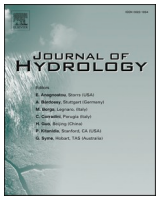
Journal of Hydrology 657 (2025) 133073

Contents lists available at ScienceDirect

Journal of Hydrology

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



Multi-Scale domain adaptation for high-resolution soil moisture retrieval from synthetic aperture radar in data-scarce regions

Liujun Zhu [a,b], Qi Cai [a], Junliang Jin [a,d,*], Shanshui Yuan [a,c], Xiaoji Shen [a,b], Jeffrey P. Walker [b]

[a] National Key Laboratory of Water Disaster Prevention, Yangtze Institute for Conservation and Development, Hohai University, Nanjing 210098 China
[b] Department of Civil Engineering, Monash University, Clayton, Vic. 3800, Australia
[c] Key Laboratory of Hydrologic-Cycle and Hydrodynamic-System of Ministry of Water Resources, Hohai University, Nanjing 210098 China
[d] Research Center for Climate Change of Ministry of Water Resources, Nanjing 210029, China

A B S T R A C T

Remote sensing of soil moisture plays an important role in advancing various hydrology applications, with Synthetic Aperture Radar (SAR) being the most promising technique for high-resolution soil moisture estimation. The growing adoption of machine learning methods has further enhanced this field, though their effectiveness heavily relies on the availability and quality of in-situ measurements. A recent study has demonstrated that pretrained models at 9-km resolution, based on the Soil Moisture Active Passive (SMAP) soil moisture products, can be transferred to 1-km using fewer in-situ measurements. Despite the success of this cross-resolution framework, its performance in data-scarce regions at high resolutions remains poor. To address this limitation, a multi-scale domain adaption (MSDA) method was proposed for soil moisture retrieval from data-scarce regions at a resolution of 50 m, taking the pretrained 9 km models as the starting points. Two modifications were made: the integration of multi-scale losses at both 9-km and 50-m resolutions, and the application of domain loss to bridge the gap between training and testing datasets. The MSDA was evaluated in both transductive and inductive modes where transductive mode involves adapting the model using a portion of the unlabeled test data, and inductive mode involves generalizing the model to entirely new, unseen data. A total of 66,547 daily averaged soil moisture measurements from 480 stations of 7 networks across the Contiguous United States were used. In the transductive mode, the use of a single training station in the MSDA achieved an R and RMSE of 0.67 and 0.088 m^3/m^3 respectively, which were improved to 0.81 and 0.071 m^3/m^3 when using data from 45 training stations. An acceptable R and RMSE of 0.76 and 0.078 m^3/m^3 was achieved in the inductive mode. The joint use of the two modifications achieved significantly better results ($p < 0.01$), with a relative improvement of 5.8 – 20.0 %, overall, and a lower risk of performance deterioration in data-scarce scenarios.

1. Introduction

Near surface soil moisture monitoring at coarse resolutions (> 10 km) has been made operational by two state-of-art spaceborne microwave radiometers, i.e., NASA's Soil Moisture Active and Passive (SMAP, (Entekhabi et al. 2010) mission and ESA's Soil Moisture and Ocean Salinity (Kerr et al. 2010) mission, offering major contributions to various applications (Cao et al. 2022; Jalilvand et al. 2023). While various downscaling methods have been developed, reliable soil moisture estimation at a resolution of < 1 km is still challenging (Peng et al. 2020). Synthetic Aperture Radars (SARs) provide a promising alternative for high resolution soil moisture, with recent SAR missions (e.g., Sentinel-1) having an enhanced global revisit of < 12 days (Torres et al. 2012).

In the past decade, substantial efforts have gone to retrieving soil moisture from SAR data operationally, with the most popular methods including the change detection methods (Arias et al. 2023; Balenzano et al. 2011; Wagner et al. 1999; Zhu et al. 2022), time series methods (Fan et al. 2021; Kim et al. 2012; Zhu et al. 2019; Zhu et al. 2023) and machine learning methods (Batchu et al. 2023; Liu et al. 2022; Wang et al. 2024). Among these, machine learning methods gain increasing acceptance as they are free of complex scattering modeling and do not require rigorous preprocessing of SAR data, e.g., noise removal, incidence angle normalization, terrain correction (Zhu et al. 2024). All the pre-works of conventional soil moisture inversion can be made automatically or compensated for by the feature (explainable variables) learning from multimodal data, allowing incredible convenience in operational soil moisture mapping.

Existing applications of machine learning methods in SAR remote sensing of soil moisture can be categorized into three main groups, with the first focusing on modeling of surface scattering. Some studies have trained machine learning models using simulated data sets generated by existing scattering models, with the inversion process then made using the trained models instead of directly inverting the scattering models

* Corresponding author.
E-mail address: jljin@hhu.edu.cn (J. Jin).

https://doi.org/10.1016/j.jhydrol.2025.133073
Received 30 December 2024; Received in revised form 3 March 2025; Accepted 7 March 2025
Available online 15 March 2025

(Baghdadi et al. 2002; Dong et al. 2023; Paloscia et al. 2008). Recent development of these studies include the Physical-Informed Neural Networks (PINNs) where physical scattering models are coupled with neural networks, resulting in improved performance in both the forward modeling and soil moisture retrieval (Bai and Tan 2023).

Other studies have directly applied conventional machine learning methods to retrieve soil moisture from plot to global scales, e.g., random forest (RF, Zheng et al. 2023), support vector regression (SVR, Stamenković et al. 2015; Xu et al. 2019), artificial neural network (ANN, Ahmad et al. 2010) and eXtreme Gradient Boosting (XGboost, Nguyen et al. 2022). Apart from directly using the raw backscatter, parameters derived by polarimetric decomposition (Chen et al. 2021), difference between or ratios of different polarizations (Wang et al. 2022), or two successive observations (Xu et al. 2019) were also used, being part of the feature extraction step. The auxiliary inputs (predictor attributes) have varied among the aforementioned studies but commonly include terrain features, soil texture, normalized difference vegetation index (NDVI), as well as precipitation and/or evaporation from either optical sensors or reanalysis datasets (Karamouz et al. 2022). The performance of these methods was commonly compared with the naïve version of RF, SVR, ANN and XGboost, with the ensemble methods (i.e., RF and XGboost) being found to generally perform better (Chen et al. 2021; Miller et al. 2022; Nguyen et al. 2022; Zhu et al. 2020).

The recent advances in deep learning and the considerable ground soil moisture measurements now available through the International Soil Moisture Network (ISMN, Dorigo et al. 2021) allow more powerful machine learning models be developed for the remote sensing of soil moisture, built either on the long short-term memory (LSTM, Celik et al. 2022), convolution neural network (CNN, Batchu et al. 2023) and/or transformer architecture (Wang et al. 2024). The global validation of these methods showed a much better accuracy than both the conventional machine learning methods and the time series methods, providing a promising solution to operational global soil moisture mapping at high spatial resolution ($<$ 1 km). However, these deep learning methods require a large amount of training samples and the transfer of trained models to the periods and/or areas with scarce samples is still challenging (Zhu et al. 2025). In our previous study (Zhu et al., 2024), a cross-resolution transfer learning approach was proposed, which assumes that the coarse soil moisture products produced by the SMAP mission can be used to pretrain robust models for a coarse grid (9 km), and be finetuned to a higher resolution of 1 km. Validation based on the ISMN showed a root mean square error (RMSE) of 0.06 m^3/m^3 when finetuning to 2 % of the $\sim$ 190,000 available ground samples. Despite the ability to substantially reduce the training samples, the ability to transfer to locations/areas without samples was still poor. Moreover, the performance of the transferred models was found to be worse than the trained model with 9 km data for scenarios with scarce finetuning samples.

The poor generalization capability partly arises from the domain shift (Tan et al. 2018), where the data distribution in the training set (source domain) differs from that in the testing set (target domain). Apart from finetune, multi-task learning, few-shot learning, and unsupervised domain adaption can be adapted to address the domain shift, with numerous successful cases in computer vision (Tan et al. 2018) and remote sensing classifications (Ma et al. 2024). In view of the training process, the negative effect of transfer learning can be ascribed to overfitting of the limited training samples. Widely used solutions include having a smaller learning rate, or freezing part of the network and applying another regularization scheme (Abdalla et al. 2019). While these strategies are useful, substantial efforts are still required for reliable high resolution soil moisture retrieval over data-scarce regions using deep transfer learning.

In this study, a multi-scale domain adaption (MSDA) approach was proposed for high resolution retrieval over data-scarce regions, taking the pretrained 9 km models as the starting point (Zhu et al., 2024). The target "high" resolution was set to 50 m because i) SAR data at a higher resolution can have large speckle noise; ii) most auxiliary data are available at a relatively coarser grid (e.g., the soil texture); and iii) it meets the requirements of most applications including farm-scale water management (Peng et al. 2020). Two modifications were made to the cross-resolution finetune method of Zhu et al. (2024) for improved generalization capability and lower risk over data scarce scenarios, i.e., utilizing the knowledge from both 50 m and 9 km and including the domain shift loss between the source and target domains. The proposed method was ultimately evaluated using samples collected from 7 soil moisture networks across the Continental United States (CONUS), covering both agriculture types and others. Two evaluation modes of transductive and inductive were involved for further suggestions in agricultural water management. The transductive mode makes predictions only for the unlabeled samples of the SCAN network and the input features of these unlabeled samples were utilized to minimize the domain loss, while the inductive mode used the models trained for SCAN to predict soil moisture on completely unseen data. Moreover, the role of each modification was addressed.

2. Data and pre-processing

The proposed multi-scale domain adaption (MSDA) approach is built on the pretrained Densely connected networks for Soil Moisture (DenseSM) at 9 km (Zhu et al., 2024). The required input variables of the DenseSM model are summarized in Table 1, including 3 time-invariant categories of soil texture, terrain features and locations, as well as 3 dynamic categories of SAR observations, vegetation descriptors and weather variables. Sentinel-1 SAR data is the main input, with the other parameters prepared based on the acquisition date of the SAR data. Since MSDA requires data at both 9 km and 50 m, all datasets were processed accordingly on the Equal-Area Scalable Earth grid V2.0 (EASE-2.0). SMAP soil moisture is natively distributed on the EASE-2.0 grid, while other datasets were either resampled or upscaled to match the 9 km and 50 m grids.

