, including 8, 16, 32, 64 and 128, while d was set from 1 to 5 with an interval of 1, resulting in 25 DenseSM models. The trainable parameters of these models therefore ranged from 5316 (DenseSM-1-8) to 377,785 (DenseSM-5-128).

In the training of coarse models, 90% of the 383,596 samples were randomly selected as the training set, while the other 10% were used for validation. Independent validation was not undertaken because the pre-trained models were used in the same area and period as for the

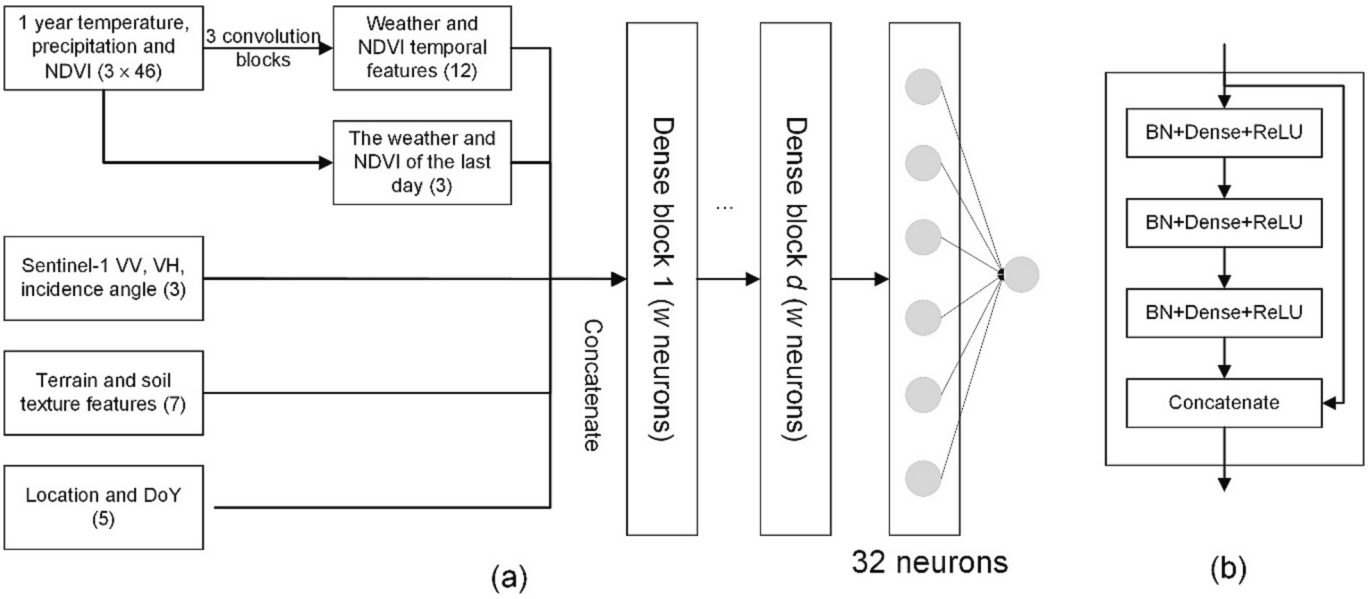

Fig. 2. The architecture of the densely connected network for soil moisture (DenseSM) (a) and the dense block (b). The BN, Dense and ReLU are batch normalization, a fully connected dense layer and an activation layer of ReLU, respectively. The numbers in the brackets are the dimensions of the variables. A DenseSM contains d dense blocks, with the number of neurons in each dense layer being w.

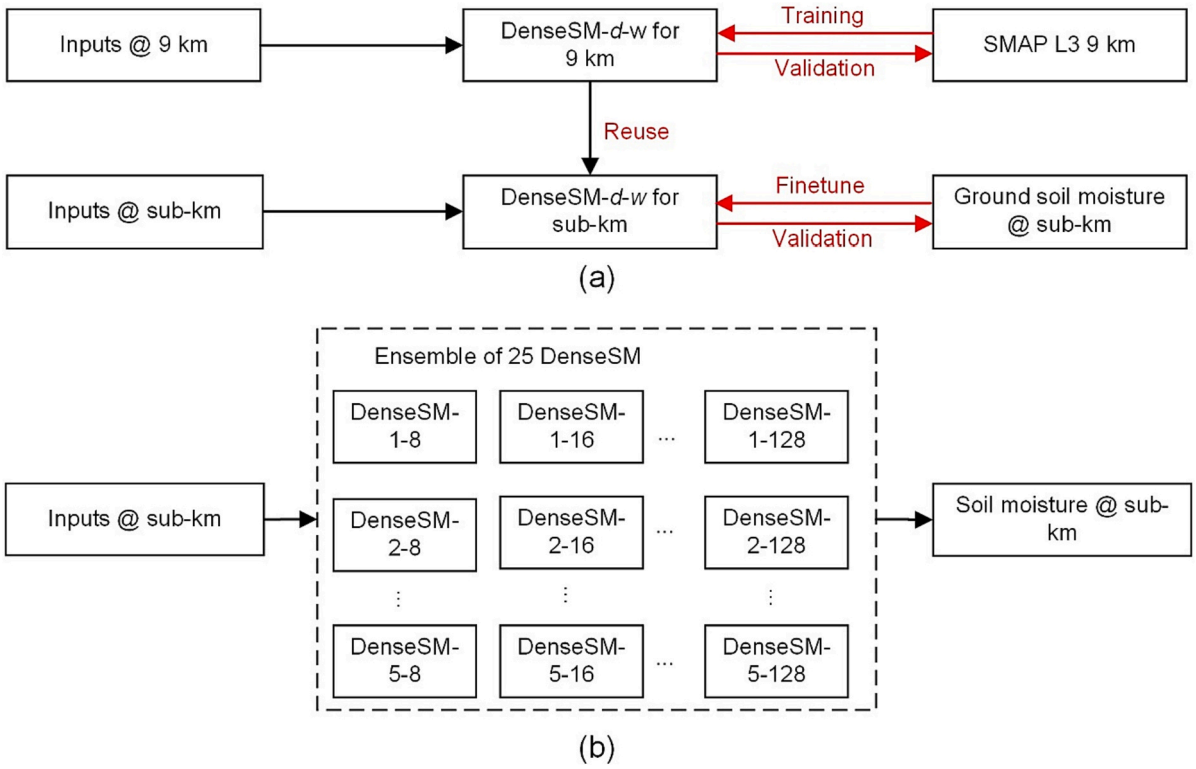

Fig. 3. The transfer learning framework from the 9 km grid to sub-km grids using a finetune method (a) and the ensemble of the 25 finetuned DenseSM for sub-km soil moisture retrieval (DenseSM-E).

following transfer learning, and thus a simple check of overfitting was sufficient. However, an extra scenario of using 30% samples in validation was made and included in the supplementary material (Fig. S1), with marginal difference in RMSE (<0.001 m^3/m^3) and R (< 0.01). The models were trained using a loss function of RMSE and the Adam optimizer with default parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$) (Kingma and Ba, 2014). The batch size and learning rate were set to 128 and 0.01, considering the balance of training accuracy and efficiency. Each model was trained for 100 epochs without early stopping, and the trainable parameters of the last 10 models (epochs) were averaged as the parameters of the trained model. This is known as snapshot ensemble of models (Huang et al., 2017a), allowing a more stable and reliable model.

In the transfer stage, the 25 trained coarse models were finetuned for a finer grid respectively (Fig. 3 a). Finetune is a re-training process that modifies the trainable parameters slightly for a task different from the original task, generally using a small number of samples and a small learning rate (Zhuang et al., 2020). In this study, each coarse model was retrained for 1000 epochs, using a small learning rate of $1e^{-5}$. The other settings were consistent with those of training a coarse model, including the snapshot ensemble of the last 10 epochs. The finetuned models were then used to estimate soil moisture separately, with the 25 estimations of each sample being taken as the ensemble average (Fig. 3 b). The ensemble of 25 DenseSM is named as DenseSM-E for short hereafter, while DenseSM-*d*-*w* is used to denote a single model with a width of *w* and a depth of *d*.

3.3. Transfer scenarios and validation metrics

The training and validation were focused on three transfer scenarios with sparse training samples, being the 1) dependent scenario, 2) independent on-site scenario and 3) independent off-site scenario. Since there is no clear definition of sparse training samples, a few cases with different training sizes were made in each scenario. In the dependent scenario, coarse models were finetuned using a small set of random training samples, ranging from 0.5% to 4.5% of the sample set with an interval of 1%. This is a much smaller number compared to other studies that used >50% of samples in training (Batchu et al., 2023; Karthikeyan and Mishra, 2021) and is expected to provide an estimate of the performance when ~1 to 8 thousand samples are available globally. Experiments with more training samples can still be interesting, but less critical as the main target of this study was scenarios with sparse training samples. Moreover, more training samples means fewer validation samples, and thus it can be challenging to compare the results of different cases. In this scenario, the training and validation sets are expected to be from the same input space, with limited difference in the distribution of each input variable.

The independent on-site scenario was designed to show the effect of available archived samples on the estimation at the same locations but in different calendar years. The grid cells involved in this scenario should thus have samples in at least two different calendar years. The training samples were all selected from 2021 because 1) samples from 2021 or 2016 have the largest time gap when compared with those from other calendar years, enabling to investigate the effect of varying time gaps, and 2) and the number of valid 1 km grid cells for 2021 was 635, being much larger than that of 2016 (381). More specifically, the same number of training samples (from 1 to 11) were selected from each grid cell with an interval of 2, allowing to reflect the performance of using a small varying sample size. The validation was made on the same grid cells but for the period 2016–2020.

