Remote Sensing of Environment 339 (2026) 115408

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

DeepProfile: An inverse fusion framework for root zone soil moisture profile estimation

Liujun Zhu [a,b], Yi Tan [a,c], Shanshui Yuan [a,c], Junliang Jin [a,*], Zhengyang Tang [d], Jeffrey P. Walker [b]

[a] State Key Laboratory of Water Disaster Prevention, Yangtze Institute for Conservation and Development, Hohai University, Nanjing 210098, China
[b] Department of Civil and Environmental Engineering, Monash University, Clayton, VIC 3800, Australia
[c] College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China
[d] China Yangtze Power Co., Ltd. (CYPC), Wuhan, Hubei 430000, China

ARTICLE INFO

Editor: Jing M. Chen

Keywords:
Soil moisture
Root zone
Profile
Deep learning

ABSTRACT

Root zone soil moisture (RZSM) is a critical variable for understanding land–atmosphere interactions, hydrological processes, and agricultural productivity. Direct remote sensing of RZSM remains challenging due to the shallow sensing depth and the ill-posed nature of inversing a profile, with the existing global RZSM products mainly derived from model-based data assimilation. These products offer valuable information but exhibit inconsistent accuracy and disparate vertical discretizations. As no single root-zone soil moisture product is superior globally, fusing them offers a means to integrate their complementary strengths into a unified and consistent framework. In this study, a DeepProfile framework was proposed for estimating the continuous soil moisture profile throughout the top 100 cm layer of soil by integrating three widely used RZSM products; Soil Moisture Active Passive level 4 (SMAP L4), Global Land Data Assimilation System (GLDAS) version 2, and the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis (ERA5-land). Unlike traditional fusion methods requiring harmonized inputs, DeepProfile treats parent products as learning targets, optimizing the integral of a polynomial profile to match heterogeneous layers without enforcing identical vertical or spatiotemporal coverage. This yields a continuous analytical profile for flexible depth extraction, utilizing location-specific triple collocation weights to optimally balance product contributions. Evaluation against in-situ measurements from 2373 stations across 45 global networks demonstrated strong agreement in near-surface and intermediate layers ($\leq$50 cm), with median RMSE values below 0.06 m^3/m^3 and correlation coefficients (R) exceeding 0.72. The use of SMAP near-surface soil moisture was found critical for the satisfactory results for the top 50 cm. The model also showed promising performance at deeper layers of >50 cm ($R > 0.65$), although accuracy declined with depth due to weaker observational constraints. The proposed DeepProfile offers a scalable and transferable solution for generating depth-resolved soil moisture estimates, with potential applications in hydrological modeling, drought monitoring, and weather forecasting.

1. Introduction

Root zone soil moisture plays a crucial role in understanding land-atmosphere interactions, agricultural productivity, and hydrological processes (Bengough, 2012; Stocker et al., 2023). Accurate information on root zone soil moisture is vital for predicting droughts (Hirschi et al., 2014), optimizing irrigation, and improving water resource management (Zhang and Long, 2021). It also serves as a critical parameter in land surface models (Shrestha et al., 2018). Monitoring and modeling root zone soil moisture is therefore essential for advancing understanding of the terrestrial water cycle and ensuring effective adaptation to climate variability and change.

Remote sensing techniques have been widely used for soil moisture monitoring using optical/thermal, passive (radiometer) and/or active (radar) microwave sensors (Peng et al., 2020), with most studies focusing on the surface soil moisture (< 5 cm) due to their limited capabilities of penetrating into deeper soil layers. While some recent studies have made successful retrievals of soil moisture to 60 cm depth

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

https://doi.org/10.1016/j.rse.2026.115408
Received 14 November 2025; Received in revised form 31 March 2026; Accepted 3 April 2026
Available online 8 April 2026

using the longer wavelength P-band (Brakhasi et al., 2024; Li et al., 2024; Shen et al., 2024; Tabatabaeenejad et al., 2015) observations, they suffer from a varying reliable sensing depth in time and space as a result of the varying soil moisture profile (Brakhasi et al., 2023; Shen et al., 2020). Moreover, remote sensing of root zone soil moisture is challenging even with the knowledge of sensing depths, as the number of parameters for describing a soil moisture profile should be smaller than that of the independent observations for a stable inversion. While some constraints may be used in the inversion (Zhu et al., 2023), most studies have used 2–3 parameters to provide an over-simplified soil moisture profile (Brakhasi et al., 2024; Sadeghi et al., 2016; Tabatabaeenejad et al., 2015).

The alternative has been to estimate the soil moisture profile using a hydrology or land surface model (Hamman et al., 2018; Rodell et al., 2004), which divide the soil column into multiple layers representing different depths, and calculating the soil moisture content at each layer by solving the governing equations of soil water redistribution. Assimilation of in-situ and/or remote sensing measurements into these models has been shown to be critical for reliable estimates, with some commonly used methods including direct insertion, statistical correction, Newtonian Nudging, Kalman filtering and variational approaches (Vereecken et al., 2008). Several global root zone soil moisture products have been produced in this framework, such as the NASA Soil Moisture Active Passive (SMAP) level-4 soil moisture (Reichle et al., 2019), the Global Land Data Assimilation System version 2 (GLDAS-2) root zone soil moisture (Rodell et al., 2004), and the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis (ERA5) soil moisture (Hersbach et al., 2020).