2.1. Soil moisture at 50 m and 9 km

The soil moisture truth at 50 m was generated using in-situ soil moisture measurements from 7 networks across the CONUS (Table 2), being the SCAN, USCRN, SOILSCAPE, PHO_2HO, iRON, FLUXNET-AMERIFLUX and ARM respectively. Another national-wide network (SNOTEL) was excluded due to the large terrain variations around its stations, which could introduce relatively larger spatial representativeness errors (Peng et al. 2024). While pre-selecting networks helps mitigate scale mismatch between point measurements and the 50 m reference data, it does not entirely eliminate the issue. However, a more reliable preprocessing step is not feasible due to the lack of measurements in nearby regions. All available records of these networks from 2016 to 2021 were collected from the ISMN, but only the stations with $>$ 5 coincident Sentinel-1 acquisitions were kept, resulting in 480 stations. While Sentinel-1 data was collected at $\sim$ 6 AM (ascending orbit) and $\sim$ 6 PM (descending orbit) local solar time, the daily average of the in-situ soil moisture measurements ($<=$ 5 cm) within the same 50-m grid cell was treated as the truth for simplicity, suggesting that the models and method proposed in this study are designed for daily averaged soil moisture. However, the soil moisture difference between the daily averaged value and the value of the descending or ascending hour was only $\sim$ 0.002 m^3/m^3. All 480 stations were located in separate grid cells. While some stations having multiple sensors at the same depth, data was collected from a small area of a few m^2, being much smaller than the target grid cell (50 $\times$ 50 m^2). Consequently, the soil moisture truth at 50 m may deviate from the "real" soil moisture, with the uncertainty of scale mismatch remaining unclear.

Soil moisture at 9 km were extracted from the SMAP L3 Radiometer Global Daily 9 km Soil Moisture, Version 5 (O'Neill et al. 2021). While this product is provided on a 9 km grid, it does not really reflect the

Table 1
The input variables of the DenseSM.

Category	Variable name	Range for normalization	# Input variables	Grid size	Temporal resolution
SAR observation	VV [dB]	[-30, 5]	1	10 *m*	6 or 12 days
	VH [dB]	[-35, 0]	1		
	Incidence angle [°]	[29.1, 46]	1		
	Sentinel-1 DoY*	[1, 365]	2		
Soil texture	Sand [%]	[0, 100]	1	250 *m*	Static (released in 2020)
	Clay [%]	[0, 100]	1		
	Bulk density [g/cm^3]	[1, 1.8]	1		
Terrain attributes	Elevation [m]	[0, 5500]	1	30 *m*	Static and the data source was collected at 1999
	Slope [°]	[0, 40]	1		
	Aspect [°]*	[0, 360]	2		
Location	Grid row number	[0, 1624]	1	9000 *m*	Static
	Grid column number*	[0, 3856]	2		
Vegetation	NDVI	[-0.2, 1]	46	250 *m*	8 days
Weather	Temperature [°C]	[-10, 35]	46	0.1°	1 h
	Precipitation [m/d]	[0, 0.3]	46		

* Projected to the cosine and sine space.

Table 2
The number of samples for 9 km and 50 m from the 7 soil moisture networks located in the CONUS.

Network	Number of stations	Number of samples at 50 *m*	Number of samples at 9 km
ARM (Galle et al. 2018)	17	4,055	4,396
FLUXNET-AMERIFLUX	5	875	2,947
iRON (Osenga et al. 2021)	9	1,263	2,183
PBO_H2O (Larson et al. 2008)	134	6,398	49,498
SCAN (Schaefer et al. 2007)	187	33,542	50,782
SOILSCAPE (Moghaddam et al. 2016)	25	245	3,476
USCRN (Bell et al. 2013)	103	20,169	26,014
Total	480	66,547	139,296

spatial variations of soil moisture variations at 9 km, as each value was interpolated from the oversampled SMAP native 36 km product using the Backus-Gilbert optimal interpolation algorithm. This product is a daily composite of multiple SMAP L2 half-orbit soil moisture, with both ascending (6 PM LST) and descending (6 AM LST) data. The bias of ascending and descending data was removed in this study by taking the descending data as the benchmark and the observations from both orbits then averaged as the truth of 9 km, being consistent with the calculation of the daily averaged in-situ measurements. The pretrained 9 km model was trained for global application with the grid cells of training samples covering all ISMN stations (Zhu et al., 2024), while only the 480 grid cells covering the aforementioned 7 networks were used in this study for providing local information.

2.2. Static input variables of models

Static variables include the terrain features, soil texture and locations. Three terrain features of elevation, slope and aspect were extracted or calculated from the Shuttle Radar Topography Mission Digital Elevation Data Version 3 (SRTM DEM V3) at a resolution of 30 m, which were resampled to 50 m and 9 km using an Area-based Weight Aggregation (AWA) method. For a target grid cell resolution of 50 m, the AWA first determined the DEM pixels overlapped with the target grid cell and then averaged the overlapped pixels taking the overlapped areas as the weights. The calculated elevation, slope and aspect were normalized to [0, 1] using the range listed in Table 1 and the max–min normalization method. Specially, the aspect angle is a cyclic variable where 0 equals to 2π and the numerical difference of two aspect angles cannot represent the difference of terrain orientations. Accordingly, sine and cosine of aspect angles were used to determine the directional difference of terrain orientations. In this study, all cyclic variables were normalized in a similar way.

Soil texture were extracted from the global gridded soil information (SoilGrids 250 m, Poggio et al. 2021). The sand ratio, clay ratio and bulk density of the top 5 cm were reprojected and resampled to the two target grids, first using the AWA method and then normalized to [0, 1] as part of the model inputs. The location of each grid cell was critical information that allowed the DenseSM to learn location dependent features and was thus included. Following the inputs of the pretrained DenseSMs, the row and column number of the 9 km EASE-2.0 grid was used for both the 50 m and 9 km resolution grids. This means that the locations of multiple 50 m grid cells within the same 9 km grid cell are assigned the same value.

2.3. Dynamic input variables of models

The dynamic inputs include the Sentinel-1 SAR observations, one-year temperature, precipitation, and Normalized Difference Vegetation Index (NDVI) preceding the acquisition date of Sentinel-1. The google earth engine (GEE) archive of Sentinel-1 was used in this study, collected in the Interferometric Wide (IW) mode and pre-processed to Ground Range Detected (GRD) images using a standard processing flow. The VV and VH backscatter coefficients (sigma null) and the approximate incidence angle at 10 m were upscaled to 50 m and 9 km respectively using the AWA method. All available Sentinel-1 GRD data was used, resulting in 66,547 SAR observations from the 480 grid cells at 50 m resolution (Table 2). In contrast, the number of samples prepared for the 9 km grid was 139,296 as the SMAP has a shorter revisit of 2–3 days. Since SAR observations are the main inputs, the acquisition date of Sentinel-1 in day of year (DoY) was also included as an auxiliary variable, allowing models to learn time dependent features. Both the Sentinel-1 and SMAP L3 9 km were prepared using the GEE without downloading the acquisitions; the codes are shared at https://github.com/rszlj/Transfer-DenseSM-E.

The time series NDVI used in this study was from two Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI products (MOD13Q1 and MYD13Q1, V6) with a resolution of 250 m. The MOD13Q1 and MYD13Q1 were two 16-day composite NDVI products produced from the data collected by two identical satellites Terra and Aqua respectively. They were combined to provide the temporal dynamic of vegetation every 8 days in this study (Didan 2015). A one-year time series preceding the acquisition date of each SAR observation were extracted from the GEE using the AWA method, with the extracted time

series having a length of 46 and thus not being exactly one year (46 × 8 days).

The time series temperature and precipitation were extracted from the reanalysis land dataset (ERA5-land, Muñoz-Sabater et al. 2021). The original ERA5-land is an hourly dataset with a resolution of 0.1°. They were first resampled to 50 m and 9 km using the AWA method. Different from many studies that take weather variables as the forcing data of soil moisture, the time series temperature and precipitation were treated as indicators of climate types in the DenseSM and so were aggregated to a coarser temporal resolution of 8 days for a smaller size of trainable parameters (Zhu et al., 2024). More specifically, the temperature is the average air temperature at 2 m of 8 successive days from the first day of the year, while the daily aggregation sum of total precipitation was calculated first, with the average of the same 8 days being the precipitation required in DenseSM. The pre-processed temperature and precipitation preceding each SAR observation thus also had a length of 46, being consistent with the NDVI time series.

3. Methodology

3.1. Pretrained models at 9 km

The proposed multi-scale domain adaption (MSDA) approach was built on the DenseSM. The inputs of DenseSM are summarized in Table 1 with all the variables required to be normalized to the interval [0, 1] using the min–max normalization method. The DenseSM consists of a feature extraction block, a few dense blocks, and a regression block (Fig. 1 a). In the feature extraction block, a one-year time series of temperature, precipitation and NDVI were fed into a temporal feature extractor to extract 15 features representing the vegetation and climate conditions. They were then concatenated with the corresponding SAR data (4 variables), static terrain features (4 variables), soil textures (3 variables) and location attributes (3 variables). The 29 concatenated features were fed into *d* (depth) densely connected blocks (Huang et al. 2017), each of which had 3 fully connected layers with *w* (width) neurons in each layer. The outputs of the dense blocks were then fully connected to a bottleneck layer of 32 neurons, followed by a regression layer to predict the soil moisture; refer to Zhu et al. (2024) or the shared pre-trained models for more details about the temporal feature extractor and dense block.

The size of DenseSM is controlled by the value of *w* and *d*, with a larger *w* and *d* meaning a wider and deeper network that is more powerful when sufficient training samples are available. However, the performance was variable for scenarios with sparse samples (Zhu et al., 2024). Determining the "optimal" values for *w* and *d* using *k*-fold cross-validation can be challenging when the training set is too small to be split effectively. Additionally, the "optimal" hyperparameters based on a small training set may not be reliable for the target domain. So instead of searching for the optimal hyperparameters, *w* was set to 8, 16, 32, 64 and 128, while *d* was set to 1, 2, 3, 4 and 5, resulting in 25 DenseSM models. Soil moisture was then estimated independently using the 25 models and the ensemble averaged as the output, being denoted as DenseSM-E (Fig. 1 b). The DenseSM-E was found to have better results than any single model and to be insensitive to the hyperparameters (Zhu et al., 2024). Accordingly, the proposed MSDA was also applied to the DenseSM-E of 25 DenseSM models in this study.

3.2. Multi-Scale domain adaption (MSDA) method

The proposed MSDA has two main branches (Fig. 2). The inputs of the first branch are labeled samples at 50 m and 9 km, with the labels being the soil moisture truth at 50 m and 9 km resolution respectively. The inputs of the second branch are unlabeled samples at 50 m resolution from the target domain. Since the SMAP L3 9 km resolution soil moisture was treated as the truth of the 9 km grid and are available globally, the samples at the 9 km resolution of the first branch were taken from both the source and target domains. Both branches have identical architecture to the DenseSM, with the initial trainable parameters being copied from the pretrained DenseSM (Zhu et al., 2024). The outputs of the first branch include the predicted soil moisture at both 9 km and 50 m resolutions. Moreover, the outputs of the bottleneck layer were extracted for the 50 m resolution, being the learned feature representations of the source domain for regression. In contrast, the second branch only has the feature representations of the target domain as its output.

Three loss items are calculated using the estimated soil moisture and the extracted feature representations, being the estimation loss at 50 m ($\mathscr{L}_{50m}$) and 9 km ($\mathscr{L}_{9km}$) resolutions, and the domain discrepancy between the source and target domains. The soil moisture estimation losses were calculated using Mean Squared Error (MSE):

$$\mathscr{L}_s = \frac{1}{n}\sum_{i=1}^{n}(y_{s,i} - \hat{y}_{s,i})^2 \quad (1)$$

where $y_{s,i}$ and $\hat{y}_{s,i}$ are the truth and estimated soil moisture of the *i*th sample, with *s* being either 50 m or 9 km resolution, and *n* is the number of labeled samples for training, being the batch size in deep learning.