In the independent off-site scenario, the samples from a small number of grid cells were randomly selected as the training samples, with the rest being used for validation. Since the soil, vegetation and climate types of a few selected cells can be substantially different from that of the validation set, this scenario is the most challenging transfer scenario. For example, the input features representing the locations of the training set will have a much smaller range than that of the validation set and the transfer models can be overfitted easily, resulting in values <0 m^3/m^3 or > 1 m^3/m^3. Accordingly, samples from a few grid cells were not

expected to support robust retrievals globally, being different from the dependent scenario and on-site scenario.

In this study, the off-site scenario was focused on two regional networks. Specifically, the Soil Climate Analysis Network (SCAN, Schaefer et al., 2007) and SNOwpack TELemetry network (SNOTEL) from the continental US (CONUS) were selected considering the relatively large number of stations, large spatial coverage and the long continuity of data records. While both networks were located in the CONUS, the SNOTEL stations were mainly located in the western mountainous areas focusing on snowpack monitoring and thus the local terrain features, vegetation and climate near a SNOTEL station can be more complex than that of a SCAN station. A model that works for a few SNOTEL stations can thus be useless for other SNOTEL or ISMN stations. Accordingly, validation on the SCAN and SNOTEL were expected to provide a comprehensive evaluation in a general to challenging scenario. Following the other two scenarios, the performance of this scenario was also evaluated for a few cases with small but varying training sample sizes. The grid cells used in training were set to be <10% of all grid cells, with a maximum number of ~17. The training samples were thus selected from 1 to 17 grid cells with an interval of 4, with the rest of the grid cells of the same network being used for validation. Specially, a same number of training samples from 9 km was used in the transferring stage and thus the batch normalization can be maintained. The commonly used leave-one-out and *k*-fold cross-validation was not used considering the small number of samples or grid cells used in training.

Four commonly used accuracy metrics, i.e., bias, correlation coefficient (R), RMSE and unbiased RMSE (ubRMSE), were calculated for all the validation samples, being denoted as the overall performance. Moreover, the statistics were also calculated for each grid cell having >15 samples, to show the spatial variation of the performance, being denoted as the cell-wise performance hereafter. Moreover, a model can be overfitted during the transfer learning resulting in soil moisture values <0 m^3/m^3 or > 1 m^3/m^3. Only the values within a validity range of 0–0.6 m^3/m^3 was considered and another indicator of inversion rate (0–1) was calculated to show the ratio of valid retrievals to the whole validation set.

All the training and validation were repeated 10 times in the transfer training phase using different random seeds, but only the mean statistics of the 10 replications were presented for simplicity. Though a few machine learning methods were proposed in literature for soil moisture retrieval or downscaling (Batchu et al., 2023; Celik et al., 2022), a direct comparison with the proposed framework was not feasible, as they were not developed for small training sets and had different inputs. In this study, high-resolution models without transfer learning were also trained and validated using the same training and validation samples, being the benchmark in the comparison.

4. Results

4.1. Evaluation of pre-trained coarse models

The training and validation results of three coarse models with varying depth (*d*) and width (*w*) are depicted in Fig. 4, with the results of the other 22 coarse models being similar. The three coarse models achieved a near zero bias in the training and validation phase. Moreover, the training and validation had almost the same accuracy statistics for all three models, suggesting that the coarse models were trained properly without overfitting or underfitting. As expected, both the training and validation accuracy was improved when the *w* and *d* increased, confirming that a bigger model is more powerful for scenarios with sufficient training samples. The poorest model was DenseSM-1-8 with a training and validation RMSE of ~0.04 m^3/m^3, being much better than

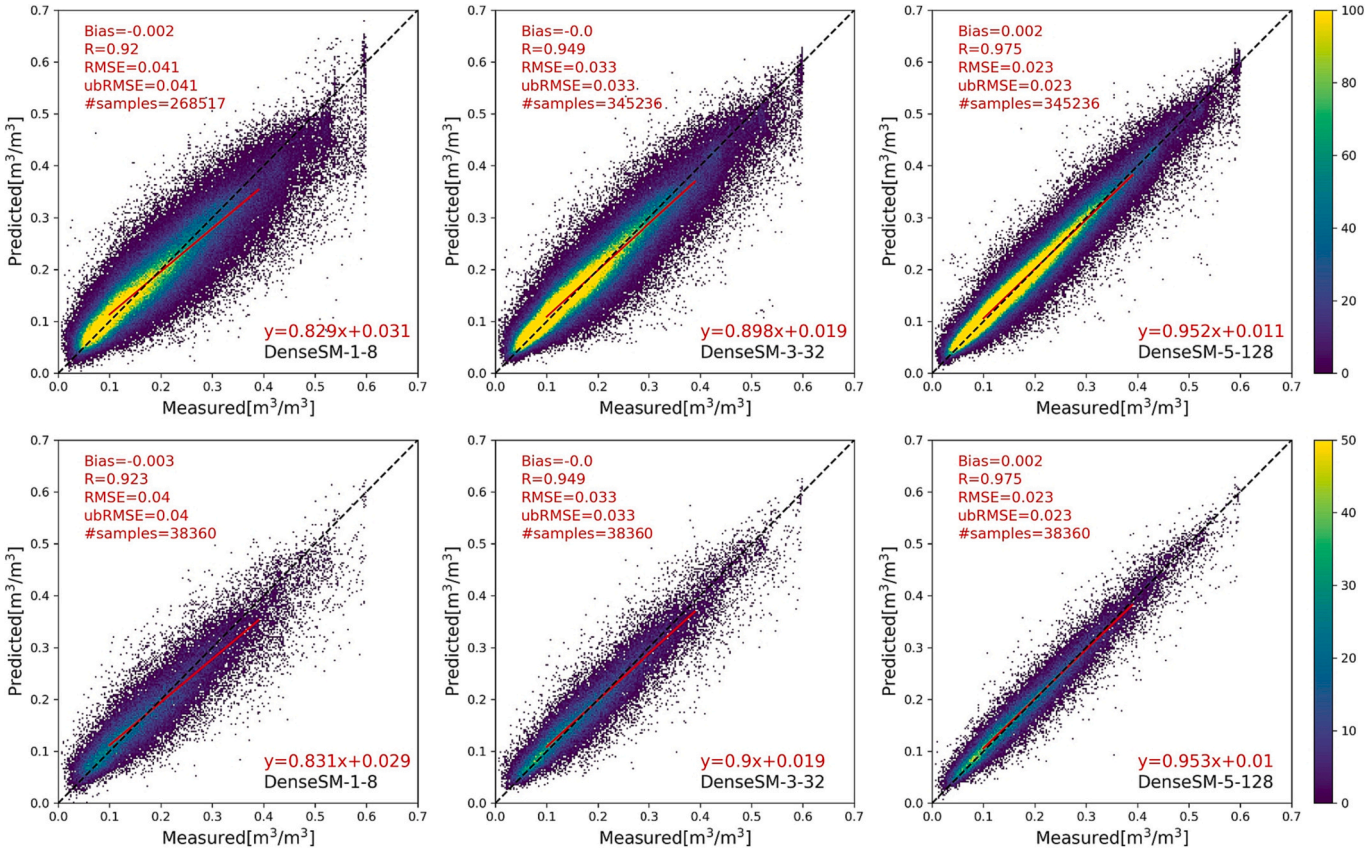

Fig. 4. The training (top) and validation (bottom) results of three pre-trained coarse models at 9 km, with the model width and depth increased from left to right. Notably, the scales of the color bars are different between the training and validation.

the general SAR retrieval target of 0.06 m^3/m^3. This suggests that all the trained DenseSMs well captured the complex relationships between SAR data and SMAP soil moisture at 9 km. Though the "students" (DenseSMs) cannot outperform their "teacher" (SMAP L3 9 km products) in the current training framework, they at least grained some valuable knowledge about the soil moisture estimation, being critical for reducing the required number of training samples at higher resolutions.

The performance of the coarse models on the three fine grids were also evaluated. The RMSEs of the 25 coarse models were 0.111–0.143 m^3/m^3, 0.106–0.120 m^3/m^3 and 0.105–0.114 m^3/m^3 for 0.1 km, 0.5 km and 1 km respectively, while the corresponding R was 0.258–0.398, 0.387–0.470 and 0.410–0.469 respectively. The ensemble of the 25 models (DenseSM-E) outperformed any single DenseSM-*d*-*w*, with the RMSE being 0.111 m^3/m^3, 0.106 m^3/m^3 and 0.105 m^3/m^3 for 0.1 km, 0.5 km and 1 km respectively (Fig. 5). An RMSE of >0.1 m^3/m^3 is much larger than the target of 0.06 m^3/m^3, being only slightly better than a guess of using the average value of the ground truth (RMSE: 0.110 m^3/m^3 for 1 km). Accordingly, the direct use of 9 km models for higher resolution estimation is questionable. However, the 9 km model performed much better than a random guess in view of R (> 0.4), being a valuable starting point for high resolution soil moisture retrievals.

4.2. Evaluation in the dependent scenario

The performance of the transfer DenseSM-E in the dependent scenario was depicted in Fig. 6. Since the retrieval bias of all three grids (0.1, 0.5 and 1 km) was near zero in all cases, the ubRMSE was very close to the corresponding RMSE and thus was not included. As expected, the accuracy of all the models was improved as the ratio of training samples increased. A compatible R of 0.71 and RMSE of 0.076 m^3/m^3 was achieved using only 936 training samples (0.5% of the whole sample set), improving to 0.83 and 0.06 m^3/m^3 respectively for a training ratio of 3%. The models of 500 m resolution achieved the best results, followed by that of 1 km and 0.1 km. However, the difference in R and RMSE among the three grid resolutions were negligible, being <0.008 and 0.001 m^3/m^3 respectively. Since the performance at the three grid resolutions were generally consistent, only the results of 1 km are presented below, being comparable to most existing studies on retrieving soil moisture from Sentinel-1 (Balenzano et al., 2021; Bauer-Marschallinger et al., 2018; Zhu et al., 2022; Zhu et al., 2023).