These products have been evaluated at a regional to global scale using in-situ soil moisture (Das and Mohanty, 2006; Liu et al., 2024; Xu et al., 2021), being mainly based on the International Soil Moisture Network (ISMN, Dorigo et al., 2021). Statistical methods such as triple collocation (TC) and three-cornered hat have also been used for evaluation, which do not require any ground truth (Xu et al., 2021). While different evaluation results have been reported for varying regions under different evaluation schemes (Fan et al., 2022; Liu et al., 2024), none of the products clearly outperformed another, and thus multiple products are required in order to get an "optimal" root zone soil moisture estimate. However, a challenge of jointly using these products is the different spatial grids and layer thicknesses for which soil moisture is calculated, requiring a fusion approach in order to achieve a consistent data framework.

Fusion of multiple products is widely acknowledged as an effective solution to produce a more reliable product with high spatial-temporal continuity. A few studies have fused existing root zone soil moisture products using either the TC or three-cornered hat methods (Tian and Zhang, 2025), mostly requiring a pre-processing step that converts the source data to a consistent spatiotemporal grid and layer/depth system. In contrast, machine learning methods are more flexibly able to integrate data with varying spatial and temporal resolutions, such as in-situ soil moisture, soil proprieties, optical and thermal data (Fuentes et al., 2022). Despite the great flexibility, the soil moisture data included for fusion were generally used as inputs (Babaeian et al., 2021; Sahaar and Niemann, 2024), with spatial and/or temporal discontinuities and different time coverage. Thus, spatial and/or temporal gap filling along with temporal extrapolation is required. This commonly requires a complex, multi-stage workflow, often applying different interpolation or re-gridding techniques for each data source, thus restricting the ability to seamlessly integrate new products with differing spatiotemporal characteristics.

In this study, an inversion fusion strategy was proposed for estimating the soil moisture profile. The inversion fusion starts from building a deep learning model with the output being the soil moisture profile from 0 to 100 cm. Three root zone soil moisture products (ERA5-land, SMAP L4 and GLDAS) were treated as learning targets, while the state variables included in the SMAP L4 geophysical dataset were used as the input variables of the deep learning model. During the training stage, the model learned the root-zone soil moisture dynamics captured by these three products, with the TC method applied to determine the appropriate learning weights for each target based on its relative reliability. The trained model was then employed to produce a continuous, spatially distributed representation of the soil moisture profile. Unlike existing approaches that treat parent products as inputs and require identical start and end dates and continuous global coverage, the proposed approach does not require the parent products to share the same global spatial and temporal extent. Because the products serve as learning targets rather than inputs, the model can be trained on valid overlapping samples wherever they exist, without requiring the entire dataset to be aligned. This eliminates the need for spatial gap-filling or temporal interpolation to force the datasets into a uniform space-time cube, thereby avoiding the introduction of artificial uncertainties associated with these harmonization steps. Another advantage of inversion fusion is that it provides an analytical soil moisture profile estimation, and thus discrete soil moisture values for specific depths or layers can be calculated directly, seamlessly matching various applications that have different depth systems. A comprehensive evaluation of the proposed inversion fusion method has been made using the soil moisture stations available in the ISMN.

2. Data and preprocessing

2.1. Root zone soil moisture products

The SMAP L4, ERA-5 land and GLDAS soil moisture were selected for model training, considering their relatively higher resolutions and accuracy among the existing root zone soil moisture products (Xu et al., 2021). The SMAP L4 root zone soil moisture dataset is generated by integrating SMAP L-band brightness temperature (Tb) observations into the catchment land surface model using the ensemble Kalman filter data assimilation method. The assimilation algorithm combines L-band Tb measurements with the water and energy balance constraints embedded in the land surface model, along with meteorological forcing data. Since SMAP Tb was required to produce this product, the temporal coverage starts from the first release date of SMAP data (March 31, 2015). The SMAP L4 contains three sub-products and the geophysical data product (SPL4SMGP, Reichle et al., 2019), which was additionally used in this study. This product provides three soil moisture values every 3 h on the 9 km Equal-Area Scalable Earth Grids version 2 (EASE-2.0), being near-surface soil moisture (0–5 cm), root zone soil moisture (0–100 cm) and entire soil moisture profile being from the soil surface to the bedrock depth. Moreover, 43 extra variables are provided, including surface meteorological forcing variables, soil temperature, evapotranspiration, net radiation and error estimates for a few state variables. Since this study was focused on the daily averaged soil moisture profile, daily average values of 39 these variables were calculated (Table A1 in Appendix).

ERA5-land is an hourly product of surface variables with a spatial resolution of ~9 km (Muñoz-Sabater et al., 2021). The data is a replay of the land component of the ERA5 climate reanalysis, with a temporal coverage from 1950. ERA5-land provides four soil moisture values in m^3/m^3, being for the 0–7 cm, 7–28 cm, 28–100 cm and 100–289 cm layers. The hourly soil moisture of the first three layers were averaged on a daily basis, and the daily average values of these three layers projected to the EASE-2.0 grid of SMAP L4 using a bilinear interpolation.