A few methods/metrics have been proposed to estimate the domain discrepancy, including the Maximum Mean Discrepancy (MMD, (Dziugaite et al. 2015), Kullback-Leibler Divergence and CORrelation ALignment (CORAL, Sun et al. 2017). In this study, CORAL was selected

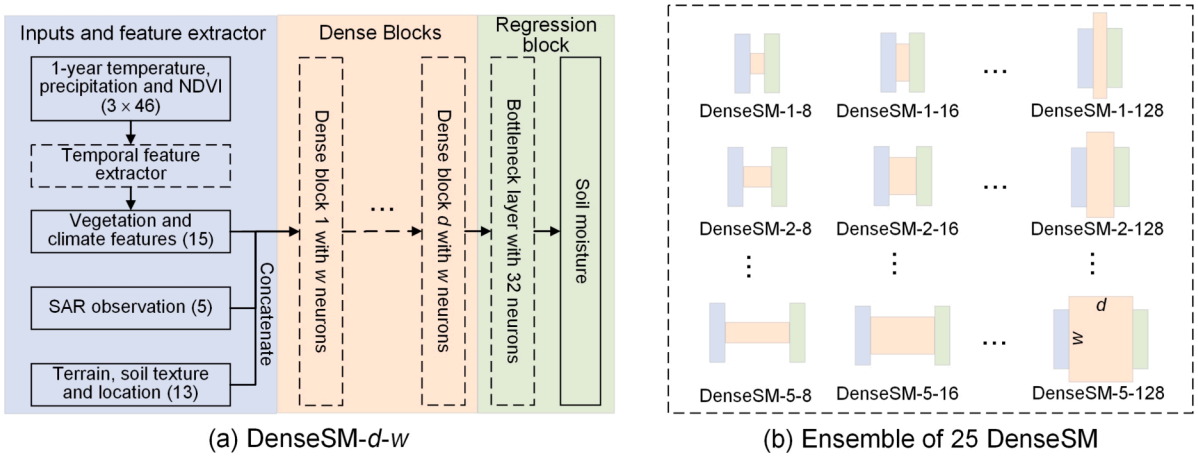


Fig. 1. The architecture of the DenseSM with the numbers in parentheses being the number of variables or feature representations (a) and the ensemble concept of 25 DenseSM-*d*-*w*, with the *w* and *d* being the width and depth of the dense blocks respectively (b).

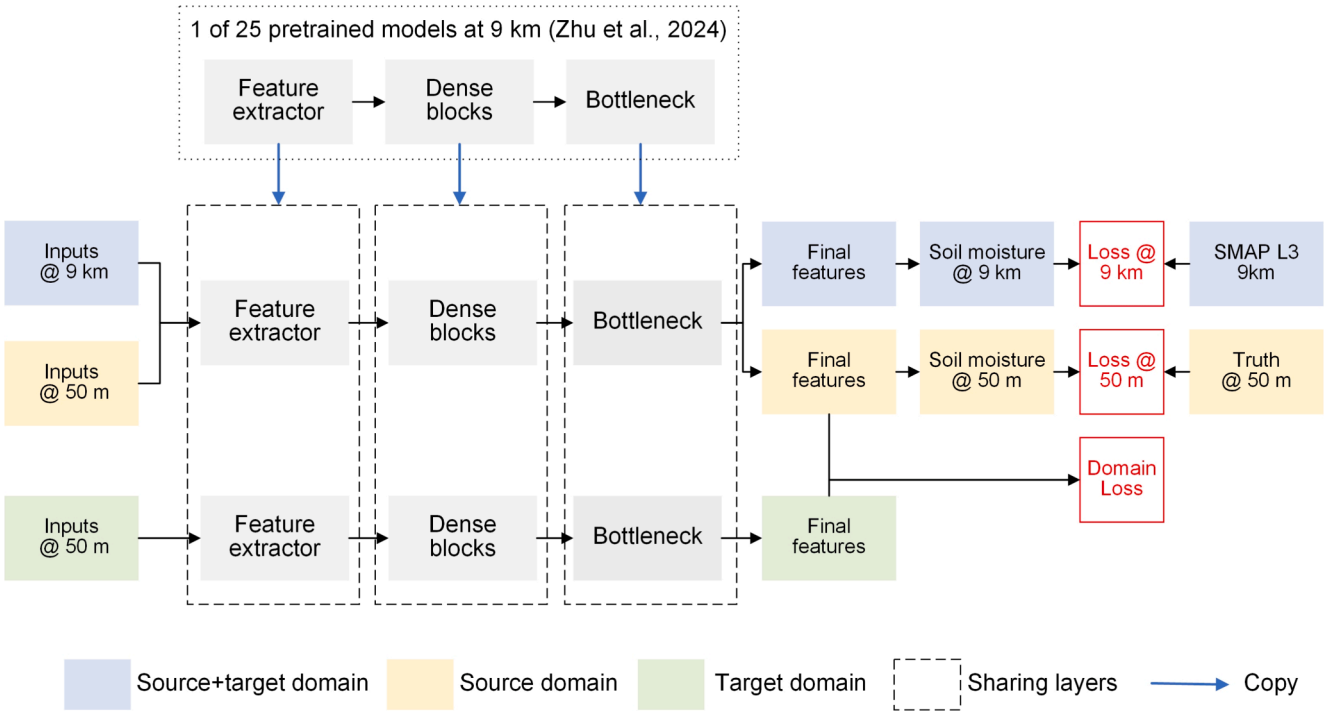


Fig. 2. The proposed multi-scale domain adaption, taking the pretrained 9 km models as the starting points. The source and target domain refer to the training and testing sites.

as it is computationally efficient and straightforward to be implemented without requiring complex kernel functions or hyperparameter tuning. The simplicity is from the assumption of linear relationships between the source and target domains, being not valid for most scenarios. However, the CORAL loss was calculated for the final feature representations in this study. Aligning second-order statistics is sufficient and more straightforward than considering higher-order statistics, as the final feature representations were fed into a linear regression layer for soil moisture retrieval. The domain loss in CORAL of two feature presentations can be calculated using:

$$\mathscr{L}_{\text{domain}} = \frac{1}{4d^2}\|C_S - C_T\|_F^2 \tag{2}$$

where C_S and C_T are the covariance matrices of the source and target feature representations respectively, $\|\ \|_F$ denotes the Frobenius norm and d is the dimensionality of the features, being the number of neurons of the bottleneck layer in this study, i.e., 32.

A joint cost function was calculated using the three cost items:

$$\mathscr{L} = \alpha\mathscr{L}_{50\text{m}} + (1-\alpha)\mathscr{L}_{9\text{km}} + \beta\mathscr{L}_{\text{domain}} \tag{3}$$

where α and β are two empirical parameters, controlling the weight of each item, with α having a range of [0, 1]. The special case of $\alpha = 0$ means that only the loss at 9 km resolution was considered and thus ground truth at 50 m resolution is not required, being an unsupervised domain adaption method. In contrast, the special case of $\alpha = 1$ only uses data from 50 m resolution and thus becomes a single scale domain adaption method. Since the estimation loss and domain loss have different units and are challenging to be normalized, it is impossible to determine a unique β. Previous experiments on various object recognition tasks have suggested that maintaining a balanced classification loss and CORAL loss after training lead to feature representations that are both discriminative and have minimized discrepancy between the source and target domains (Sun and Saenko 2016). However, it is still not straightforward for a determined β that $\mathscr{L}_{50\text{m}}$ and $\mathscr{L}_{\text{domain}}$ have a similar value at the end of training. In contrast, the balance between estimation and domain can be made easily at the beginning of the training, allowing a rough balance at the end of training. Accordingly, β was determined using:

$$\beta = \alpha\mathscr{L}_{50\text{m}}^{\text{pretrained}} / \mathscr{L}_{\text{domain}}^{\text{pretrained}} \tag{4}$$

where $\mathscr{L}_{50\text{m}}^{\text{pretrained}}$ and $\mathscr{L}_{\text{domain}}^{\text{pretrained}}$ are the estimation and domain loss of the pretrained 9 km model at 50 m resolution. Eq. (4) therefore allows a similar value of domain loss and estimation loss at the beginning of the training process.

3.3. Training and validation methods

A pretrained DenseSM model was finetuned using the loss function described in Eq. (3) and the Adam optimizer with default parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ Kingma and Ba 2014). Training was conducted for 50 epochs, and the trainable parameters from the last 10 epochs averaged to form the final trained model (Zhu et al. 2021). The batch size was set to 128, meaning each batch included 128 labeled training samples at 50 m resolution, 128 labeled samples at 9 km resolution, and 128 unlabeled samples at 50 m resolution. Given the substantially smaller number of labeled samples at 50 m resolution compared to those at 9 km resolution and the unlabeled samples at 50 m resolution, samples from the 9 km resolution set or the unlabeled 50 m resolution set were randomly selected without replacement until all samples were used. This iteration method addresses the discrepancy between the small source domain and the large target domain, as well as the estimation loss of 9 km resolution over a larger area and/or a longer period. A relatively small learning rate of $5e^{-4}$ was used to reduce the risk of overfitting during the early epochs. Moreover, there are 25 pretrained 9 km resolution DenseSM models and so the MSDA was applied to each model independently, resulting in 25 transfer DenseSM models at 50 m resolution. The ensemble average of these 25 models was then considered the final output, referred to as the transfer DenseSM-E. Notably, this study only focused on the transfer version of DenseSM-E.

Four metrics of bias were used to assess the performance of the transfer models; Pearson correlation coefficient (R), RMSE and unbiased

RMSE (ubRMSE). Since model performance depends on the selected training sets, especially in scenarios with few stations, training and evaluation were conducted 10 times using different random training stations. The proposed MSDA was evaluated in both transductive and inductive inference modes. In the transductive mode, the testing samples were used to calculate the feature representations of the target domain, while in the inductive mode the models trained for the transductive mode were applied without any further training. Accordingly, the accuracy in the transductive mode is generally expected to be higher than that in the inductive mode.

€The evaluation of the transductive mode was focused on the SCAN network as it has the largest number of stations among those listed in Table 2 and a long continuity of data records, with around 85 % stations having > 100 samples (Fig. 3 a). To evaluate the MSDA in different scenarios with scarce stations, the number of training stations was increased from 1 to 45 with intervals of 4 or 8 stations, accounting for 0.5 % to 24 % of the SCAN network. The non-selected stations of each scenario were used for evaluation. Moreover, the CONUS was divided into 6 longitudinal regions and 6 latitudinal regions, each with a similar number of stations (~ 30). Accuracy statistics were calculated for each region to highlight the spatial discrepancy in performance. Moreover, the cross-resolution finetune method (Zhu et al., 2014) was treated as the benchmark to show the contribution of the proposed MSDA. Since two modifications were made in the MSDA, including multi-scale loss and domain loss, two methods incorporating each as a single modification was used for comparison. The method using multi-scale loss and domain loss were denoted as multi-scale finetune (MS) and domain adaption (DA) respectively. A comparison of four transfer learning methods was therefore conducted on the SCAN network with varying numbers of training stations, providing experimental insights into the impact of these losses and/or constraints.