The R and RMSE of the 25 DenseSM-*d*-*w* in the transfer and benchmark modes are depicted in Fig. 7a. The transfer models outperformed the corresponding benchmarks in all cases, and even the worst transfer DenseSM-*d*-*w* achieved better results than the benchmark DenseSM-*d*-*w* for a training ratio of <3%, confirming the effectiveness of the transfer learning. The benefit of using transfer learning decreased when using more training samples, with an overlap for a training ratio > 3%. This can be explained by the fact that more training samples can better represent the whole sample set, and thus knowledge transferred from the coarse models became less critical. Moreover, the performance ranges of the transfer DenseSM-*d*-*w* were smaller than those of the benchmark models, suggesting that the transfer learning can also enhance the stability of models. Fig. 7a also provides a comparison of the DenseSM-E and DenseSM-*d*-*w* in the transfer and benchmark modes. Here the DenseSM-E outperformed any a single DenseSM-*d*-*w* in both the transfer mode and the benchmark modes, confirming the effectiveness of the ensemble concept used in this study.

The accuracy statistics of each 1 km grid cell were calculated for the DenseSM-E (Fig. 7b), with the results of each main landcover and climate type being provided in the supplementary material (Fig. S2 and S3), showing a similar dependence on the size of training sets. The models achieved a near zero median bias for all cases expect the case of directly using the coarse models (0%), but the range of bias reduced –substantially as training samples increased, indicating a stable overall bias and a reduced spatial heterogeneity in bias. The median of the grid cell-wise R ranged from 0.652 to 0.758 for the transfer DenseSM-E, being much smaller than that of the overall R (0.704–0.853, Fig. 7a). This was also true for the benchmark DenseSM-E. The median RMSE achieved in the transfer mode was <0.06 m^3/m^3 when the training ratio reached 2.5%, while the median ubRMSE was <0.06 m^3/m^3 even for a training set of 0% (i.e., the coarse models). In contrast, the benchmark DenseSM-E also achieved improved statistics for a large training set but performed worse than the coarse models for a training ratio of 0.5% in view of R and ubRMSE. This suggests that the knowledge provided by the coarse models can be more valuable than that from a small training set.

The performance difference between the transfer and benchmark DenseSM-E in the case of 0.5% training samples was calculated, with the results of the CONUS and Europe being depicted in Fig. 8. The transfer DenseSM-E achieved higher R than the benchmark on 92.1% and 92.6% of the grid cells in the CONUS and Europe respectively, while the grid cells with a lower RMSE accounted for 80.0% and 78.2%. The red points, being where the performance deteriorated in the transfer mode, were generally distributed randomly in the CONUS. While the grid cells in north Europe (> 50° N) seemingly had worse results in the transfer mode, the limited number of grid cells are insufficient to make convincing conclusions.

4.3. Evaluation in the on-site scenario

The overall results of the on-site scenario were depicted in Fig. 9, having similar patterns to that of the dependent scenario. The accuracy

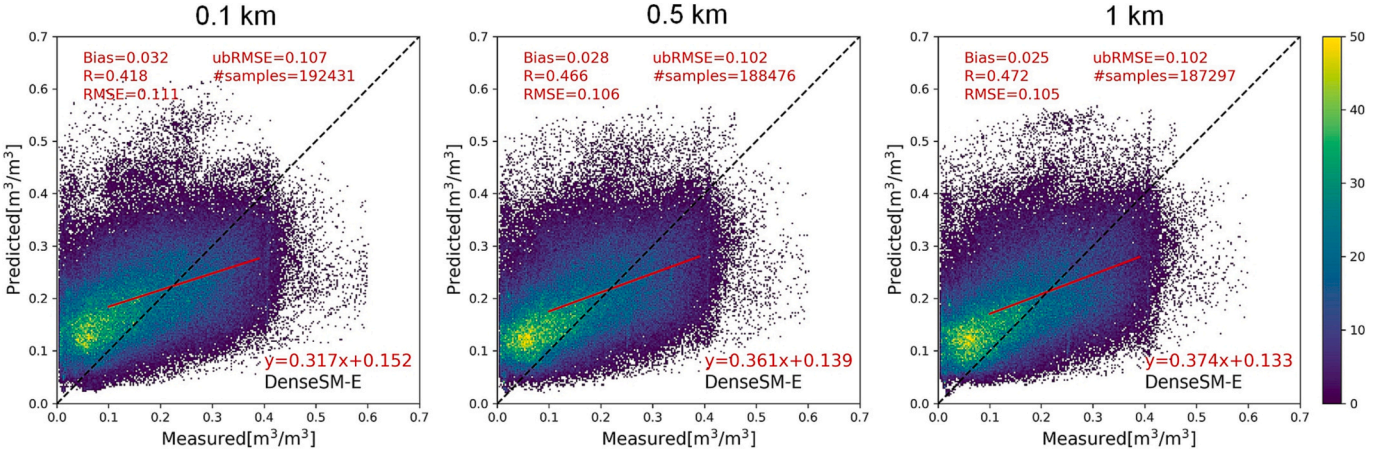

Fig. 5. The performance of the 9 km DenseSM-E at 0.1 km, 0.5 km and 1 km.

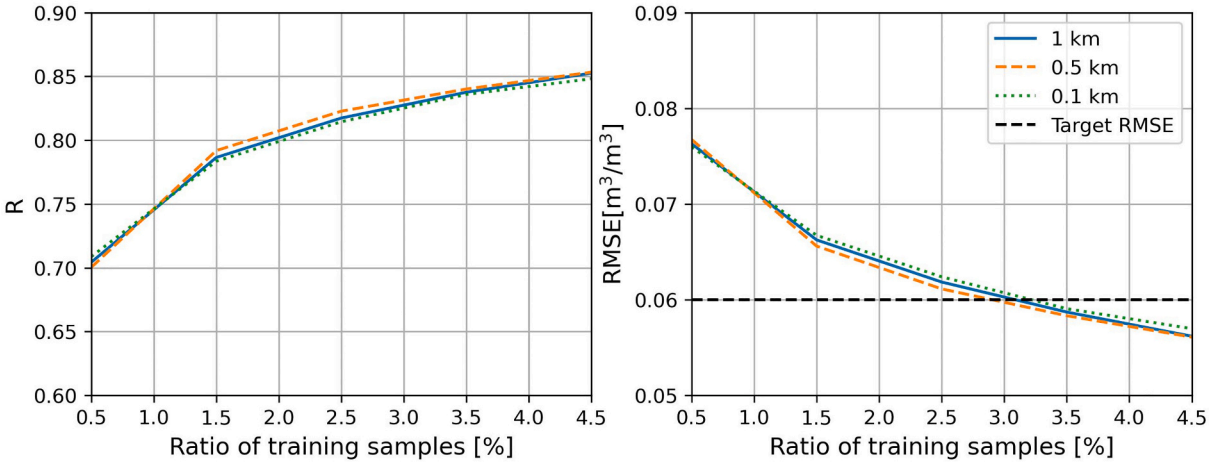

Fig. 6. Performance of the transfer DenseSM-E at 0.1, 0.5, and 1 km in the dependent scenario. The retrieval bias was near zero in all cases and the ubRMSE was very close to the corresponding RMSE, being thus not included.

of all models was improved using more samples and changed little when the number of training samples of each grid cell was >9. While the benefit from the transfer learning slightly decreased using an increasing number of training samples, it was much larger than the benefit observed in the dependent scenario. For example, the ΔR of the transfer and benchmark DenseSM-E was 0.061 for a training set of 7007 (11 per grid cell × 637 grid cells) in the on-site scenario, while the ΔR for a similar size of training set (3.5%, 6610) was 0.031 in the dependent scenario. However, the transfer DenseSM-E only achieved an RMSE of 0.081–0.068 m^3/m^3 for a grid cell specific sample size of ≤11, requiring more samples to reach the target of 0.06 m^3/m^3. This is related to the fact that the independent on-site scenario is more challenging than the dependent scenario. The models can benefit more from the transfer learning but require more training samples to reach a higher accuracy.

Fig. 9b shows a comparison of the benchmark and transfer DenseSM-E in view of grid cell-wise statistics, with the comparison for each landcover and climate type being provided in the supplementary material (Fig. S4 and Fig. S5). The transfer DenseSM-E achieved relatively smaller biases than the benchmark DenseSM-E in all cases, but the median biases were all close to 0. The transfer DenseSM-E achieved significant better results than the benchmark DenseSM-E in R, RMSE and ubRMSE, with the improvements in the median values being 0.119–0.237, 0.009–0.014 m^3/m^3 and 0.007–0.013 m^3/m^3 respectively. The R and ubRMSE of the transfer DenseSM-E was observed to be less sensitive to the number of samples per grid cell, with the median ubRMSE being all <0.06 m^3/m^3. In contrast, the median RMSE decreased significantly when the grid cell-wise samples were < 7 and gradually approached the target of 0.06 m^3/m^3 for a larger sample size. Similar to the results of the dependent scenario (Fig. 7b), the improvement in RMSE was mainly from the reduced biases across the grid cells.