GLDAS-2 simulations have three options, being GLDAS-2.0, GLDAS-2.1 and GLDAS-2.2, with the main difference being the varying forcing data and whether data assimilation is applied (Rodell et al., 2004). GLDAS-2.2 is produced using a data assimilation method and is thus expected to have a better accuracy than the other two options, but its simulation was forced with the ECMWF data which is consistent with the ERA5-land. Thus it was not used in this study, to ensure that more independent information was incorporated in the learning system. GLDAS-

2.0 was forced with the Princeton meteorological forcing input data covering only a historic period of 1984–2014, while GLDAS-2.1 was forced with a more comprehensive combination of model and observation data from 2000 to present. Accordingly, GLDAS-2.1 was selected in this study. GLDAS-2.1 has four soil layers being 0–10 cm, 10–40 cm, 40–100 cm and 100–200 cm, and so the soil moisture of the first three layers were used. These soil moisture values are provided in kg/m^3 and have a spatial resolution of 0.25 deg. They were converted to m^3/m^3 and reprojected to the EASE-2.0 grid of SMAP L4 using a bilinear interpolation.

2.2. Ground soil moisture observations

To evaluate the accuracy of the estimated soil moisture profiles, in-situ measurements from 2373 stations of 63 monitoring networks around the word were collected, covering the period 2015–2024. There were 2167 station records derived from the ISMN with the remaining 206 stations from the QLB-NET (Chai et al., 2024), Saihanba (Jiang et al., 2020), RAAM (Benninga et al., 2018), OzNet (Smith et al., 2012), Little Washita (Cosh et al., 2006) and Fort Cobb networks. Since this study focused on estimating daily soil moisture profiles at a spatial resolution of 9 km, soil moisture measurements collected from the same depth and calendar date were averaged as the truth. However, soil moisture values measured at a point can have large discrepancies with the 9 km estimates and the uncertainty caused by the scale mismatch varies in time and space (Peng et al., 2024). The existing calibration/validation activities of satellite missions commonly have a few core validation sites where dense stations are available for soil moisture upscaling (Colliander et al., 2017). However, they could not be directly used in this study as they were focused on the top 5 cm.

In this study, the daily averaged soil moisture values of stations within the same 9 km grid cell were averaged according to measuring depths. Two criteria were applied to select grid cells with sufficient ground truth: 1) the maximum sensing depth of the grid-averaged soil moisture must exceed 10 cm, and 2) the number of valid daily averaged measurements must be $\geq$ 180 (approximately six months). The selected grid cells were then grouped into “core” and “extra” categories, with core grid cells defined as those containing at least three working stations for spatial average. The filtering resulted in 54 core grid cells and 984 extra grid cells across 45 monitoring networks. Evaluation on the core grid cells is expected to be more reliable than that on the extra grid cells. A summary of these grid cells by network is provided in Table 1, with their spatial distribution shown in Fig. 1. Twelve networks containing core grid cells were selected for detailed analysis, with three networks chosen from each of the four major regions, i.e., North America, Europe, Asia, and Oceania. Measurement depths across the networks were categorized into 9 groups in the results section: 0–5 cm, 5–10 cm, 15 cm, 20 cm, 20–30 cm, 40–45 cm, 50 cm, 60–80 cm, and 100 cm. Notably, the 0–5 cm group represents the soil moisture content within the top 5 cm of the soil profile, whereas the remaining groups correspond to the specific depths at which measurements were taken. These 9 groups were specifically defined to aggregate in-situ stations into bins with sufficient sample sizes, which was necessary to robustly analyze the relationship between model accuracy and soil depth.

2.3. Preparation of samples for training and validation

A sample was defined as a combination of input variables X and target soil moisture values Y. The commonly used input variables used for soil moisture estimation include soil texture, terrain features, weather variables, vegetation descriptors, optical and/or microwave remote sensing observations (Zhu et al., 2025a; Zhu et al., 2025b). These factors are well represented by the geophysical variables provided in SMAP L4, which were therefore used as the primary inputs. These variables include meteorological forcing and a set of surface state variables, enabling global estimation of root zone soil moisture within a consistent

Table 1
A summary of the core and extra grid cells selected from 45 networks.