The models trained on the SCAN stations were further evaluated in the inductive mode using 6 other networks (Fig. 3b) to demonstrate their ability to be generalized over areas without stations. These include the PBO_H2O network located in the western CONUS, the USCRN that is evenly distributed across the CONUS similar to the SCAN, and the iRON, ARM, FLUXNET-AMERIFLUX and SOILSCAPE which are clustered in small regions. The different characteristics of these networks allows a more comprehensive evaluation of the MSDA, and thus progress towards operational soil moisture retrieval over areas without stations.

4. Results

4.1. Performance of the MSDA in the transductive mode

The performance of the MSDA on the SCAN network was evaluated first. Since 10 random implementations were made, only the model with the median RMSE (the 5th of the 10 implementations) is presented in Fig. 4. As expected, directly applying a pretrained 9 km resolution model ($N = 0$) on a fine resolution of 50 m led to poor results, with an RMSE of 0.098 m^3/m^3. However, the accuracy was improved substantially when data from a single station was involved in the transfer learning, leading to RMSE and R improvements of ~ 0.01 m^3/m^3 and 0.037 respectively. The performance was further improved by using more training stations, reaching an RSME of 0.075 m^3/m^3 when $N = 21$ (~11 % of the SCAN network). However, the additional benefit of including more than 21 training stations was relatively small, with the RMSE decreasing only slightly to 0.071 m^3/m^3 for $N = 45$, aligning with the finetune method (Zhu et al., 2024). The absolute values of biases remained < 0.02 m^3/m^3, with a shift to negative bias for $N > 21$. Consequently, removing the systematic biases contributed only marginally to accuracy improvement, with the difference between the RMSE and ubRMSE being < 0.002 m^3/m^3. Notably, the scatter patterns of measured versus predicted soil moisture remained consistent across different N values. Soil moisture values exceeding 0.3 m^3/m^3 tended to be underestimated, whereas values below 0.03 m^3/m^3 were substantially overestimated. Since most samples had a relatively small soil moisture of < 0.1 m^3/m^3, dense data clusters were observed in this range in each subplot. While biases at low values could not be fully eliminated, increasing the number of training stations resulted in a substantial overall improvement in model performance.

The station-specific bias, R and RMSE were calculated to show the spatial performance of MSDA, with the statistics averaged over 10 random implementations. The average accuracy metrics for stations within each longitudinal and latitudinal region are presented in Fig. 5a. Specially, the absolute values of bias were used to evaluate the model's capability in capturing the spatial variation of soil moisture. The RMSE and absolute bias of the pretrained models exhibited a "U" trend with longitude, showing higher values in the western and eastern regions, while R was observed to decrease from west to east. The retrieval error in the western regions was mainly attributed to station-specific bias, indicating that the pretrained 9 km resolution models can capture the overall trend of soil moisture at 50 m resolution but struggled to accurately recover the spatial variation of soil moisture levels within a 9 km grid. In contrast, the pretrained 9 km models not only failed to capture the relationship between input features but also made consistently

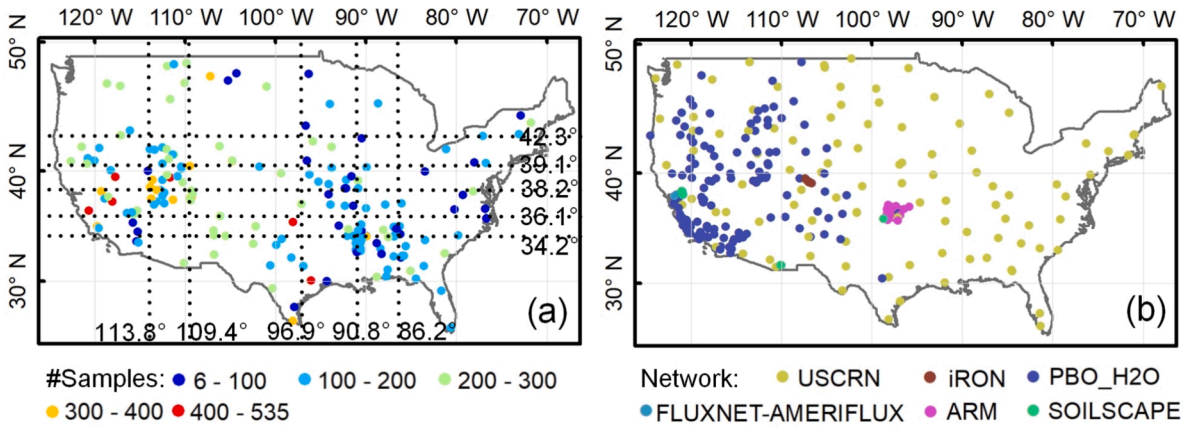


Fig. 3. The number of available samples for each SCAN station that used in transductive mode (a) and the spatial distribution of the other networks used in the inductive mode (b). The dashed-lines in (a) divide the CONUS into 6 longitudinal regions and 6 latitudinal regions, each with a similar number of SCAN stations (~30 stations).

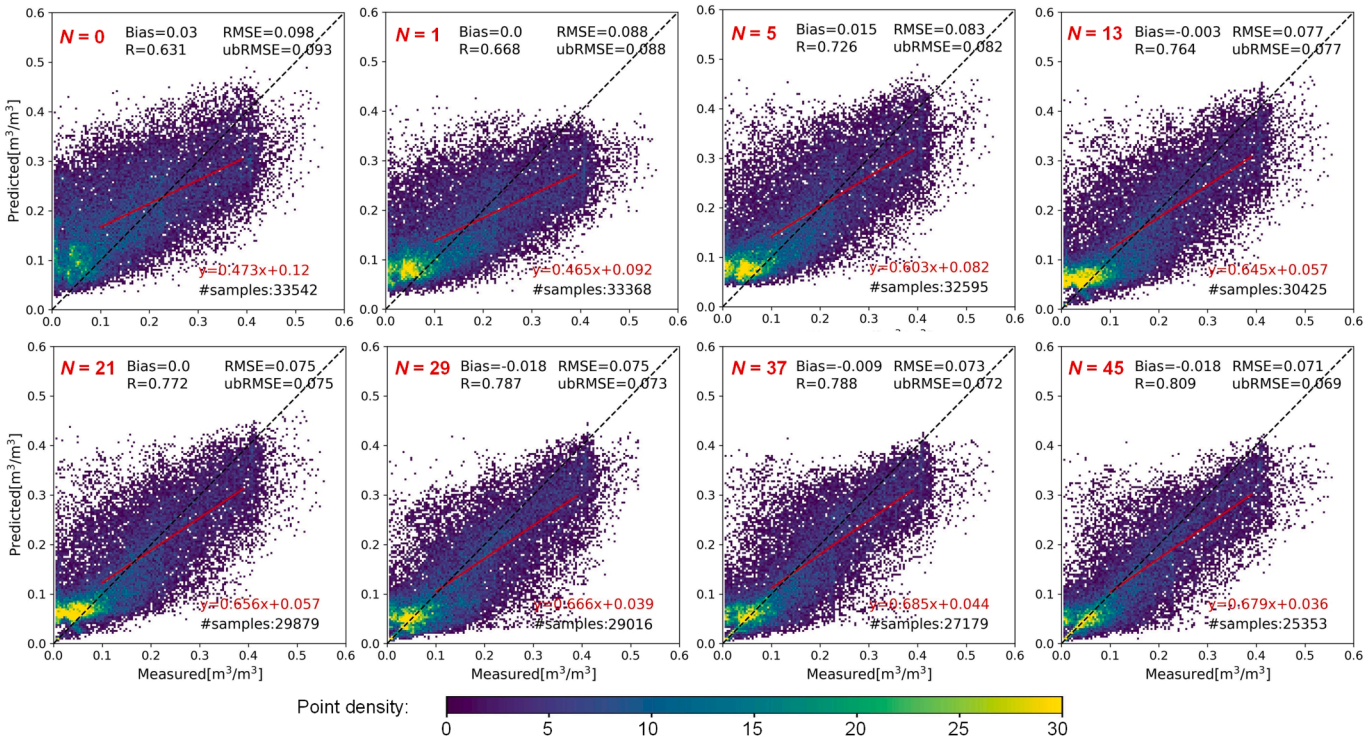


Fig. 4. Performance of the MSDA in the transductive mode with a varying number of training stations (N). Only the results of the models with the median RMSE were included. The number of testing samples (# samples) decreased as N increased, with the case of $N = 0$ equating to the performance of the pretrained 9 km DenseSM-E on 50 m. Using a single training station resulted in an RMSE of 0.088 m^3/m^3 at 50 m. However, the accuracy became acceptable when 45 stations were included in the training process.

inaccurate predictions in the eastern regions. The MSDA with a few training stations could considerably reduce biases, with the western regions benefiting more than the eastern regions, resulting in an increasing trend of RMSE from west to east. This trend was related to the increasing soil moisture levels from west to east, except for some special regions like the Pacific Northwest. The error caused by underestimating high values was larger than that caused by the overestimation of low values (Fig. 4). Moreover, the RMSE of the pretrained model along the latitude was dominated by the site-specific bias, with the central regions having relatively smaller biases. In contrast, the R of the pretrained model generally decreased from north to south. After applying the MSDA, the performance across all three metrics was improved, but the pattern with respect to the latitude was maintained. This indicates that the performance discrepancy of MSDA along latitude was primarily influenced by the performance of the pretrained model.

Fig. 5 b shows the station-specific RMSE of the pretrained model ($N = 0$) and the MSDA across 5 transfer learning scenarios. The pretrained model achieved an RMSE of > 0.1 m^3/m^3 for 59 of the 187 stations. The number of stations with large RMSE (> 0.1 m^3/m^3) was substantially reduced from 19 to 9 in the west regions (125°W–110°W) when the MSDA was applied with just one training station. The RMSE was further reduced in the western regions, with > 62 % stations achieving an RMSE of < 0.06 m^3/m^3 for $N >= 9$. However, not all the stations benefited from the MSDA; e.g., the four stations marked within the black dashed ellipse. The central and eastern CONUS also saw substantial improvements with a larger N, but still had a considerable number of stations with a large RMSE of > 0.1 m^3/m^3 (27 stations for $N = 45$), especially in the southeastern regions.

4.2. The contributions of multi-scale and domain loss

A comparison of the pretrained, finetune, MS, DA and MSDA methods was conducted on the SCAN network (Fig. 6). It was observed that the performance of the 4 transfer learning approaches improved gradually in view of RMSE and R as more training stations were used, with the largest improvements occurring at small N values of < 13. However, the bias initially decreased but gradually increased for $N > 5$ with the values generally remaining within a small range of 0 to −0.03 m^3/m^3. The difference between the RMSE and ubRMSE of all the methods was thus marginal. As N increased, the performance ranges of all methods gradually reduced, indicating a reduced sensitivity to the combination of the training stations.