Fig. 10 shows the year specific accuracy statistics of the benchmark and transfer DenseSM-E, with the evaluation on 2021 being also included for comparison. Not surprisingly, the models achieved the best results in 2021 as the training samples were from the same year. The models based on >5 grid cell-wise training samples met the target of 0.06 m^3/m^3. For the period of independent validation (2016–2020), the benchmark DenseSM-E achieved the best results in 2016, followed by 2018, 2017, 2020 and 2019. The largest difference in R and RMSE were observed between 2016 and 2019, being 0.050 and 0.010 m^3/m^3 respectively. In contrast, the transfer DenseSM-E achieved the best results in 2020, followed by 2016, 2018, 2017 and 2019, with the largest difference among the 5 years being 0.033 and 0.004 m^3/m^3 in R and RMSE respectively. The small interannual variation of the performance suggests a good stability of the transfer DenseSM-E, although the results differ from expectations that a small time-gap between training and validation set can lead to better results. The limited interannual difference also suggested that the retrieval scenarios of the five validation years were similar and that a model trained for either calendar year can have stable validation results on the remaining years. This is not surprising because the climate and surface conditions change little in 5 years globally. However, surface conditions (e.g., crop types) of a small area can have substantial changes and the findings of this section can be not valid.

4.4. Evaluation in the off-site scenario

The off-site validation results of the SCAN sites are depicted in Fig. 11a. Similar to the results of the other two scenarios, the performance of all models was improved using more samples (grid cells here) in the training, with marginal benefits when the number of grid cells was >5. The 25 benchmark DenseSM-*d-w* achieved a median RMSE of 0.195–0.095 m^3/m^3, being much worse than their counterparts in the transfer mode and the coarse DenseSM-E. Moreover, the benchmark DenseSM-*d-w* obtained a low inversion rate of 0.40–0.89, suggesting a substantial overfitting on the training set. This also suggests that direct training of the DenseSM-*d-w* cannot result in acceptable models if the training set was from a small number of grid cells. However, while the ensemble of them (i.e., the benchmark DenseSM-E) led to better accuracy and an inversion rate of 1, it was still worse than the coarse DenseSM-E when the number of grid cells was <17. In contrast, the transfer DenseSM-E achieved an acceptable R and RMSE of 0.653–0.767 and 0.102–0.078 m^3/m^3, respectively, being only worse than the coarse DenseSM-E in the case of 1 grid cell. This confirms the benefit of using transfer learning in the off-site scenario, but the transfer learning based on samples from only 1 grid cell can deteriorate the performance of the pre-trained models through overfitting.

The off-site validation results of the SNOTEL sites are provided in Fig. 11b, with the patterns of the four metrics being similar to that of the SCAN sites (Fig. 11a). However, all the models achieved much poorer results over the SNOTEL sites. The R and RMSE of the transfer DenseSM-E was 0.271–0.509 and 0.126–0.098 m^3/m^3. The benefit in RMSE from transfer learning was marginal for DenseSM-E, being 0.004–0.012 m^3/m^3, while a relatively large improvement was observed in R, being 0.101–0.222. Similarly, the transfer DenseSM-E only slightly outperformed the coarse DenseSM-E in RMSE (< 0.012 m^3/m^3) when >1 grid cells were involved in the training, while relatively large improvement was made in R, being up to 0.239.

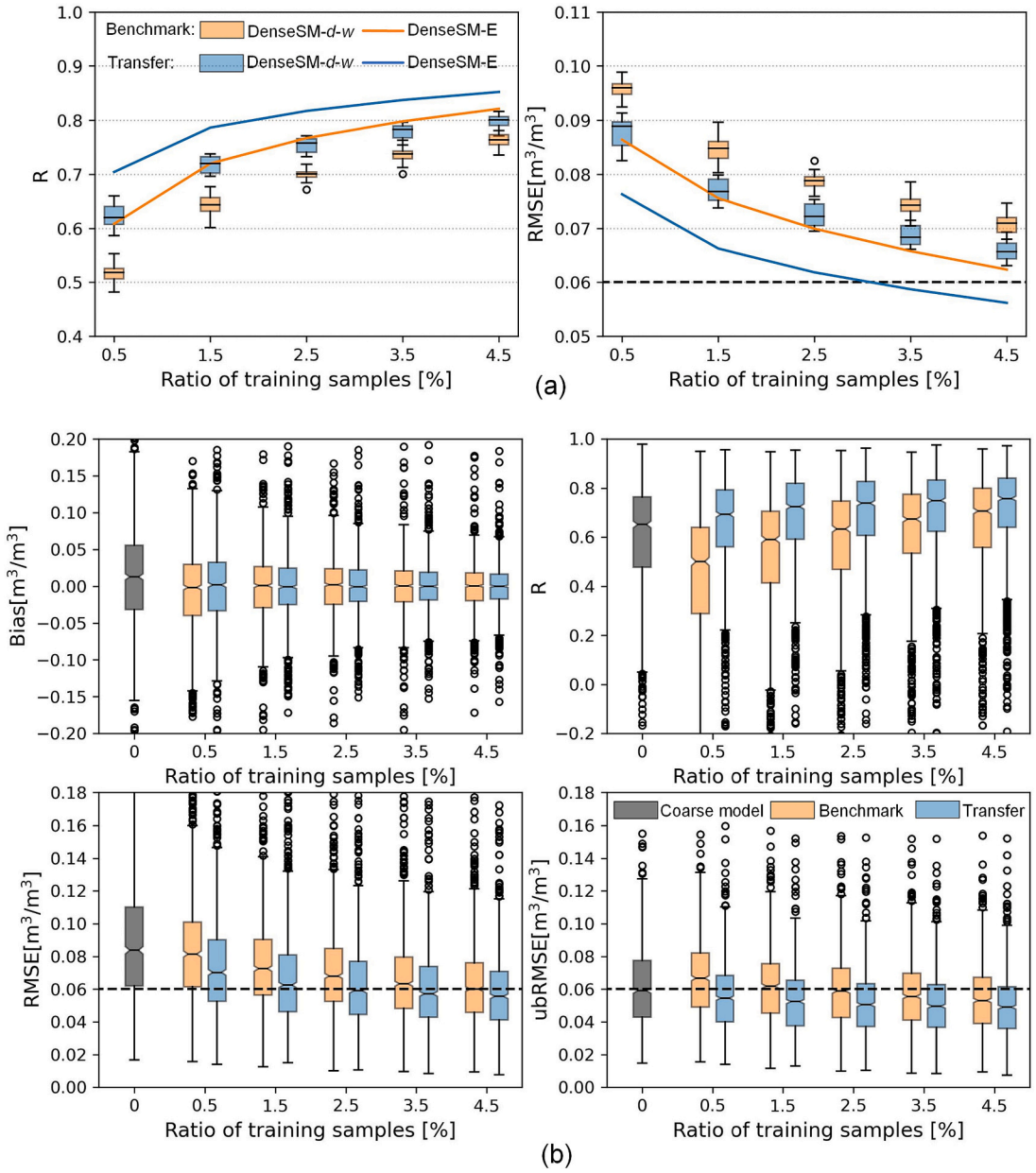

Fig. 7. The overall performance of the DenseSM-*d-w* and DenseSM-E on all validation samples (a) and the grid cell-wise statistics achieved by the DenseSM-E (b). All the models achieved a near zero overall bias and thus the RMSE and ubRMSE were similar, being not presented in (a). Some outliers in (b) were not depicted to have a proper range in the y-axis. The black dashed lines are the target accuracy of 0.06 m^3/m^3. The grid cell-wise RMSE of main landcover (Fig. S2) and climate types (Fig. S3) can be found in the Supplementary documents.

5. Discussion

The transfer learning framework proposed here starts from building coarse models using the SMAP L3 passive soil moisture products, resulting in 25 robust pre-trained models at 9 km resolution. Despite the high training and validation accuracy (Fig. 4), they could not outperform their teacher and inherit the uncertainties of the SMAP products, e.g., lower accuracy over densely vegetated areas (O'Neill et al., 2012). Moreover, the 9 km SMAP enhanced L3 radiometer soil moisture was resampled from the 36 km products, being not able to fully reflect the soil moisture at 9 km. However, even moderate coarse models can help as they can provide valuable knowledge about the nonlinear relationships between Sentinel-1 and soil moisture. Further improvements to the pre-trained models are expected to further improve results, which can include 1) learning from multiple teachers (Liu et al., 2022), e.g., the low-resolution soil moisture from land surface models or other satellites; and 2) extending the training set from the current grid cells to a larger area. Moreover, the root mean square difference between the estimation from the coarse model and the SMAP products was <0.03 m^3/m^3. The coarse model may thus be used to enhance the spatial-temporal continuity of the archived SMAP products (Fang et al., 2019; Fang et al., 2017).