Network	Depths [cm]	# Stations per grid cell*		# Grid cells		Reference
		Core	Extra	Core	Extra	
BIEBRZA_S-1	5–50	9	9–18	1	2	(Cook, 2016)
CTP_SMTMN	0–50	3–7	1–4	3	26	(Yang, 2013)
Fort Cobb	5–45	3–4	3–4	2	2	
HOAL	5–50	32	4	1	1	(Blöschl et al., 2016)
IMA_CAN1	10–10	12	–	1	–	
KIHS_CMC/SMC	10–90	18–19	–	2	–	
Little Washita	5–45	6	–	1	–	(Cosh et al., 2006)
MySMNet	0–100	5	2	1	1	(Kang et al., 2016)
NAQU	5–80	3	1–4	1	4	(Su et al., 2011)
OZNET	0–90	3-8	1–4	6	18	(Smith et al., 2012)
PTSMN	7–43	9-11	–	2	–	(Hajdu et al., 2022)
QLB-NET	5–30	3–15	3–4	12	2	(Su et al., 2011)
RAAM	5–80	3-5	–	2	–	(Bogena et al., 2018)
SKKU	5–40	11	4	1	1	(Nguyen et al., 2017)
SMN-SDR	3–50	8	1	1	23	(Zhao et al., 2020)
SNOTEL	5–50	3	1–2	2	372	(Service, 2022)
SOILSCAPE	4–75	4–33	–	11	–	(Moghaddam et al., 2016)
SONTE-China	5–40	10	–	2	–	(Li et al., 2022)
Saihanba	5–10	12	–	1	–	
iRON	5–50	3	1–2	1	6	(Osenga et al., 2021)
AMMA-CATCH	5–135	–	1–2	–	5	(Galle et al., 2018)
ARM	2–175	–	1–3	–	15	(Galle et al., 2018)
COSMOS	0–101	–	1–4	–	10	(Zreda et al., 2008)
COSMOS-UK	0–50	–	1–2	–	49	(Cooper et al., 2021)
FLUXNET-AMERIFLUX	2–20	–	1–2	–	2	
FR_Aqui	5–90	–	1–3	–	3	(Al-Yaari et al., 2018)
HOBE	0–55	–	1–5	–	11	(Jensen and Refsgaard, 2018)
LABFLUX	10–40	–	1–2	–	3	
MAQU	5–80	–	1–3	–	14	(Su et al., 2011)
MOL-RAO	8–150	–	1	–	2	
NGARI	5–80	–	1–3	–	13	(Su et al., 2011)
NVE	10–60	–	1	–	2	
RISMA	0–150	–	1–4	–	14	(Ojo et al., 2015)
Ru_CFR	5–100	–	2	–	1	
SCAN	5–203	–	1–3	–	198	(Schaefer et al., 2007)
SD_DEM	5–200	–	1	–	1	(Ardö, 2012)
SMOSMANIA	5–30	–	1–3	–	20	(Calvet et al., 2016)
TAHMO	5–60	–	1–2	–	5	
TERENO	5–50	–	1	–	5	(Bogena et al., 2018)
TWENTE	5–80	–	1–5	–	20	(van der Velde et al., 2023)
UMBRIA	5–45	–	1	–	3	(Brocca et al., 2011)
USCRN	5–100	–	1–2	–	110	(Bell et al., 2013)

(continued on next page)

Table 1 (*continued*)

Network	Depths [cm]	# Stations per grid cell*		# Grid cells		Reference
		Core	Extra	Core	Extra	
VDS	1–20	–	1	–	4	(Fuchsberger et al., 2021)
WEGENERNET	20–30	–	1–2	–	2	
XMS-CAT	5–100	–	1–2	–	14	
Total	–	–		54	984	–

* The number of working stations within each grid cell vary in time and can be smaller than the values listed.

SMAP L4 framework. Moreover, elevation, slope and aspect were extracted from the Shuttle Radar Topography Mission Digital Elevation Data Version 3 (SRTM DEM V3) and used as supplementary inputs. Slope and aspect were first calculated at the native 30 m resolution and then resampled to the 9 km EASE-2.0 grid to preserve fine-scale terrain variations. The target variable Y consisted of soil moisture values from the aforementioned three root zone soil moisture products, covering 8 distinct layers.

The input variables were globally available from March 2015, resulting in billions of potential samples. While all samples could be used for model training, the benefit of additional data becomes negligible once a sufficient training size is reached. In this study, 2500 "other" grid cells were randomly selected, and their input variables extracted for use in the training and validation set. Since the core and extra grid cells are densely distributed in the Contiguous United States (CONUS), the 2500 other grid cells were selected from outside the CONUS region (Fig. 1). In total, approximately 11.5 million samples were extracted from the 54 core, 984 extra, and 2500 other grid cells. All variables were normalized to the range 0–1 using the min–max normalization method.

3. Methods

3.1. Architecture of DeepProfile model

The proposed DeepProfile architecture (Fig. 2) consists of a temporal feature extractor, a dense block and a regression block, following several successful architectures for surface parameter retrieval (Zhu et al., 2024; Zhu et al., 2021). The temporal feature extractor includes four sequential one-dimensional (1D) convolution (Conv1D) layers, each followed by batch normalization, a Linear rectification function (ReLU) activation layer, and an average pooling layer. The number of filters for the four Conv1D layers is 32, 32, 16 and 8 respectively, with corresponding average pooling sizes of 2, 2, 2 and 5. An approximately one-year long time series (360 days) of 39 SMAP L4 geophysical variables was fed into the temporal feature extractor to learn the temporal dynamics of meteorological forcing and surface state variables. After the pooling operations, the time series length was reduced sequentially to 180 (360/2), 90 (180/2), 45 (90/2) and finally 9 (45/5). The resulting temporal feature size was 72, calculated as 8 filters × 9 time steps.

The extracted temporal features were then concatenated with 6 auxiliary features and fed into the dense block, consisting of 6 sequential fully connected layers. Each layer was then batch normalized and a ReLU activation function applied. A fixed number of 128 neurons was used for all fully connected layers. Additionally, two shortcut connections were introduced to enable feature reuse and to enhance parameter efficiency (Huang et al., 2017b). The first shortcut connects the temporal feature block directly to the fourth fully connected layer, while the second passes the output of the third fully connected layer to the regression block. In total, 336 features were learned by the temporal feature extractor and the dense block, calculated as 80 + 128 + 128.