The MSDA achieved significantly smaller RMSE and higher R than the finetune method ($p < 0.01$ in the t-test), the MS method ($p < 0.05$) and the DA method ($p < 0.05$) in all scenarios. Specially, the median RMSE values of the finetune, MS and DA methods were larger than the RMSE of the pretrained model when $N = 1$, while the MSDA outperformed the pretrained model in all scenarios. This suggests that transfer learning approaches do not necessarily yield better results in the target domain because the pretrained models can be overfitted when the training set is small. However, the proposed joint use of multi-scale and domain loss methods allows for safer transfer learning. Moreover, the RMSE range of the MSDA was approximately 0.01 m^3/m^3 when one training station was used, compared to around 0.06 m^3/m^3 for the fine-tuning method (Zhu et al., 2024). This suggests that certain sites are particularly helpful in Zhu et al. (2024), but their influence diminishes when the MSDA is applied, further confirming that the proposed MSDA is less sensitive to the combination of training stations. In addition to the enhanced stability, using two loss terms resulted in an improvement of 0.005 – 0.022 m^3/m^3 in median RMSE, corresponding to relative improvement of 5.8 – 20.0 %.

The use of either multi-scale or domain loss alone also achieved better results than the finetune method, but the difference in RMSE was not significant in most scenarios ($p > 0.05$). The limited benefit from a single loss term can be partly attributed to the small number (10) of random implementations. The difference between the MS and DA

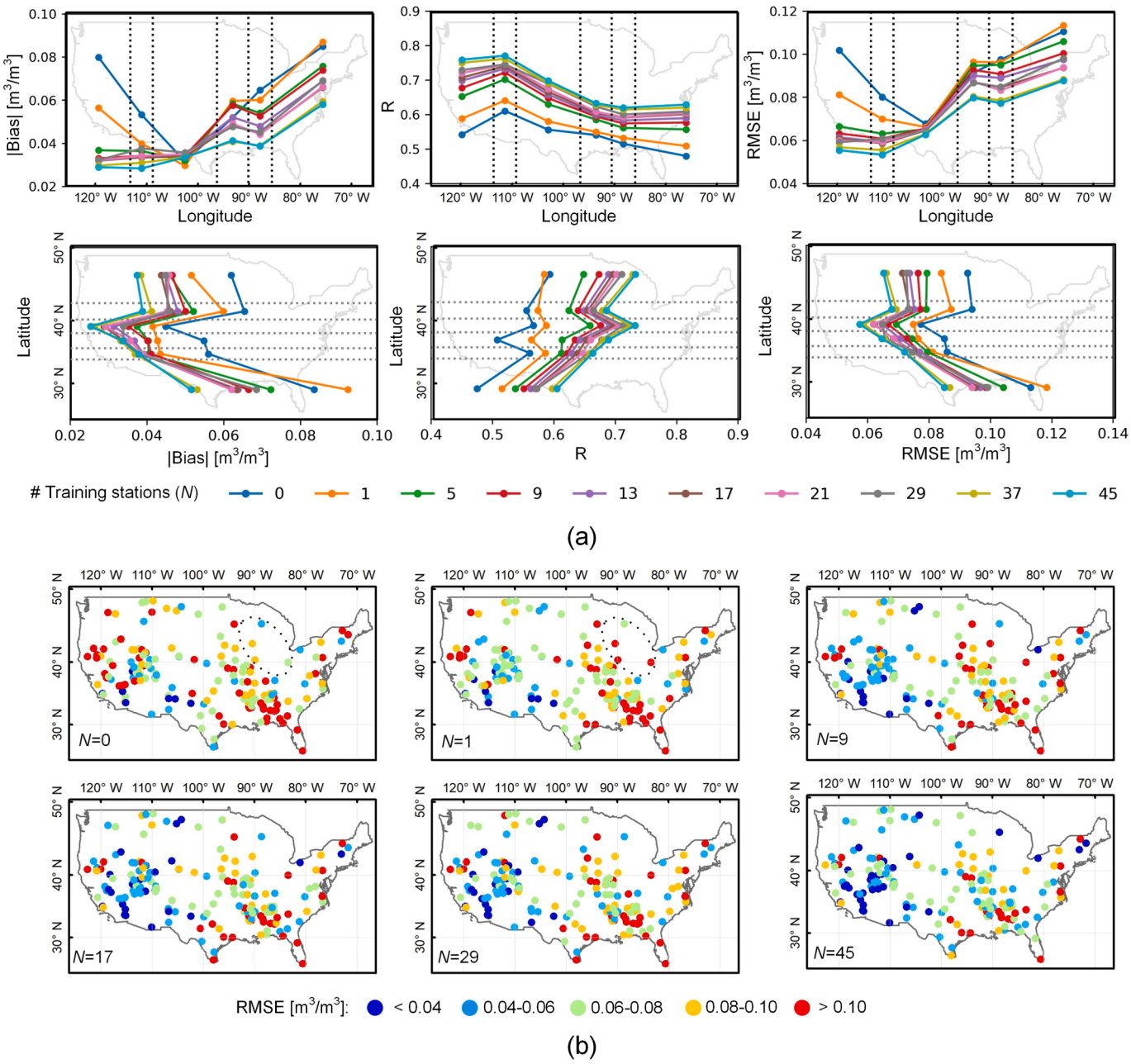


Fig. 5. Spatial performance of the MDSA using varying numbers of training stations (*N*); (a) is the region-specific accuracy metrics in latitude and longitude with the strip regions being created with a similar number of stations, and (b) is the station specific RMSE of MSDA.

methods was marginal, being < 0.006 m^3/m^3 in RMSE and < 0.06 in R. However, the MS method had relatively larger negative biases, resulting in smaller ubRMSE values. In contrast, the DA method had the smallest absolute bias in each scenario of $N > 1$. This can be explained by the minimization of the linear shift between the source and target domains. Similarly, the absolute bias of MSDA was smaller than that of MS due to the additional domain loss.

The effects of multi-scale and domain loss were further investigated from the retrieved time series soil moisture of the Monocline_Ridge station, which had the largest number of samples (Fig. 7). Fig. 7a depicts the results of the pretrained model, finetune method and MS method, along with the SMAP L3 9 km soil moisture. All the time series had a consistent temporal pattern of wet winters and dry summers, with the main difference being the average soil moisture. The time series soil moisture of the pretrained model was much smaller than the SMAP time series during the wet seasons. The finetune method brought only slight improvements to the retrieved time series, with the soil moisture being marginally higher than that of the pretrained model. In contrast, the MS method took the SMAP soil moisture as an extra target, and thus the time series of the MS was forced to be closer to the SMAP time series. This is favorable for scenarios with few training stations, as the SMAP soil moisture maintained the trend of the soil moisture at 9 km which is generally consistent with that of 50 m except for areas/periods with irrigation (Zhu et al., 2023). Consequently, the models are less likely to be overfitted when using only a few training stations.

Fig. 7b shows that the use of domain loss had little effect on the temporal trend of the retrieved soil moisture, but the retrieved time series was closer to the in-situ time series during the wet seasons. This improvement can be explained by the minimization of the domain shift in the DA method. Fig. 7c compares the in-situ, MS, DA and MSDA time series. All the models underestimated the soil moisture during the wet seasons, with the temporal trends of the retrieved time series being

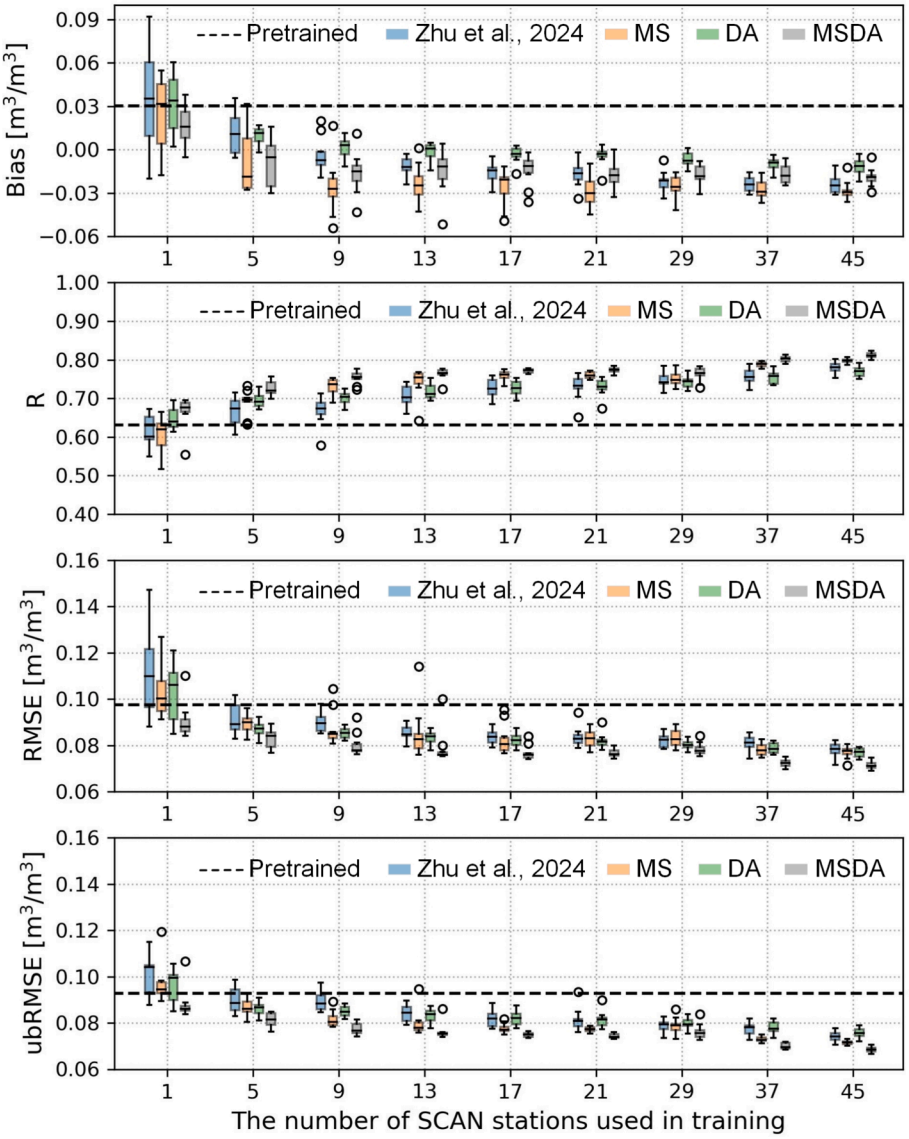


Fig. 6. The contribution of multi-scale loss (MS), domain loss (DA) and joint use of the two losses (MSDA), with the finetune method of Zhu et al. (2014) and the pretrained model as the benchmark. The transfer learning can benefit from either single modification with the largest improvement being achieved by the joint use of two modifications.

dominated by the SMAP time series (Fig. 7a). For example, the peaks of soil moisture in April to June 2016 were not captured by the SMAP mission and were also missed by all the transfer learning approaches. The time series of the MSDA was slightly smaller than that of the MS but higher than that of the DA. The MS method achieved the best results (RMSE = 0.065 m^3/m^3) in this case, followed by the MSDA (RMSE = 0.068 m^3/m^3) and DA (RMSE = 0.073 m^3/m^3). This indicates that the joint use of two loss items does not necessarily lead to better results at a specific location when compared to the use of a single loss term.