The direct use of the pre-trained coarse models at higher resolutions of 0.1–1 km achieved poor results, being far away from the target RMSE of 0.06 m^3/m^3. This is different from other studies focusing on downscaling (Xu et al., 2022; Zhao et al., 2022). While marginal, the coarse models achieved better results at the scales which are closer to the training scale (Fig. 5), suggesting a scale dependence of the proposed DenseSM. However, negligible difference was observed among the three fine grids in the transfer learning (Fig. 6), being different from the

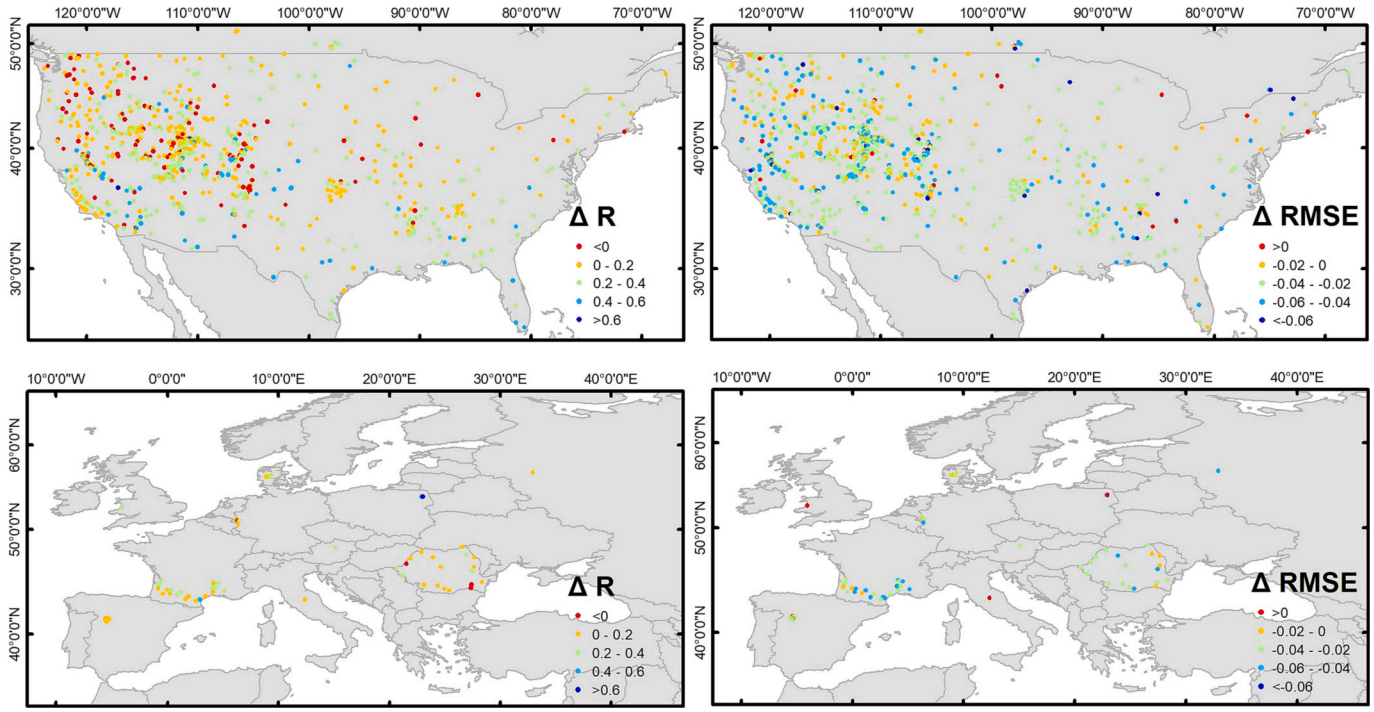

Fig. 8. The benefit of using transfer learning in the continental US and Europe in the case of 0.5% training samples. ΔR and ΔRMSE were calculated by subtracting the value of the benchmark from that of the transfer DenseSM-E. Notably, the color of ΔR and ΔRMSE are flipped and thus the orange to blue points are these benefited from the transfer learning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

expectation that soil moisture retrieval at a higher resolution is more challenging. Potential reasons can be that 1) the transfer learning phase was made independently on each scale and the models have been optimized for each scale; 2) the validation was based on in-situ measurements and the mismatch between the point scale and the three grid resolutions may introduce large uncertainties, though a few grid cells contained >1 stations (Balenzano et al., 2021); 3) the auxiliary input data at 0.1 km, 0.5 km and 1 km can be similar due to the coarse resolutions of NDVI and soil texture; and 4) deep learning models are less sensitive to the noise compared to the methods based on scattering models.

The transfer models were evaluated in three scenarios. Generally, the reported accuracy statistics of the dependent scenario were superior to those of most existing change detection methods (Balenzano et al., 2011; Palmisano et al., 2020; Wagner et al., 1999; Zhu et al., 2022) and inversion methods based on scattering models (Kim et al., 2014a; Kim et al., 2012; Zhu et al., 2019a, 2019b) using only a few thousand training samples. This demonstrates the effectiveness of the proposed multiscale transfer learning scheme for areas and periods with sparse ground samples. Since the ground in-situ measurements of a specific area are generally collected discontinuous in time and space, the three validation scenarios were simplified cases of real applications and cannot provide a clear conclusion of the real benefit. While the number of required samples can vary from site to site, Fig. 7 provides an empirical suggestion that 5000–6000 samples can be sufficient for achieving a target RMSE of 0.06 m^3/m^3. In contrast, a more straightforward recommendation can be made for areas where stations only worked for a short period. The on-site scenario suggested that 5 samples from the same year are sufficient to train a robust model for soil moisture retrieval at the same site in other periods, with a competitive R and RMSE of 0.75 and 0.07 m^3/m^3, respectively. The proposed method can also improve soil moisture retrieval in the off-site scenario. The transfer models for the SCAN sites based on 17 grid cells (~10% of all stations) achieved a moderate accuracy (R: 0.767 and RMSE: 0.078 m^3/m^3), outperforming the benchmark by 10%. However, the transfer learning failed on the SNOTEL sites, with the use of 17 SNOTEL grid cells (~20% of all stations) in training only achieving an RMSE of 0.098 m^3/m^3. Accordingly, future studies should be focused on an enhanced spatial generalization capability. Apart from the poorer spatial generalization capability of the proposed method in the mountainous areas, the poorer results from SNOTEL sites may be related to the coarse resolutions of the input data, being insufficient to capture the complex local vegetation, terrain and climate features. Moreover, the real soil moisture of a 0.1 or 1 km grid cell can deviate from the in-situ measurements of a point. Unfortunately, the uncertainties caused by the spatial mismatch between a point measurement and the grid cell cannot be made in this study as most grid cells only have one station. A recent study of Balenzano et al. (2021) suggested that the spatial representativeness error can be 0.02 m^3/m^3 for a flat area if only 1 station is available for the retrieval at 1 km. However, the spatial representativeness error can be much larger over mountainous areas considering the larger spatial variation of soil moisture. Accordingly, the evaluation results and conclusions are potentially affected by the uncertainty of the point soil moisture measurements, with the uncertainty of sensor calibration typically around 0.04 m^3/m^3 (Smith et al., 2012), while the uncertainty caused by the varying measuring depth (0–5 cm, 0–3 cm, 5 cm) across the networks still unclear. The target of this study is daily averaged soil moisture, but the root mean square difference between the descending soil moisture and the daily average values for the whole sample set was calculated as being only 0.0027m^3/m^3 and for ascending acquisitions only 0.0028 m^3/m^3, making it negligible in both cases.

The three scenarios were designed for a comprehensive evaluation of the spatial-temporal imbalance of ground observations (Table 1 and Fig. 1). However, the results were mainly focused on Europe and the CONUS as most samples were from these areas. The results of the on-site and off-site suggested that some scenario specific ground measurements are still required. Accordingly, sophisticated retrieval models were not made for global high-resolution application in this study, but rather the mode of "pre-trained + scenario" models were encouraged. In other words, users are encouraged to reuse pre-trained coarse models and

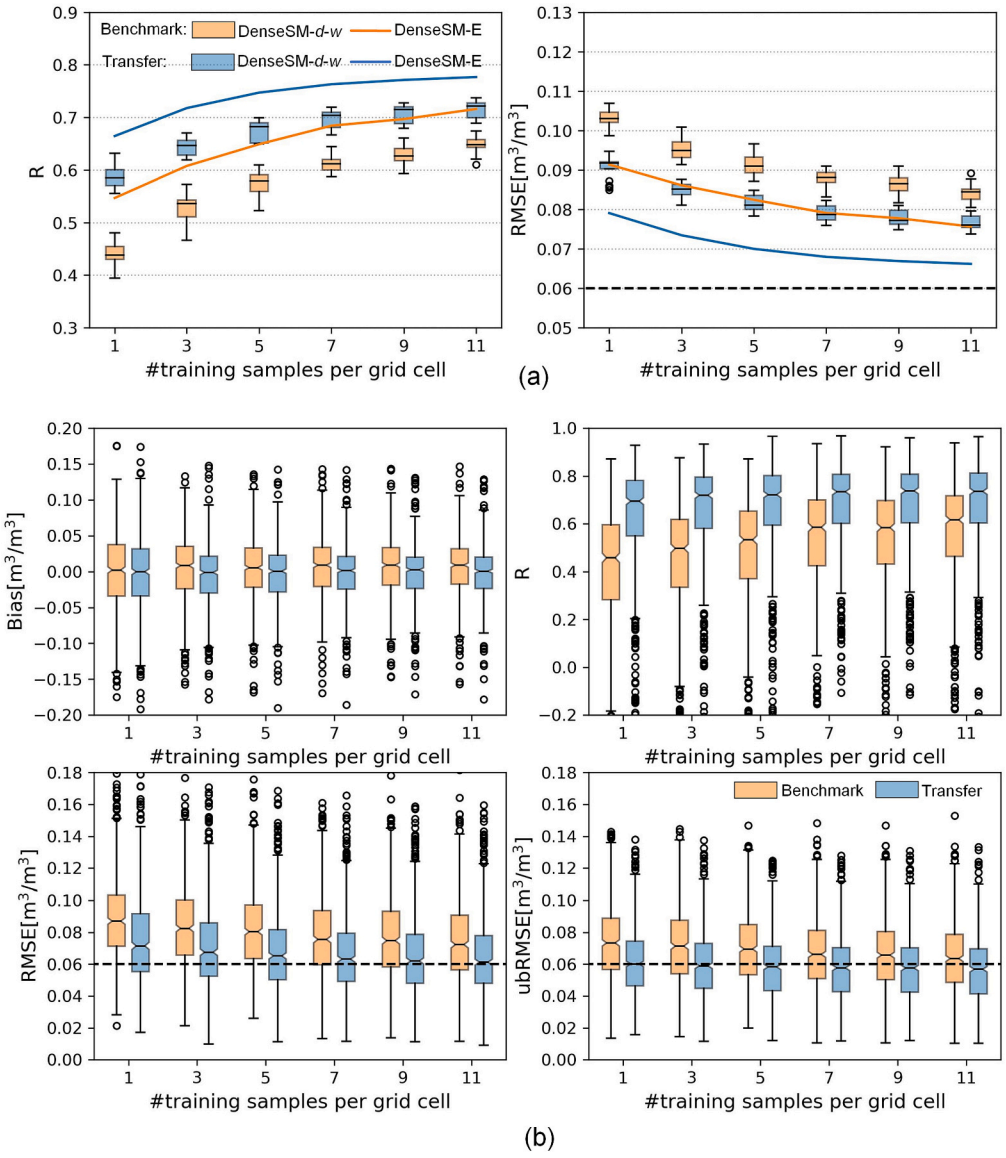

Fig. 9. Same as Fig. 7 but for the on-site scenario. The grid cell-wise statistics of 637 grid cells were presented in (b), with the validation samples being all from 2016 to 2020.