The regression block consists of n parallel regression networks. Each network shares an identical architecture comprising three fully connected layers, with batch normalization and ReLU activation applied after each layer. The first layer uses 32 neurons as a bottleneck to reduce the dimensionality of the learned features. The second layer expands to

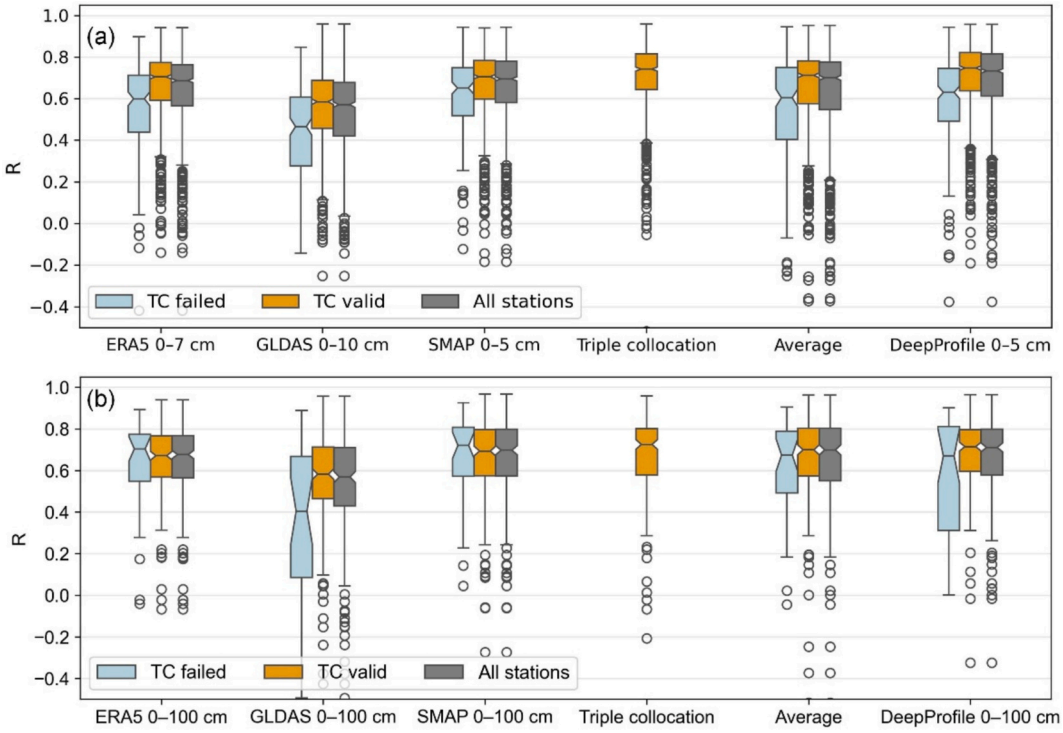

Fig. 1. Spatial distribution of core, extra, and other grid cells. Core and extra grid cells correspond to locations with in-situ measurements, while other grid cells were randomly selected for model training and validation. Twelve networks containing core grid cells are labeled, with three networks selected from each of the four major regions: North America, Europe, Asia, and Oceania.

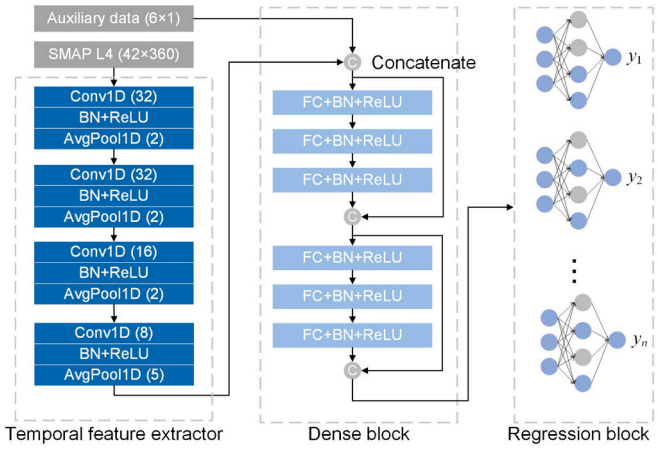

Fig. 2. The architecture of DeepProfie. The Conv1D, BN, ReLU, AvgPoorl1D and FC are the 1D convolution layer, batch normalization, activation layer of ReLU, 1D average pooling layer and fully connected layer, respectively. The numbers in the brackets are the dimensions of the input variables, referring to the number of filters for Conv1D and the pooling size of AvgPool1D. All of the FC layers in the dense block have 128 neurons, while the neuron numbers of the n identical regression networks are 32, 128 and 1 respectively.

128 neurons to capture specific regression relationships, followed by a single-neuron output layer. The outputs of DeepProfile are a set of n regression values $\{y_1, y_2, \ldots, y_n\}$, which can represent either soil moisture values at specific layers or analytical parameters describing the full soil moisture profile.

3.2. Inverse-fusion

The proposed fusion method treats multiple root zone soil moisture products as learning targets, allowing critical temporal and spatial patterns of root zone soil moisture to be learned through the back-propagation process (Fig. 3). Accordingly, this method was named as inverse-fusion, being different from most existing fusion methods where multiple sources are used as the inputs rather than targets. The target soil moisture profiles were generally represented by discrete values at specific soil layers, with varying layer systems applied across the different products leading to potential overlaps between layers. To accommodate these differences, a finer-resolution layer system was needed. As an alternative, the soil moisture profile can be described using an analytical function of depth. This study adopted the analytical approach, which offers greater flexibility for learning from heterogeneous layer structures and facilitates the future integration of additional soil moisture observations. Accordingly, the outputs of DeepProfile $\{y_1, y_2, \ldots, y_n\}$ represent the analytical parameters to be estimated.