The weight parameter α was used to balance the loss at 50 m and 9 km in training, with its effect evaluated using three different values of 0.5, 0.7 and 0.9 (Table 3). The difference among these three cases were marginal, with the lowest RMSE being achieved using an α of 0.7 in 7 of the 9 cases. While not significant, a smaller α of 0.5 was found to result in the highest accuracy when a small number of training stations were available (N = 1 or 5). More weight on the 9 km loss can force the retrieved soil moisture at 50 m to be closer to the 9 km time series. In view of transfer learning, 9 km was used for pretraining, and so a higher weight on 9 km loss can reduce the risk of the catastrophic forgetting and overfitting. However, as more training stations are incorporated, the role of the 9 km loss becomes less critical and may even have a negative impact.

4.3. Performance of the MSDA in the inductive mode

The models trained on the SCAN network were further evaluated using data from the other 6 networks, with SCAN results included for comparison (Fig. 8). In general, the MSDA achieved the best results, followed by the DA, MS and finetune method, which is consistent with the evaluation in the transductive mode. Despite the varying improvements, using more training stations from the SCAN network resulted in better results on the USCRN, PBO_H2O, SOILSCAPE, iRON and FLUXNET-AMERIFLUX, with the improvements from adding more training stations diminished for $N > 13$. Similar to the evaluation results of the SCAN network, the RMSE of the finetune method decreased using even a single training station in these networks, while the MSDA achieved better retrieval accuracy. The ARM was the only network where the RMSE increased, by up to 0.015 m^3/m^3, as N increased from 1 to 45. The increasing RMSE was partly due to increased underestimation. The mean absolute bias of the 17 ARM stations increased from 0.05 to 0.055

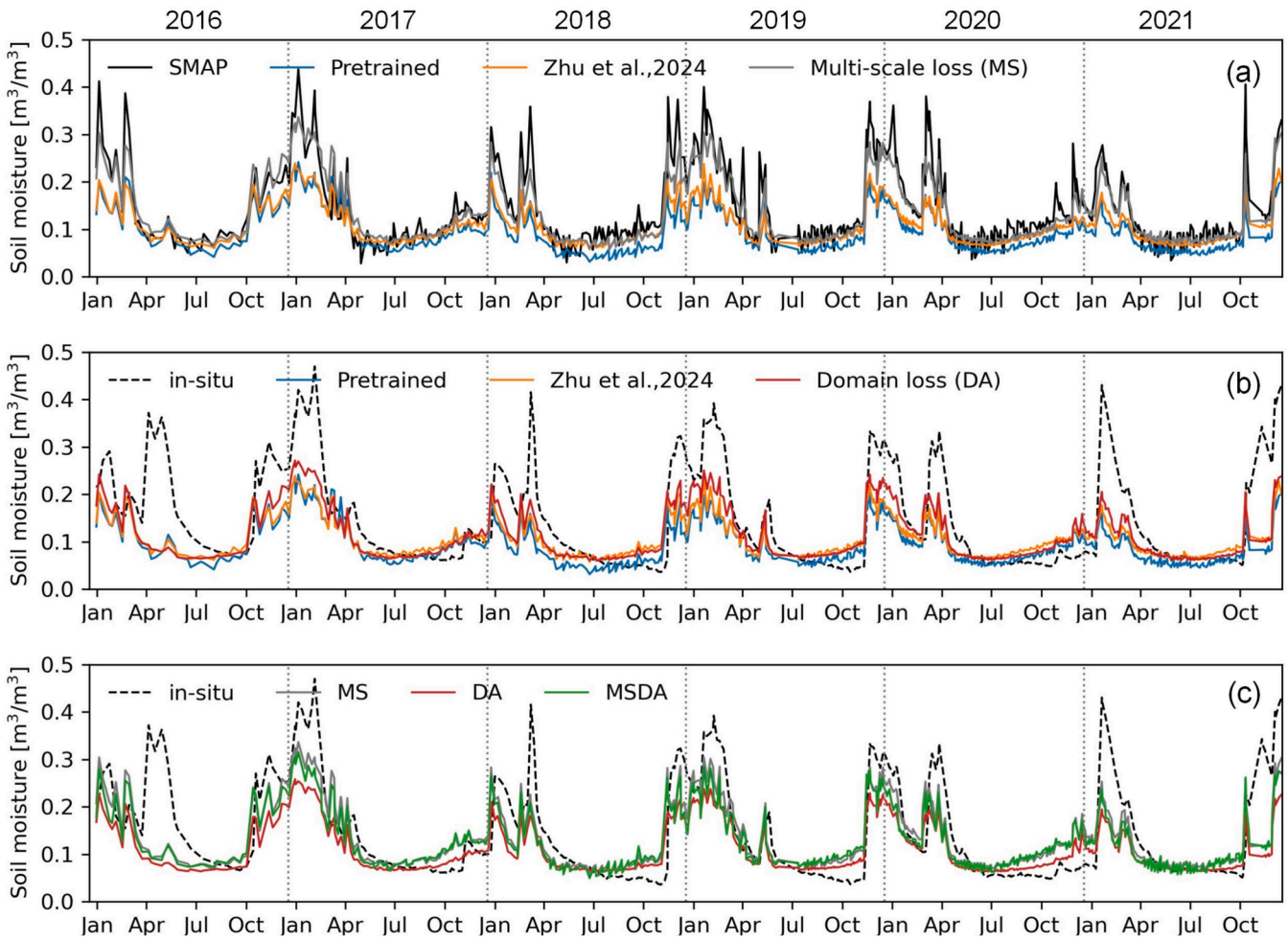


Fig. 7. The effect of multi-scale loss (MS), domain loss (DA) and the joint use of two losses (MSDA) on the retrieved time series of the Monocline_Ridge station, which has the largest number of soil moisture measurements. The finetune method of Zhu et al. (2014) and SMAP 9 km were also included for comparison. (a) shows the role of multi-scale loss, (b) shows the role of domain loss and (c) focus on the effect of the joint loss.

Table 3
The RMSE (m^3/m^3) of MSDA when compared against the SCAN network using different weight parameter of α. The A and B in A ± B are the average value and standard deviation respectively, with the best performance case being in bold.

N	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
1	**0.089 ± 0.005**	0.089 ± 0.007	0.091 ± 0.006
5	**0.081 ± 0.002**	0.082 ± 0.004	0.081 ± 0.004
9	0.081 ± 0.004	**0.079 ± 0.005**	0.084 ± 0.008
13	0.079 ± 0.002	**0.078 ± 0.007**	0.079 ± 0.004
17	0.079 ± 0.002	**0.077 ± 0.003**	0.078 ± 0.003
21	0.078 ± 0.003	**0.076 ± 0.002**	0.077 ± 0.005
29	0.077 ± 0.002	**0.075 ± 0.003**	0.076 ± 0.003
37	0.076 ± 0.003	**0.072 ± 0.002**	0.075 ± 0.002
45	0.075 ± 0.003	**0.071 ± 0.002**	0.073 ± 0.003

m^3/m^3, while the corresponding values of other stations decreased from 0.065 to 0.043 m^3/m^3.

The retrieval accuracy of each network varied considerably, showing the different generalization capabilities in space. All the transfer methods achieved relatively better results on the PBO_H2O and iRON networks. This is consistent with the results of SCAN (Fig. 5), where the western CONUS had relatively better accuracy. The USCRN was distributed across the CONUS being similar to the SCAN network, and so the RMSE of USCRN may be treated as an estimation of the overall inductive RMSE of the whole CONUS, being < 0.08 m^3/m^3 for $N > 21$. Moreover, the RMSE of USCRN was higher than that of SCAN in most cases, further demonstrating the importance of exploiting the distribution information of the target domain.

Fig. 9 shows the spatial performance of the MDSA on the 5 testing networks, with Fig. 9a illustrating the performance in 6 longitudinal regions and 6 latitudinal regions. The pattern of performance in longitude generally matched that of the SCAN network where the western regions had smaller bias, higher R and lower RMSE. Consequently, the MSDA was found to substantially reduce the bias, with more improvements observed in western regions. In contrast, the pattern along latitude was found to be more complex than that observed in the SCAN network. However, the RSME was still dominated by the bias, and the accuracy pattern along latitude changed little after using the MSDA.

Fig. 9b shows the site-specific RMSE, with the number of stations having large RMSE (> 0.1 m^3/m^3) being reduced from 68 to 35 in the western CONUS (125°W–110°W) when the MSDA was applied using one training station, further reduced to 22 when $N = 9$. However, it was 21 for $N > 17$, confirming the reduced benefit of using more training stations beyond a certain point. The RMSE of stations in the central and eastern regions also showed smaller changes than those in western regions. The number of stations with large RMSE (> 0.1 m^3/m^3) was 47, 39, 39, 35, 29 and 28 when $N = 0$, 1, 9, 17, 29 and 45 respectively. Similar to the results of the SCAN network (Fig. 5b), the RMSE of a few stations increased after applying the MSDA, as shown by the black dashed ellipses in Fig. 9b.

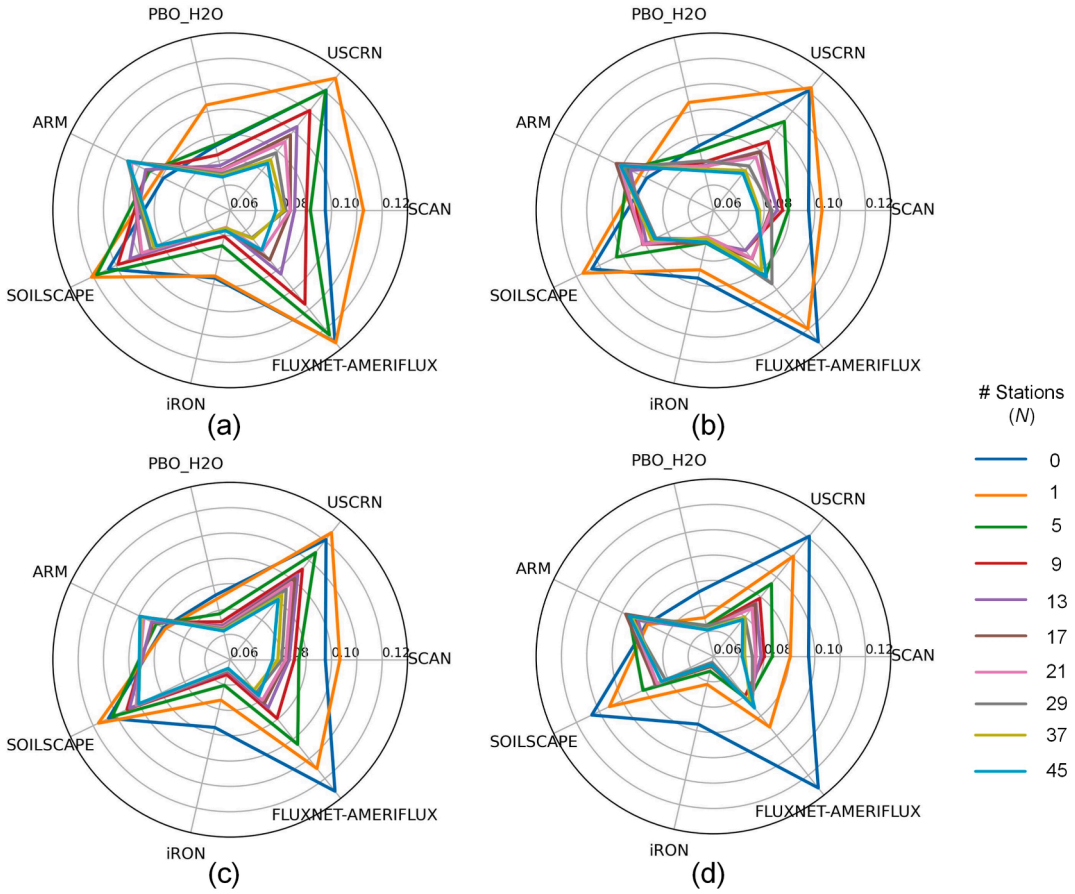


Fig. 8. Network specific RMSE of (a) Zhu et al. (2014), (b) MS, (c) DA and (d) MSDA.