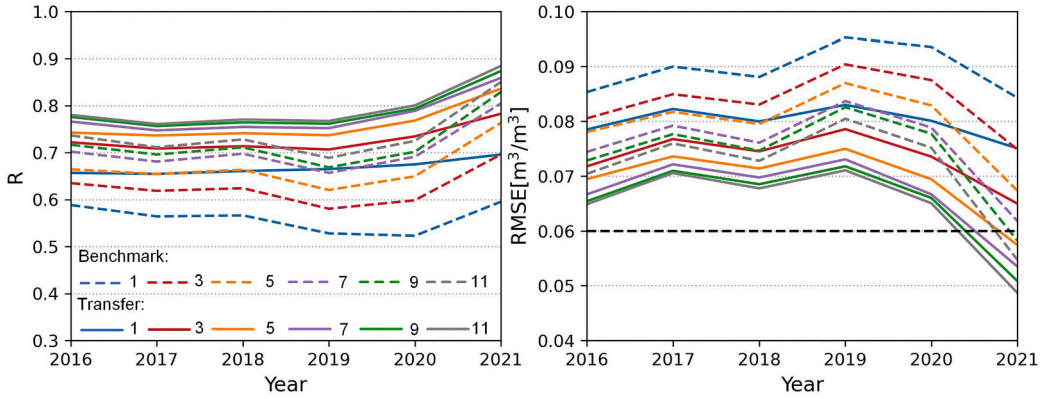

Fig. 10. Year specific performance achieved in the on-site scenario, with 1 to 11 being the number of training samples per grid cell selected from 2021. Validation was also made on 2021, being a special case similar to the dependent scenario.

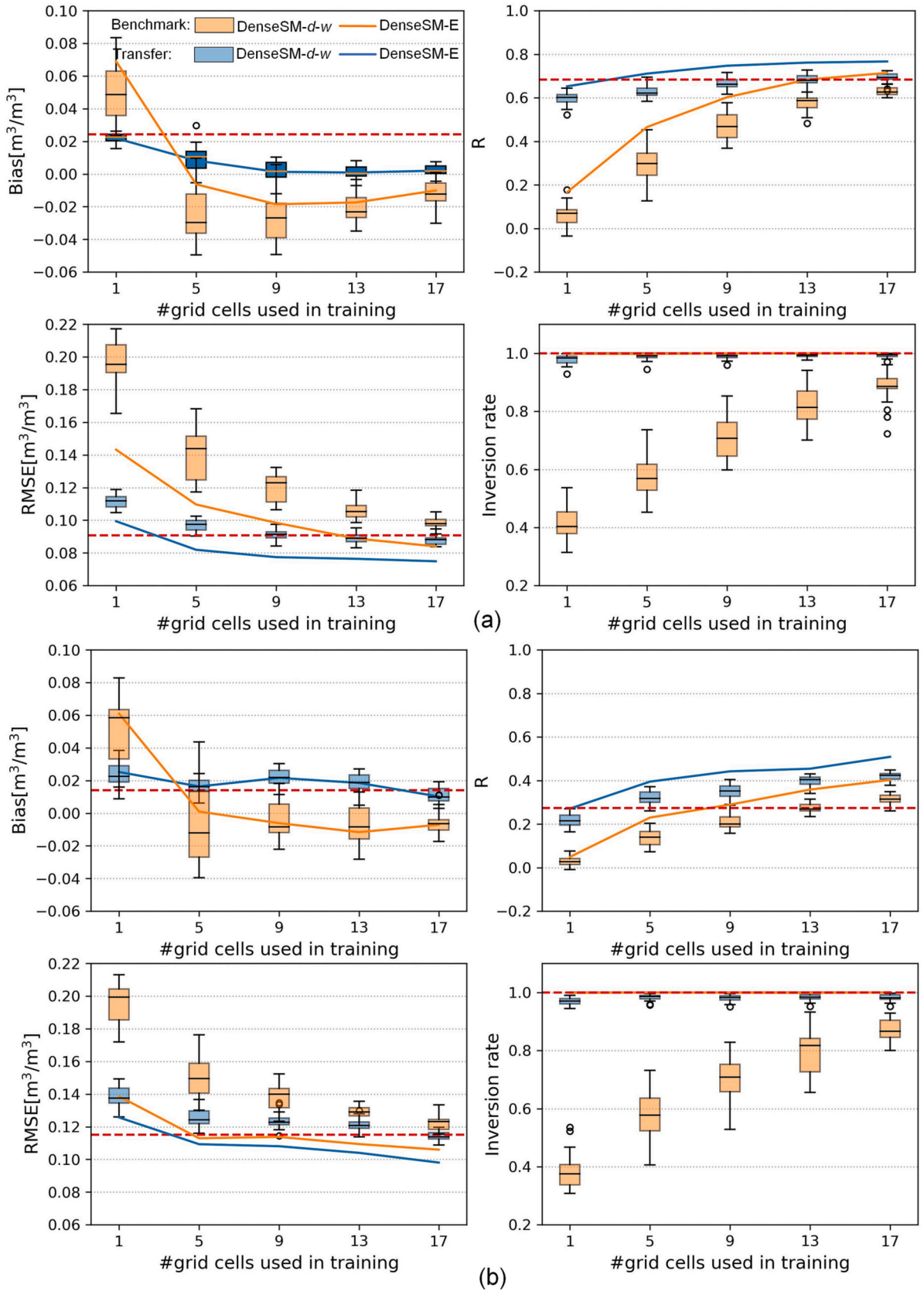

Fig. 11. The overall performance over the SCAN (a) and SNOTEL (b) sites in the off-site scenario. The inversion rate of the DenseSM-E was 1 in all case and thus lines were overlapped. The red dashed line is the value achieved by the coarse DenseSM-E. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

finetune them for the area and/or period of interest based on a few samples. The main contributions of this study are thus the pre-trained coarse models at 9 km and the multiscale transfer learning framework. The pre-trained coarse models are shared at https://github.com/rszlj/Transfer-DenseSM-E for direct use. The proposed architecture and the transfer framework were implemented using PyTorch as the backend. The codes to construct models, apply training and finetune are also provided, allowing reproduction of the presented results and further developments.

The benefit from transfer learning was determined by a comparison with the coarse models and the benchmark models, without a direct comparison with other state-of-art methods, considering the special scenarios of limited training samples. Generally, the proposed transfer framework achieved compatible accuracy statistics with other studies that used more than half of the samples in training. For example, Karthikeyan and Mishra (2021) proposed an XGBoost method to achieve multi-layer soil moisture using 80% of samples in training, leading to a regional-wise median R and ubRMSE of 0.772–0.950 and 0.024–0.054 m^3/m^3 for the top 5 cm. A more recent study achieved an average per-sensor R of 0.707 and ubRMSE of 0.055 m^3/m^3 using a similar dataset and 60% of samples in training (Batchu et al., 2023). Apart from the different training/validation scheme and the target resolution, the varying way of calculating accuracy statistics made the comparison even more challenging. For example, the median of the grid cell-wise R was 0.652–0.758 for the transfer DenseSM-E (Fig. 7b), being much smaller than the overall R of 0.704–0.853 (Fig. 7a). This is known as the Simpson's paradox, that a trend appearing in the whole data set can disappear or reverse when observed on each group (Sharma et al., 2022).

The proposed DenseSM was built on the sophisticated architecture of DenseNet, which has been used for SMAP soil moisture downscaling (Xu et al., 2022). Since an entire year of NDVI, 2-m temperature and precipitation preceding the retrieval date were used, the different phenologies of vegetation types and climate features were extracted as high-level features in the TempCNN block, with their relationships to soil moisture learned automatically. The TempCNN block works as a vegetation and climate type classifier (Zhu et al., 2021), inherently using the information related to the landcover and climate without requiring extra climate type or landcover maps. However, other surface parameters, e.g., soil organic matter content (Park et al., 2021) and soil surface temperature, can potentially be helpful and should be considered in a future version. In view of architecture, the DenseSM is somewhat old fashioned as more powerful architectures have been used in soil moisture retrieval and downscaling, e.g., the long short-term memory (Li et al., 2022), 2-dimensional CNN blocks (Batchu et al., 2023). However, the use of them is not straightforward for Sentinel-1. Sentinel-1 is the first C-band SAR mission that routinely provides global SAR observations every 6 to 12 days. However, the real temporal resolution varies in time and space, especially after the failure of Sentinel-1B in late 2021. A LSTM architecture will therefore suffer from the differing sampling rates. The 2-dimensional CNNs have demonstrated capability to extract high level image features. However, they are generally extremely time-consuming and the role of high-level image features in SAR soil moisture retrieval is still unclear as the scattering process of each pixel is generally independent.