Various analytical functions have been used to describe the soil moisture profile shape in remote sensing applications (Brakhasi et al., 2024; Njoku and Kong, 1977; Sadeghi et al., 2016; Shen et al., 2020), including exponential functions, polynomial functions and simplified solutions of the Richard Equation. The number of parameters to be determined was typically limited to less than 4 as the number of independent observations is generally less than 4 in microwave remote sensing. In contrast, DeepProfile allows for an arbitrary number of output parameters, constrained only by GPU memory. According to the Weierstrass Approximation Theorem, every continuous function defined on a closed interval can be uniformly approximated by a polynomial function. Polynomial functions were thus used in this study to represent the soil moisture profile. This avoided the need to handle multiple function types and provides flexibility for learning from diverse layer structures. The general form is

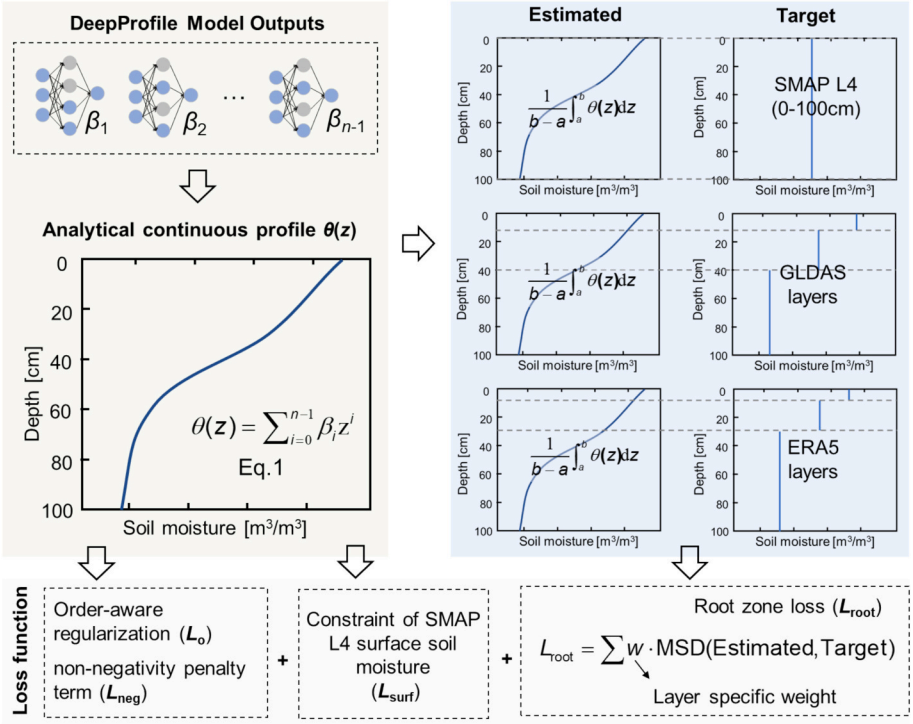

Fig. 3. Schematic of the inverse-fusion strategy implemented in DeepProfile. Unlike traditional fusion methods that merge inputs, the DeepProfile model outputs polynomial coefficients (β_i) to generate a continuous analytical soil moisture profile $\theta(z)$. This profile is then integrated over specific depth intervals to calculate layer-equivalent soil moisture values. These values are trained against the disparate target layers of ERA5-Land, SMAP L4, and GLDAS. The optimization is driven by the root zone loss (L_{root}), surface loss (L_{surf}) and two regularization terms (L_o and L_{neg}). The influence of each product layer is controlled by location-specific weights (w) derived from Triple Collocation, layer thickness in meter and the spatial scale factor. MSD refers to mean squared difference.

$$\theta(z) = \sum_{i=1}^{n-1} \beta_i z^i, z \in [0, 100] \quad (1)$$

where $\theta(z)$ is the soil moisture at depth z (0–100 cm), with the n unknown coefficients β needing to be determined for soil moisture estimation.

The three root zone soil moisture products provide the average soil moisture of several distinct layers. The corresponding estimations for each layer are expressed as

$$\theta(a,b) = \frac{1}{b-a}\int_a^b \theta(z)\mathrm{d}z \quad (2)$$

where a and b are the lower and upper boundaries of the soil layer, respectively. A total of 8 soil moisture layers were involved in this study, being the SMAP-L4 0–5 cm and 0–100 cm, ERA5-Land 0–7 cm, 7–28 cm and 28–100 cm, GLDAS 0–10 cm, 10–40 cm and 40–100 cm. A few studies have evaluated these products using ground measurements and/or statistical methods, all focusing on the average soil moisture across the full profile of 0–100 cm (Sahaar and Niemann, 2024; Tian and Zhang, 2025). Layer-specific weights cannot be directly derived from these evaluations. Accordingly, the commonly used TC method was used to estimate location-specific weights using the full records from March 2015 to February 2025.