5. Discussions

The proposed MSDA method was built on the 9 km pretrained models (Zhu et al., 2024), with two modifications, using multi-scale loss and domain loss. While the direct use of the 9 km pretrained models at 50 m achieved a relatively high RMSE of ~ 0.1 m^3/m^3, an acceptable bias of 0.03 m^3/m^3 and R of 0.631 were achieved with MSDA (Fig. 4). This confirmed that a pretrained DenseSM-E model for global application at a coarse resolution of 9 km can serve as an effective starting point for regional application at the higher resolution of 50 m, being much better than training from scratch, due to requiring lower computation costs and smaller training sets. Compared to the cross-resolution fine-tune method, which only uses samples at 50 m resolution (Zhu et al., 2024), the multi-scale loss was found to force the retrieved soil moisture to be closer to the SMAP soil moisture (Fig. 7), and so a transfer model was less likely to overfit at the beginning of the transfer learning and in scenarios with few training samples. The retrieval accuracy declined by up to 0.007 m^3/m^3 across all networks when training involved only a single station. In contrast, the MS method led to improvements in five out of the seven networks. A similar effect of multi-scale information was observed in another study (Liu et al. 2022), where coarse-grid loss acted as a constraint, reducing spurious overfitting, and in-situ data alone was found insufficient to constrain a soil moisture prediction model. Moreover, the use of multi-scale data aligned with the idea of learning from multiple teachers, found to be more robust in various applications, e.g., data assimilation (Pan et al. 2009). The incorporation of domain loss, specifically CORAL, in this study effectively reduced linear shifts between the source and target domains. This resulted in the smallest biases (<0.01 m^3/m^3) among the four transfer learning methods across all scenarios, along with notable improvements of approximately 0.05 in R. Since the CORAL is focused on reducing the first-order domain shift based on second-order statistics, further but probably marginal improvements may be achieved by considering higher-order shifts (Rahman et al. 2020). The effectiveness of jointly using the two loss terms has been demonstrated in both transductive (Fig. 6) and inductive (Fig. 8) modes. On the SCAN network, this method led to improvements in RMSE of 0.005–0.022 m^3/m^3 and in correlation (R) of 0.03–0.07. Moreover, the most favorable feature of the MSDA is the lower risk in scenarios with few training samples, outperforming the pretrained model with only 1 training station (Fig. 6). While two empirical parameters, α and β, were required to balance domain and multiscale losses (Eq. (3), an early balance approach (Eq. (4) was proposed to automatically determine β, with the results being largely insensitive to variations in α. The maximum difference caused by different α values was 0.004 m^3/m^3 (Table 3). For practical applications, a larger α (i.e., greater emphasis on the 50 m loss) is recommended in areas with a dense station network, whereas a smaller α is preferable in regions with fewer stations, as the 9 km loss helps prevent early overfitting on limited data.

A comprehensive evaluation was made in this study using the SCAN network and 6 testing networks in both transductive and inductive modes. Fig. 8 has confirmed that the evaluation results on the SCAN network were much better than those from another widespread low-density network (USCRN), with the average difference in RMSE being ~ 0.01 m^3/m^3. Accordingly, a model trained by the MSDA for a specific area/period is not recommended to be used in the inductive inference mode, as transductive learning can lead to better results. More specifically for the USCRN network, the unlabeled samples should be included in the calculation of domain loss to achieve better results. Moreover, stations with consistently high RMSE values exceeding 0.1 m^3/m^3 (e.g., those from the ARM network) were predominantly observed in the inductive mode, where RMSE was largely driven by substantial bias.

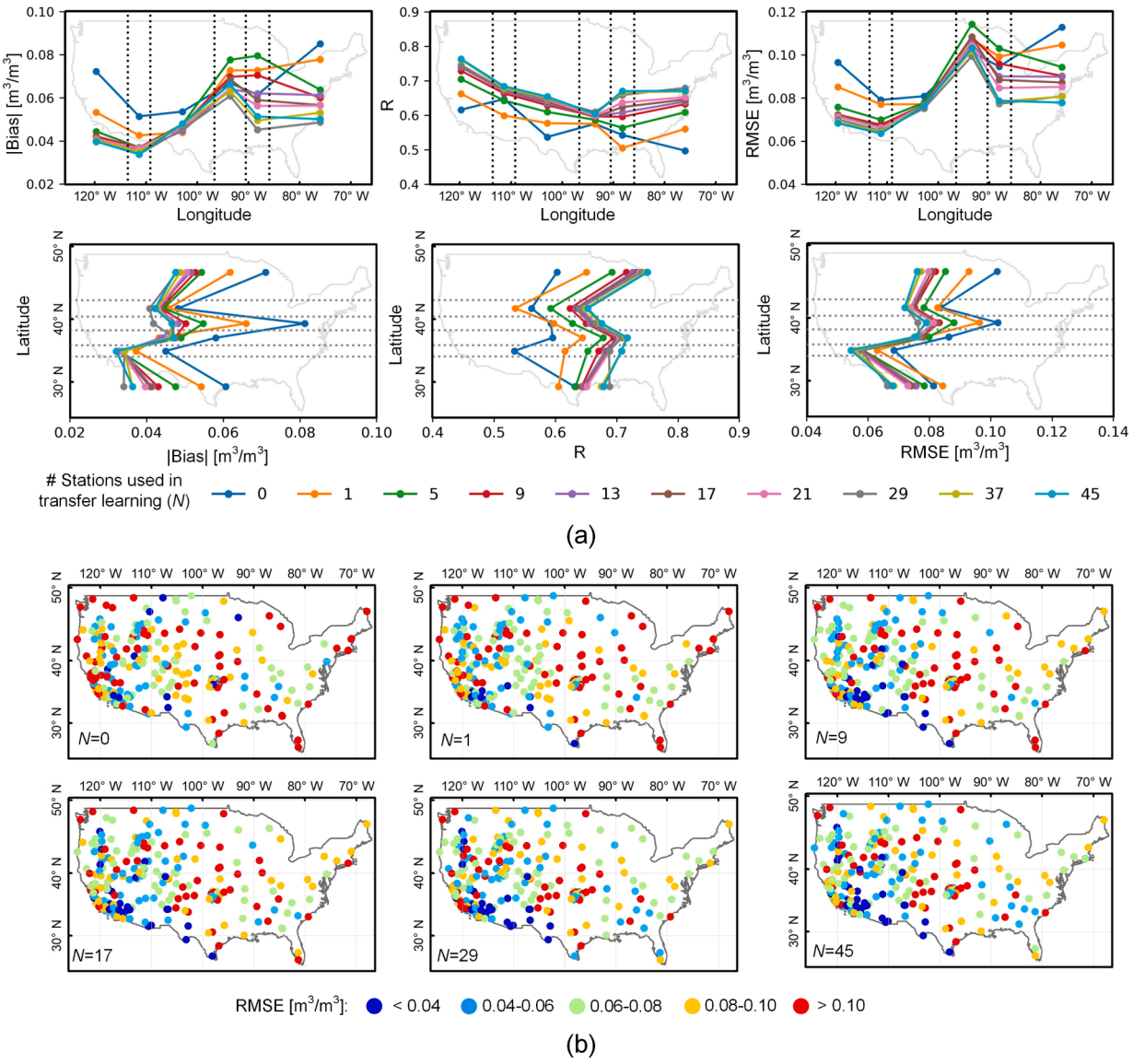


Fig. 9. Same as Fig. 5 but for the 6 testing networks in the inductive mode.

This suggests that the model trained on the SCAN network failed to capture the average soil moisture levels at these locations. Applying transductive learning, which leverages SMAP soil moisture to learn the average soil moisture distribution, is expected to reduce bias and consequently lower RMSE. Since unlabeled samples can always be prepared for a specific farm or region, the transductive mode thus is suggested for agricultural water management using the codes provided at https://github.com/rszlj/Transfer-DenseSM-E. However, this approach comes with the trade-off of a linearly increased computational cost due to learning from additional unlabeled samples.

The proposed MSDA achieved varying accuracy across the CONUS for the different networks in both evaluation scenarios. The RMSE for the drier SCAN stations in the central and western regions was significantly lower (0.058–0.065 m^3/m^3) compared to that of the wetter eastern SCAN stations (0.077–0.011 m^3/m^3). This trend was also observed across other networks; however, the relative error (%) remained comparable to that of wet areas and seasons (Fig. 5 and Fig. 9). The time series from three representative stations (Fig. 7) further confirmed larger deviations during wet seasons, indicating a systematic underestimation of high soil moisture values by the proposed MSDA method. Moreover, the experimental results indicated that the relative spatial performance of MSDA was dominated by the performance of the pretrained model (Fig. 8). Accordingly, the pretrained 9 km resolution model can be used to assess if more training samples are required for a specific area and period before applying the MSDA. The MSDA achieved an R and RMSE of 0.81 and ~ 0.07 m^3/m^3 respectively using 45 SCAN stations. This is worse than many existing studies that use more than 60 % of samples in training (Batchu et al. 2023; Karthikeyan and Mishra 2021). The performance of MSDA could be further improved using more training stations, but it suffers from the principle of diminishing marginal utility (Viering and Loog 2022). In addition, more stations in training means fewer stations in testing, making the conclusions less convincing for scenarios with scarce in-situ measurements. Another reason for not trying to achieve higher training or testing accuracy is the inherent limitation of in-situ soil moisture measurements in fully representing the soil moisture of a pixel (Balenzano et al. 2021). While the stations used in this study are generally installed in homogeneous areas with respect to soil texture, land cover, terrain features, and soil moisture—particularly within the nearby 50 m region—uncertainties remain. It is therefore questionable if a trained model perfectly matches the generated soil moisture truth at 50 m resolution.