The transfer learning method (finetune) used in this study is a fundamental method of the instance-based deep learning methods (Zhuang et al., 2020), without utilizing the distribution difference between the source and target domain. The difference between the source and target domain can be the main reason of the poor performance observed in the off-site scenario (Fig. 11). When finetuning a model for a small number of grid cells the learned features of the pre-trained models can be destroyed at early epochs and result in values outside the soil moisture bound. Accordingly, further improvements can be made by integrating such information into the cost function, e.g., the use of Maximum Mean Discrepancy (MMD) between distributions. Apart from the width and depth of the DenseSM-E, a few hyperparameters are required, e.g., the learning rate and the parameters for the Conv1D. Tunning these parameters is challenging as each sample is too valuable to be used in hyperparameter optimization. Accordingly, the hyperparameters used in this study were from existing studies (Zhu et al., 2021) and the large variations of the performance of the 25 DenseSM-*d*-w (Figs. 6, 8 and 10) has demonstrated the large uncertainty caused by the fixed hyperparameters. However, this uncertainty was greatly reduced by the ensemble of the 25 models.

6. Conclusion

A cross-resolution deep transfer learning framework was proposed for transferring the knowledge learned at 9 km to higher resolutions, with the performance being evaluated in three scenarios using a global dataset of ~190,000 samples extracted from the ISMN. The main conclusions are: 1) the SMAP 9 km soil moisture enabled the development of robust pre-trained models at 9 km resolution, providing valuable knowledge to be learned at 0.1, 0.5 and 1 km resolutions; 2) the transfer learning based on pre-trained models and 5000–6000 random in-situ soil moisture measurements resulted in reliable soil moisture retrieval at 0.1–1 km, with an RMSE of <0.06 m^3/m^3; 3) samples from a short period (2–4 months for Sentinel-1) can support the training of compatible models for a long independent period, with an overall RMSE of 0.068 m^3/m^3; 4) transfer learning using 10% stations improved the retrieval accuracy over areas without ground samples, achieving an acceptable RMSE of ~0.078 m^3/m^3 over the remaining 90% of the stations; and 5) the proposed transfer learning failed for mountainous areas without ground samples, being only slightly better than the guess of using the average value of ground measurements.

CRediT authorship contribution statement

Liujun Zhu: Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing. **Junjie Dai:** Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Yi Liu:** Conceptualization, Writing – original draft, Writing – review & editing. **Shanshui Yuan:** Writing – review & editing. **Tianling Qin:** Writing – review & editing. **Jeffrey P. Walker:** Writing – review & editing.

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.

Data availability

Data and codes will be shared at https://github.com/rszlj/Transfer-DenseSM-E after the acceptance

Acknowledgments

This work was supported by National Key Research and Development Program (2023YFC3209800), the National Natural Science Foundation of China (42101374, 42371369 and 52121006) and the Natural Science Foundation of Jiangsu Province (BK20210377). The authors express sincere thanks to the data providers and the International Soil Moisture Network for the network data listed in Table 1.

Credit author statement

Liujun Zhu: Conceptualization, Writing-Original draft preparation. **Junjie Dai**: Methodology, Writing-Reviewing and Editing. **Yi Liu**: Conceptualization, Writing-Original draft preparation, **Shanshui Yuan**:

Writing-Original draft preparation, Writing-Reviewing and Editing. **Tianling Qin**: Writing-Reviewing and Editing. **Jeffrey P. Walker**: Writing-Reviewing and Editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2023.113944.

References

Ardö, J., 2012. A 10-year dataset of basic meteorology and soil properties in Central Sudan. Dataset Papers in Geosci 2013, 297973.

Balenzano, A., Mattia, F., Satalino, G., Davidson, M.W., 2011. Dense temporal series of C- and L-band SAR data for soil moisture retrieval over agricultural crops. Ieee J Selected Topics Appl Earth Observ Remote Sens 4, 439–450.

Balenzano, A., Mattia, F., Satalino, G., Lovergine, F.P., Palmisano, D., Peng, J., Marzahn, P., Wegmüller, U., Cartus, O., Dąbrowska-Zielińska, K., 2021. Sentinel-1 soil moisture at 1 km resolution: a validation study. Remote Sens. Environ. 263, 112554.

Batchu, V., Nearing, G., Gulshan, V., 2023. A deep learning data fusion model using Sentinel-1/2, SoilGrids, SMAP-USDA, and GLDAS for soil moisture retrieval. J. Hydrometeorol 24, 1789–1823.

Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., Modanesi, S., Massari, C., Ciabatta, L., Brocca, L., 2018. Toward global soil moisture monitoring with Sentinel-1: harnessing assets and overcoming obstacles. IEEE Trans. Geosci. Remote Sens. 57, 520–539.

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.

Blöschl, G., Blaschke, A., Broer, M., Bucher, C., Carr, G., Chen, X., Eder, A., Exner-Kittridge, M., Farnleitner, A., Flores-Orozco, A., 2016. The hydrological open air laboratory (HOAL) in Petzenkirchen: a hypothesis-driven observatory. Hydrol. Earth Syst. Sci. 20, 227–255.

Bogena, H., Montzka, C., Huisman, J., Graf, A., Schmidt, M., Stockinger, M., Von Hebel, C., Hendricks-Franssen, H., Van der Kruk, J., Tappe, W., 2018. The TERENO-Rur hydrological observatory: a multiscale multi-compartment research platform for the advancement of hydrological science. Vadose Zone J. 17, 1–22.

Bracaglia, M., Ferrazzoli, P., Guerriero, L., 1995. A fully polarimetric multiple scattering model for crops. Remote Sens. Environ. 54, 170–179.

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., Smets, B., 2020. Copernicus global land cover layers—collection 2. Remote Sens. 12, 1044.

Calvet, J.-C., Fritz, N., Berne, C., Piguet, B., Maurel, W., Meurey, C., 2016. Deriving pedotransfer functions for soil quartz fraction in southern France from reverse modeling. Soil 2, 615–629.

Celik, M.F., Isik, M.S., Yuzugullu, O., Fajraoui, N., Erten, E., 2022. Soil moisture prediction from remote sensing images coupled with climate, soil texture and topography via deep learning. Remote Sens. 14, 5584.

Chaudhary, S.K., Srivastava, P.K., Gupta, D.K., Kumar, P., Prasad, R., Pandey, D.K., Das, A.K., Gupta, M., 2022. Machine learning algorithms for soil moisture estimation using Sentinel-1: model development and implementation. Adv. Space Res. 69, 1799–1812.

Chen, L., Xing, M., He, B., Wang, J., Shang, J., Huang, X., Xu, M., 2021. Estimating soil moisture over winter wheat fields during growing season using machine-learning methods. Ieee J. Selected Topics Appl. Earth Observ. Remote Sens. 14, 3706–3718.

Cook, D.R., 2016. Soil temperature and moisture profile (STAMP) system handbook. In: DOE Office of Science Atmospheric Radiation Measurement (ARM) Program.

Didan, K., 2015. MOD13Q1 MODIS/Terra vegetation indices 16-day L3 global 250m SIN grid V006. NASA EOSDIS Land Processes DAAC 10.

Dong, L., Wang, W., Jin, R., Xu, F., Zhang, Y., 2023. Surface soil moisture retrieval on Qinghai-Tibetan plateau using Sentinel-1 synthetic aperture radar data and machine learning algorithms. Remote Sens. 15, 153.

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. Hydrol. Earth Syst. Sci. 25, 5749–5804.

Eldan, R., Shamir, O., 2016. The power of depth for feedforward neural networks. In: Conference on Learning Theory. PMLR, pp. 907–940.

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., 2010. The soil moisture active passive (SMAP) mission. Proc. IEEE 98, 704–716.

Fang, K., Shen, C., Kifer, D., Yang, X., 2017. Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. Geophys. Res. Lett. 44, 11,030-011,039.

Fang, K., Pan, M., Shen, C., 2019. The value of SMAP for long-term soil moisture estimation with the help of deep learning. IEEE Trans. Geosci. Remote Sens. 57, 2221–2233.

Feng, H., Liu, Y., 2015. Combined effects of precipitation and air temperature on soil moisture in different land covers in a humid basin. J. Hydrol. 531, 1129–1140.

Flammini, A., Morbidelli, R., Saltalippi, C., Picciafuoco, T., Corradini, C., Govindaraju, R. S., 2018. Reassessment of a semi-analytical field-scale infiltration model through experiments under natural rainfall events. J. Hydrol. 565, 835–845.

Fung, A.K., Li, Z., Chen, K., 1992. Backscattering from a randomly rough dielectric surface. IEEE Trans. Geosci. Remote Sens. 30, 356–369.

Galle, S., Grippa, M., Peugeot, C., Moussa, I.B., Cappelaere, B., Demarty, J., Mougin, E., Panthou, G., Adjomayi, P., Agbossou, E., 2018. AMMA-CATCH, a critical zone observatory in West Africa monitoring a region in transition. Vadose Zone J. 17, 1–24.

Gao, L., Gao, Q., Zhang, H., Li, X., Chaubell, M.J., Ebtehaj, A., Shen, L., Wigneron, J.-P., 2022. A deep neural network based SMAP soil moisture product. Remote Sens. Environ. 277, 113059.

González-Zamora, Á., Sánchez, N., Pablos, M., Martínez-Fernández, J., 2019. CCI soil moisture assessment with SMOS soil moisture and in situ data under different environmental conditions and spatial scales in Spain. Remote Sens. Environ. 225, 469–482.

Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J.E., Weinberger, K.Q., 2017a. Snapshot ensembles: train 1, get m for free. arXiv preprint arXiv:1704.00109.

Huang, G., Liu, Z., Pleiss, G., Van Der Maaten, L., Weinberger, K.Q., 2017b. Convolutional networks with dense connectivity. In: *Proceedings of the IEEE conference on computer vision and pattern recognitionIEEE transactions on pattern analysis and machine intelligence* (pp. 4700-4708).