TC is a statistical technique that estimates the error variance and correlation structure among three mutually independent data sources measuring the same variable (Stoffelen, 1998). Consequently, the estimated error variances are used as indicators of data reliability. Prior to applying TC, all soil moisture products were aggregated to a common 0–100 cm soil column depth. For ERA5-Land and GLDAS, this was done using

$$\theta_{\mathrm{ERA5}} = (7\theta_1 + 21\theta_2 + 72\theta_3)/100 \quad (3)$$

$$\theta_{\mathrm{GLDAS}} = (10\theta_1 + 30\theta_2 + 60\theta_3)/100 \quad (4)$$

where θ_1, θ_2 and θ_3 are volumetric soil moisture values for the respective layers. Assuming that the errors in the three products are uncorrelated with each other and with the unknown truth, the error variance of ERA5 were estimated as

$$\sigma^2_{\mathrm{ERA5}} = \mathrm{cov}(\theta_{\mathrm{ERA5}}, \theta_{\mathrm{SMAP}})\mathrm{cov}(\theta_{\mathrm{ERA5}}, \theta_{\mathrm{GLDAS}})/\mathrm{cov}(\theta_{\mathrm{SMAP}}, \theta_{\mathrm{GLDAS}}) \quad (5)$$

where cov is the covariance function. Error variances for SMAP and GLDAS were estimated in a similar way. The weight for ERA5 was defined as inversely proportional to its error variance, i.e., $1/\sigma^2_{\mathrm{ERA5}}$, which was further normalized using

$$W_{\mathrm{ERA5}} = 1/\sigma^2_{\mathrm{ERA5}}/\left(1/\sigma^2_{\mathrm{ERA5}} + 1/\sigma^2_{\mathrm{SMAP}} + s/\sigma^2_{\mathrm{GLDAS}}\right) \quad (6)$$

A spatial scale factor s was introduced specifically for GLDAS to account for the mismatch in spatial information content between its coarser grid (0.25°) and the finer resolution of the target profile (9 km or 0.1°). Since the coarser product represents a spatially aggregated average, it lacks the sub-grid variance present in the finer products. To prevent the optimization from suppressing valid high-resolution details in an attempt to fit the smoother coarse-scale target, s was assigned to a fixed value of 0.4. This value corresponds to the linear resolution ratio (0.1°/0.25°), effectively down-weighting the GLDAS contribution in the loss function to reflect its lower spatial fidelity relative to the estimation grid.

The weight for each ERA5 soil moisture layer can be further determined using a scaling factor proportional to the layer length according to:

$$w^j_{\mathrm{ERA5}} = l_j \cdot W_{\mathrm{ERA5}} \quad (7)$$

where l_j is the thickness of the soil layer j in meters (i.e., 0.07, 0.21, 0.72 m). According to Eqs. (6) and (7), the summation of all layer's weights is 1. A joint loss function (L_{root}) was defined taking the three root zone soil moisture products as the learning targets:

$$L_{\mathrm{root}} = w_{\mathrm{SMAP}}\mathrm{MSD}(\theta, \widehat{\theta}) + \sum_1^3 w^j_{\mathrm{ERA5}}\mathrm{MSD}(\theta_j, \widehat{\theta}_j) + \sum_1^3 w^j_{\mathrm{GLDAS}}\mathrm{MSD}(\theta_j, \widehat{\theta}_j) \quad (8)$$

where MSD refers to the mean squared difference between the estimated soil moisture (θ_j) and the corresponding value from products ($\widehat{\theta}_j$).

Apart from the 0–1 m root zone soil moisture, SMAP L4 also provides near-surface soil moisture for the 0–5 cm depth. This value was primarily derived from SMAP L3 near-surface observations and has been extensively validated worldwide. Incorporating reliable near-surface soil moisture as input has been shown to improve root zone soil moisture estimation (Babaeian et al., 2021). Accordingly, the MSD between the SMAP L4 and the estimated near-surface soil moisture was included as a second loss term L_{surf}. Moreover, to avoid overfitting and reduce the influence of high-order terms that may introduce oscillatory artifacts or hinder generalization, an order-aware regularization term was add to the loss function as

$$L_{\mathrm{o}} = \sum_{i=2}^{n} i^2 \cdot \mathrm{E}\left[\beta_i^2\right] \quad (9)$$

where n is the maximum polynomial degree in Eq. (1). $\mathrm{E}\left[\beta_i^2\right]$ represents the mean squared value of the i-th coefficient across samples. This term applies a weighted L2 penalty to the polynomial coefficients starting from the second-order term, with higher-order coefficients penalized more heavily based on the square of their degree. This approach limits excessive curvature in the fitted profile while preserving flexibility to capture key nonlinear patterns in soil moisture distribution.

To discourage negative soil moisture values, a non-negativity penalty term was introduced using the ReLU function

$$L_{\mathrm{neg}} = \sum \max(0, -\theta_{100\mathrm{cm}}) \quad (10)$$

where $\theta_{100\mathrm{cm}}$ is the predicted soil moisture at 1 m depth. Together with L_{surf}, this term helped ensure non-negative soil moisture values at both the surface and bottom boundaries, providing sufficient constraint for a physically realistic profile.

The final loss function used in this study is

$$L = L_{\mathrm{surf}} + L_{\mathrm{root}} + L_{\mathrm{neg}} + \lambda L_{\mathrm{o}} \quad (11)$$

where λ is the regularization coefficient, set to a small value such as 1e-3. The term L_{neg} typically becomes zero after a small number of training batches as the presence of negative values leads to a large penalty. This strong signal drives the model to eliminate negative predictions early in the training process. Eq. (11) also suggests that L_{surf} and L_{root} are equally important, considering that near-surface soil moisture provides a reliable constraint for profile estimation. The code and trained models are shared at https://github.com/rszlj/DeepProfile.