Whilst the training and evaluation was focused on the 480 stations across the CONUS, the MSDA was developed with an operational global application in mind. The inputs of MSDA can be easily prepared using

the GEE Python/API, and running a trained model in the inference mode requires negligible computation resource. However, it is still challenging to produce an operational or even experimental soil moisture product, as GEE's computational capacity is not sufficient to support a deep learning-based production pipeline. Utilizing weather data and optical/thermal data offers a more straightforward approach for generating a global product at 1 km resolution, as demonstrated by recent work using traditional machine learning methods (e.g., Zheng et al. 2023). Another challenge is that a model without further training, i.e., running the model in the inductive mode, cannot achieve satisfactory results as demonstrated by the results of the pretrained model and the testing of MSDA on the USCRN. Applying the pertained 9 km resolution model and the MSDA to the area/periods of interest using a few stations can be more feasible for reliable soil moisture mapping. Fortunately, transfer learning has seen rapid advancements in recent years and thus presents a more powerful approach that requires less human intervention (Zhuang et al. 2020). Moreover, the MSDA target is surface soil moisture of ≤ 5 cm, whereas values for deeper layers are desired in agriculture water management (Jalilvand et al. 2023). Future works can thus explore extending the MSDA to a cross-layer domain adaption method, with the pre-trained models being trained using the existing root zone soil moisture products.

6. Conclusion

A multiscale domain adaption method was proposed with two modifications, including multiscale loss and domain loss. The performance was evaluated in 7 networks across the CONUS. The use of multiscale loss was found to reduce the risk of overfitting, while the use of domain loss reduced the retrieval bias. The joint use of the two loss items in the proposed MSDA achieved significantly better results ($p < 0.01$) and a lower risk of deterioration in scenarios with scarce in-situ measurements compared to the cross-resolution finetune method of Zhu et al. (2024). The overall improvement on the SCAN network was 0.005 – 0.022 m^3/m^3 in RMSE (5.8 % – 20.0 % in relative terms), and 0.004–0.014 m^3/m^3 (5.7 % – 14.8 % in relative terms) over the 6 testing networks. The performance of the MSDA varied in different scenarios, with the median RMSE being 0.088 m^3/m^3 and 0.071 m^3/m^3 when using 1 and 45 stations respectively. The direct use of the trained model on the SCAN network achieved an RMSE of $\sim$ 0.08 m^3/m^3 on the 6 testing networks across the CONUS.

CRediT authorship contribution statement

Liujun Zhu: Writing – original draft, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Qi Cai:** Writing – review & editing, Writing – original draft, Validation, Methodology, Data curation. **Junliang Jin:** Writing – review & editing, Resources, Conceptualization. **Shanshui Yuan:** Writing – review & editing, Funding acquisition. **Xiaoji Shen:** Writing – review & editing, Validation. **Jeffrey P. Walker:** Writing – review & editing, Investigation.

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.

Acknowledgments

This work was supported by the National Key Research and Development Program (2023YFC3209800) and the National Natural Science Foundation of China (42101374, 42371369 and 52121006). The authors express sincere thanks to the International Soil Moisture Network and its data providers listed in Table 2.

Data availability

Already shared at https://github.com/rszlj/Transfer-DenseSM-E

References

Abdalla, A., Cen, H., Wan, L., Rashid, R., Weng, H., Zhou, W., He, Y., 2019. Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure. *Comput. Electron. Agric.* 167, 105091.

Ahmad, S., Kalra, A., Stephen, H., 2010. Estimating soil moisture using remote sensing data: A machine learning approach. *Adv. Water Resour.* 33, 69–80.

Arias, M., Notarnicola, C., Campo-Bescós, M.Á., Arregui, L.M., Álvarez-Mozos, J., 2023. Evaluation of soil moisture estimation techniques based on Sentinel-1 observations over wheat fields. *Agric Water Manag* 287, 108422.

Baghdadi, N., Gaultier, S., King, C., 2002. Retrieving surface roughness and soil moisture from synthetic aperture radar (SAR) data using neural networks. *Can. J. Remote. Sens.* 28, 701–711.

Bai, X., Tan, S., 2023. Layered Soil Remote Sensing With Multichannel Passive Microwave Observations Using a Physics-Embedded Artificial Intelligence Framework: A Theoretical Study. *IEEE Trans. Geosci. Remote Sens.* 61, 1–12.

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. Sel. Top. Appl. Earth Obs. 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. *Journal of*. Hydrometeorology.

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.

Cao, M., Chen, M., Liu, J., Liu, Y., 2022. Assessing the performance of satellite soil moisture on agricultural drought monitoring in the North China Plain. *Agric Water Manag* 263, 107450.

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. (Basel)* 14, 5584.

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. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 3706–3718.

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. (Basel)* 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.

Dziugaite, G.K., Roy, D.M., & Ghahramani, Z. (2015). Training generative neural networks via maximum mean discrepancy optimization. *arXiv preprint arXiv: 1505.03906*.

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.

Fan, D., Zhao, T., Jiang, X., Xue, H., Moukomla, S., Kuntiyawichai, K., Shi, J., 2021. Soil Moisture Retrieval From Sentinel-1 Time-Series Data Over Croplands of Northeastern Thailand. *Ieee Geoscience and Remote Sensing Letters*.

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.

Huang, G., Liu, Z., Pleiss, G., Van Der Maaten, L., Weinberger, K.Q., 2017. Convolutional networks with dense connectivity. In: In, *Proceedings of the IEEE Conference on Computer Vision and Pattern recognitionIEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 4700–4708.

Jalilvand, E., Abolafia-Rosenzweig, R., Tajrishy, M., Kumar, S.V., Mohammadi, M.R., Das, N.N., 2023. Is It Possible to Quantify Irrigation Water-Use by Assimilating a High-Resolution Satellite Soil Moisture Product? *Water Resour. Res.* 59, e2022WR033342.

Karamouz, M., Alipour, R.S., Roohinia, M., Fereshtehpour, M., 2022. A remote sensing driven soil moisture estimator: Uncertain downscaling with geostatistically based use of ancillary data. *Water Resour. Res.* 58, e2022WR031946.

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.

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.

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

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.

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.

Ma, Y., Chen, S., Ermon, S., Lobell, D.B., 2024. Transfer learning in environmental remote sensing. *Remote Sens. Environ.* 301, 113924.

Miller, L., Zhu, L., Yebra, M., Rüdiger, C., Webb, G.I., 2022. Multi-modal temporal CNNs for live fuel moisture content estimation. *Environ. Model. Softw.* 156, 105467.

M. Moghaddam A. Silva D. Clewley R. Akbar S. Hussaini J. Whitcomb R. Devarakonda R. Shrestha R. Cook G. Prakash Soil Moisture Profiles and Temperature Data from SoilSCAPE Sites 2016 USA, ORNL DAAC, Oak Ridge, Tennessee, USA. In.

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.

Nguyen, T.T., Ngo, H.H., Guo, W., Chang, S.W., Nguyen, D.D., Nguyen, C.T., Zhang, J., Liang, S., Bui, X.T., Hoang, N.B., 2022. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. *Sci. Total Environ.* 833, 155066.

P. O'Neill S. Chan E. Njoku T. Jackson R. Bindlish J. Chaubell A. Colliander SMAP Enhanced L3 Radiometer Global and Polar Grid Daily 9 Km Ease-Grid Soil Moisture Version 5 2021 Boulder, Colorado USA.

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.

Paloscia, S., Pampaloni, P., Pettinato, S., Santi, E., 2008. A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR images. *IEEE Trans. Geosci. Remote Sens.* 46, 3274–3284.

Pan, M., Wood, E.F., McLaughlin, D.B., Entekhabi, D., Luo, L., 2009. A multiscale ensemble filtering system for hydrologic data assimilation. Part I: Implementation and synthetic experiment. *J. Hydrometeorol.* 10, 794–806.

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.

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.

Rahman, M.M., Fookes, C., Baktashmotlagh, M., Sridharan, S., 2020. On minimum discrepancy estimation for deep domain adaptation. *Domain Adaptation for Visual Understanding* 81–94.

Schaefer, G.L., Cosh, M.H., Jackson, T.J., 2007. The USDA natural resources conservation service soil climate analysis network (SCAN). *J. Atmos. Oceanic Tech.* 24, 2073–2077.

Stamenković, J., Ferrazzoli, P., Guerriero, L., Tuia, D., Thiran, J.-P., 2015. Joining a discrete radiative transfer model and a kernel retrieval algorithm for soil moisture estimation from SAR data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8, 3463–3475.

Sun B. Feng J. Saenko K. Domain adaptation in computer vision applications 2017 153 171.

Sun, B., & Saenko, K. (2016). Deep coral: Correlation alignment for deep domain adaptation. In, *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14* (pp. 443-450): Springer.

C. Tan F. Sun T. Kong W. Zhang C. Yang C. Liu A Survey on Deep Transfer Learning 2018 Springer 270 279.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., 2012. GMES Sentinel-1 mission. *Remote Sens. Environ.* 120, 9–24.

Viering, T., Loog, M., 2022. The shape of learning curves: a review. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 7799–7819.

Wagner, W., Lemoine, G., Borgeaud, M., Rott, H., 1999. A study of vegetation cover effects on ERS scatterometer data. *IEEE Trans. Geosci. Remote Sens.* 37, 938–948.

Wang, J., Wu, F., Shang, J., Zhou, Q., Ahmad, I., Zhou, G., 2022. Saline soil moisture mapping using Sentinel-1A synthetic aperture radar data and machine learning algorithms in humid region of China's east coast. *Catena* 213, 106189.

Wang, Y., Shi, L., Hu, Y., Hu, X., Song, W., Wang, L., 2024. A comprehensive study of deep learning for soil moisture prediction. *Hydrol. Earth Syst. Sci.* 28, 917–943.

Xu, W., Zhang, Z., Qin, Q., Hui, J., Long, Z., 2019. Soil moisture estimation with SVR and data augmentation based on alpha approximation method. *IEEE Trans. Geosci. Remote Sens.* 58, 3190–3201.

Zheng, C., Jia, L., Zhao, T., 2023. A 21-year dataset (2000–2020) of gap-free global daily surface soil moisture at 1-km grid resolution. *Sci. Data* 10, 139.

Zhu, L., Dai, J., Jin, J., Yuan, S., Xiong, Z., Walker, J.P., 2025. Are the current expectations for SAR remote sensing of soil moisture using machine learning over-optimistic? *IEEE Trans. Geosci. Remote Sens.* 63, 4501815.

Zhu, L., Dai, J., Liu, Y., Yuan, S., Qin, T., Walker, J.P., 2024. A cross-resolution transfer learning approach for soil moisture retrieval from Sentinel-1 using limited training samples. *Remote Sens. Environ.* 301, 113944.

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., Walker, J.P., Shen, X., 2020. Stochastic ensemble methods for multi-SAR-mission soil moisture retrieval. *Remote Sens. Environ.* 251, 112099.

Zhu, L., Walker, J.P., Tsang, L., Huang, H., Ye, N., Rüdiger, C., 2019. Soil moisture retrieval from time series multi-angular radar data using a dry down constraint. *Remote Sens. Environ.* 231, 111237.

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., 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.