Jensen, K.H., Refsgaard, J.C., 2018. HOBE: the Danish hydrological observatory. Vadose Zone J. 17, 1–24.

Karthikeyan, L., Mishra, A.K., 2021. Multi-layer high-resolution soil moisture estimation using machine learning over the United States. Remote Sens. Environ. 266, 112706.

Katz, D.M., Bommarito, M.J., Gao, S., Arredondo, P., 2023. Gpt-4 passes the bar exam. Available at SSRN, 4389233.

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.

Kim, S.-B., Tsang, L., Johnson, J.T., Huang, S., Van Zyl, J.J., Njoku, E.G., 2012. Soil moisture retrieval using time-series radar observations over bare surfaces. IEEE Trans. Geosci. Remote Sens. 50, 1853–1863.

Kim, S.-B., Moghaddam, M., Tsang, L., Burgin, M., Xu, X., Njoku, E.G., 2014a. Models of L-band radar backscattering coefficients over global terrain for soil moisture retrieval. IEEE Trans. Geosci. Remote Sens. 52, 1381–1396.

Kim, S.-B., van Zyl, J., Dunbar, S., Njoku, E., Johnson, J., Moghaddam, M., Shi, J., Tsang, L., 2014b. SMAP L2 & L3 Radar Soil Moisture (Active) Data Products Revision (A. In).

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kolassa, J., Reichle, R., Liu, Q., Alemohammad, S., Gentine, P., Aida, K., Asanuma, J., Bircher, S., Caldwell, T., Colliander, A., 2018. Estimating surface soil moisture from SMAP observations using a neural network technique. Remote Sens. Environ. 204, 43–59.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15, 259–263.

Lal, P., Singh, G., Das, N.N., Entekhabi, D., Lohman, R., Colliander, A., Pandey, D.K., Setia, R., 2023. A multi-scale algorithm for the NISAR mission high-resolution soil moisture product. Remote Sens. Environ. 295, 113667.

Larson, K.M., Small, E.E., Gutmann, E.D., Bilich, A.L., Braun, J.J., Zavorotny, V.U., 2008. Use of GPS receivers as a soil moisture network for water cycle studies. Geophys. Res. Lett. 35.

Li, Q., Wang, Z., Shangguan, W., Li, L., Yao, Y., Yu, F., 2021. Improved daily SMAP satellite soil moisture prediction over China using deep learning model with transfer learning. J. Hydrol. 600, 126698.

Li, Q., Zhu, Y., Shangguan, W., Wang, X., Li, L., Yu, F., 2022. An attention-aware LSTM model for soil moisture and soil temperature prediction. Geoderma 409, 115651.

Liu, J., Rahmani, F., Lawson, K., Shen, C., 2022. A multiscale deep learning model for soil moisture integrating satellite and in situ data. Geophys. Res. Lett. 49 e2021GL096847.

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.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., 2021. ERA5-land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst Sci. Data 13, 4349–4383.

Oh, Y., Sarabandi, K., Ulaby, F.T., 2002. Semi-empirical model of the ensemble-averaged differential Mueller matrix for microwave backscattering from bare soil surfaces. IEEE Trans. Geosci. Remote Sens. 40, 1348–1355.

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.

O'Neill, P., Chan, S., Njoku, E., Jackson, T., Bindlish, R., 2012. Soil moisture active passive (SMAP) algorithm theoretical basis document (ATBD). SMAP level 2.

O'Neill, P., Chan, S., Njoku, E., Jackson, T., Bindlish, R., Chaubell, J., Colliander, A., 2021. SMAP enhanced L3 radiometer global and polar grid daily 9 km ease-grid soil moisture version 5. In: NASA National Snow and Ice Data Center Distributed Active Archive Center (Ed.), Boulder, Colorado USA. DOI, 10.

Osenga, E.C., Vano, J.A., Arnott, J.C., 2021. A community-supported weather and soil moisture monitoring database of the roaring fork catchment of the Colorado River headwaters. Hydrol. Process. 35, e14081.

Palmisano, D., Mattia, F., Balenzano, A., Satalino, G., Pierdicca, N., Guarnieri, A.V.M., 2020. Sentinel-1 sensitivity to soil moisture at high incidence angle and the impact on retrieval over seasonal crops. IEEE Trans. Geosci. Remote Sens. 59, 7308–7321.

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. Hydrol. Earth Syst. Sci. 25, 6407–6420.
Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644.
Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M.H., Crow, W.T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M.W., 2020. A roadmap for high-resolution satellite soil moisture applications–confronting product characteristics with user requirements. Remote Sens. Environ. 112162.
Petropoulos, G.P., McCalmont, J.P., 2017. An operational in situ soil moisture & soil temperature monitoring network for West Wales, UK: the WSMN network. Sensors 17, 1481.
Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7, 217–240.
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.
Saarela, M., Jauhiainen, S., 2021. Comparison of feature importance measures as explanations for classification models. SN Appl. Sci. 3, 1–12.
Schaefer, G.L., Cosh, M.H., Jackson, T.J., 2007. The USDA natural resources conservation service soil climate analysis network (SCAN). J. Atmos. Ocean. Technol. 24, 2073–2077.
Service, U.N.R.C, 2022. SNOwpack TELemetry network (SNOTEL).
Sharma, R., Garayev, H., Kaushik, M., Peious, S.A., Tiwari, P., Draheim, D., 2022. Detecting simpson's paradox: a machine learning perspective. In: *Database and Expert Systems Applications: 33rd International Conference, DEXA 2022, Vienna, Austria, August 22–24, 2022, Proceedings, Part I* (pp. 323-335). Springer.
Smith, A., Walker, J., Western, A., Young, R., Ellett, K., Pipunic, R., Grayson, R., Siriwardena, L., Chiew, F., Richter, H., 2012. The Murrumbidgee soil moisture monitoring network data set. Water Resour. Res. 48, W07701.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn Res. 15, 1929–1958.
Su, Z., Wen, J., Dente, L., Van Der Velde, R., Wang, L., Ma, Y., Yang, K., Hu, Z., 2011. The Tibetan plateau observatory of plateau scale soil moisture and soil temperature (Tibet-Obs) for quantifying uncertainties in coarse resolution satellite and model products. Hydrol. Earth Syst. Sci. 15, 2303–2316.
Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A survey on deep transfer learning. In: In, *International Conference on Artificial Neural Networks*. Springer, pp. 270–279.
Wagner, W., Lemoine, G., Rott, H., 1999. A method for estimating soil moisture from ERS scatterometer and soil data. Remote Sens. Environ. 70, 191–207.
Wu, Z., Shen, C., Van Den Hengel, A., 2019. Wider or deeper: revisiting the resnet model for visual recognition. Pattern Recogn. 90, 119–133.
Xu, M., Yao, N., Yang, H., Xu, J., Hu, A., de Goncalves, L.G.G., Liu, G., 2022. Downscaling SMAP soil moisture using a wide & deep learning method over the continental United States. J. Hydrol. 609, 127784.
Yadav, V.P., Prasad, R., Bala, R., Vishwakarma, A.K., 2020. An improved inversion algorithm for spatio-temporal retrieval of soil moisture through modified water cloud model using C-band sentinel-1A SAR data. Comput. Electron. Agric. 173, 105447.
Yinglan, A., Wang, G., Hu, P., Lai, X., Xue, B., Fang, Q., 2022. Root-zone soil moisture estimation based on remote sensing data and deep learning. Environ. Res. 212, 113278.
Yu, J., Zhang, X., Xu, L., Dong, J., Zhangzhong, L., 2021. A hybrid CNN-GRU model for predicting soil moisture in maize root zone. Agric. Water Manag. 245, 106649.
Zhao, T., Shi, J., Lv, L., Xu, H., Chen, D., Cui, Q., Jackson, T.J., Yan, G., Jia, L., Chen, L., 2020. Soil moisture experiment in the Luan River supporting new satellite mission opportunities. Remote Sens. Environ. 240, 111680.
Zhao, H., Li, J., Yuan, Q., Lin, L., Yue, L., Xu, H., 2022. Downscaling of soil moisture products using deep learning: comparison and analysis on Tibetan plateau. J. Hydrol. 607, 127570.
Zhu, L., Walker, J.P., Tsang, L., Huang, H., Ye, N., Rüdiger, C., 2019a. A multi-frequency framework for soil moisture retrieval from time series radar data. Remote Sens. Environ. 235, 111433.
Zhu, L., Walker, J.P., Tsang, L., Huang, H., Ye, N., Rüdiger, C., 2019b. Soil moisture retrieval from time series multi-angular radar data using a dry down constraint. Remote Sens. Environ. 231, 111237.
Zhu, L., Walker, J.P., Ye, N., Rüdiger, C., 2019c. Roughness and vegetation change detection: a pre-processing for soil moisture retrieval from multi-temporal SAR imagery. Remote Sens. Environ. 225, 93–106.
Zhu, L., Webb, G.I., Yebra, M., Scortechini, G., Miller, L., Petitjean, F., 2021. Live fuel moisture content estimation from MODIS: a deep learning approach. ISPRS J. Photogramm. Remote Sens. 179, 81–91.
Zhu, L., Si, R., Shen, X., Walker, J., 2022. An advanced change detection method for time series soil moisture retrieval from Sentinel-1. Remote Sens. Environ. 279, 113137.
Zhu, L., Yuan, S., Liu, Y., Chen, C., Walker, J.P., 2023. Time series soil moisture retrieval from SAR data: multi-temporal constraints and a global validation. Remote Sens. Environ. 287, 113466.
Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q., 2020. A comprehensive survey on transfer learning. Proc. IEEE 109, 43–76.