3.3. Training and testing

The prepared samples were extracted from 54 core, 984 extra, and 2500 other grid cells. Half of the extra and other grid cells (1742) were randomly selected, resulting in approximately 5.74 million samples for the training set. To evaluate the impact of training size on model performance, an additional 30% of extra and other grid cells were designated as candidate training samples. The remaining grid cells, together with all core grid cells, were assigned to the validation set. Model training was conducted using Eq. (11) as the loss function together with the Adam optimizer having default settings (Kingma and Ba, 2014). To balance computational efficiency and convergence stability, a batch size of 128 and a learning rate of 10^{-3} were adopted. Training proceeded for 50 epochs without early stopping. To enhance robustness and reduce variance, the final model parameters were obtained by averaging those

from the last five epochs, being a strategy known as snapshot ensemble (Huang et al., 2017a).

The training and validation fusion losses were first analyzed to evaluate the performance of the trained model and assess the influence of key hyperparameters. Specifically, the potential benefit of increasing the number of training samples was examined by increasing the training samples from 50% to 80% of the extra and other grids. The joint effect of n in Eq. (1) and λ in Eq. (11) was investigated, as these two parameters together control the shape and smoothness of the estimated soil moisture profile. Based on the validation fusion loss, an 'optimal' trained model was selected for subsequent analysis.

The 'optimal' trained model was further evaluated using ground truth data. Samples from the core grid cells were entirely unseen during training and were associated with reliable ground measurements. Four commonly used accuracy metrics were calculated for these samples, i.e., bias, correlation coefficient (R), RMSE and unbiased RMSE (ubRMSE). For the extra grid cells, which typically contain only one or two stations, the correlation coefficient was used as the evaluation metric to assess consistency between estimated and observed soil moisture. Other metrics were excluded due to their higher sensitivity to uncertainties in the ground truth data.

4. Results

4.1. Location specific weights of the three products

Fig. 4 presents the spatial distribution of weights assigned to the three source soil moisture products used during model training. Each grid cell was color-coded using an RGB fusion scheme, where the red, green, and blue channels represent the relative contributions of ERA5, GLDAS, and SMAP, respectively. The boxplots in the bottom-left corner of the figure summarize the distribution of these weights. SMAP achieved the highest median weight (0.367), followed by ERA5 (0.316), and GLDAS (0.258). Although SMAP had a slightly higher median value, the differences among the three products were relatively modest, suggesting that the fusion model maintained a balanced integration of the three sources across most regions. GLDAS had the lowest median weight due to the scale factor applied in Eq. (6). However, its error standard deviation estimated by TC was lower than that of SMAP and ERA5, being consistent with Xu et al. (2021).

A notable feature in the map was the presence of black points, which accounted for approximately 22.1% of the total grid cells. These points represented locations where weights could not be determined using TC. Unlike the RGB-colored areas, these regions lacked consistent data across the three products, making TC application infeasible. The spatial distribution of these black points did not show any clear geographic pattern, appearing scattered across various climatic zones and land cover types. As a result, the fusion model defaulted to using the median weights of ERA5, GLDAS, and SMAP for these locations.

The RGB fusion map revealed clear spatial patterns in the dominance of individual products. In tropical rainforest regions such as the Amazon Basin, GLDAS contributed the most, likely due to its robust land surface modeling in humid, vegetated environments. In contrast, arid and sparsely populated regions like the Sahara Desert were dominated by ERA5. SMAP showed higher weights in regions such as the western United States, where its extensive ground soil moisture measurement networks (Fig. 2) were already partly used in the production of the products. Similarly, southern China exhibited a strong GLDAS signal, consistent with its performance in subtropical climates with dense vegetation and frequent precipitation. These spatial variations in product weighting highlight the importance of region-specific fusion strategies. In this study, such adaptability was achieved by assigning grid-cell-specific weights, even though the overall contributions of each product remained relatively balanced.

4.2. Behaviors of training and validation loess

Fig. 5a shows the training and validation loss curves across 50 epochs, offering insights into the model's learning dynamics. Both losses decreased steadily as training progressed, with convergence generally achieved around epoch 15. The consistently narrow gap between training and validation losses indicates strong generalization performance and minimal overfitting. Notably, the validation loss was slightly lower than the training loss, which may be attributed to the regularization effects or the stochastic nature of the validation sampling. The losses for surface soil moisture (L_s) and root zone soil moisture (L_r) decreased in a synchronized manner, highlighting the model's ability to jointly optimize both targets under the collaborative loss framework. In contrast, the loss related to profile shape (L_o) dropped sharply during the initial 5 epochs but then fluctuated and slightly increased, suggesting that the model struggled to consistently capture the structural patterns of soil moisture profiles. Among the three loss terms, L_r remained the largest, primarily due to inconsistencies among the three source products used in training.

Fig. 5b illustrates the impact of an increased training sample size on the model's validation losses for L_s, L_r, and L_o. Contrary to expectations, the variation in validation losses remained minimal as more training samples were added, indicating that the model's generalization performance did not substantially improve with additional training data. In some cases, the effect was slightly negative, suggesting diminishing returns. This phenomenon aligns with the observation in Fig. 5a, where the training and validation loss curves nearly overlap, implying that the model had already achieved a high level of generalization with the initial dataset. The results suggest that approximately 50% of the training grid cells exceed the saturation threshold, beyond which further data contributes little to performance gains. Given that the grid cells were randomly selected across space, this also points to a degree of spatial redundancy in the training data.

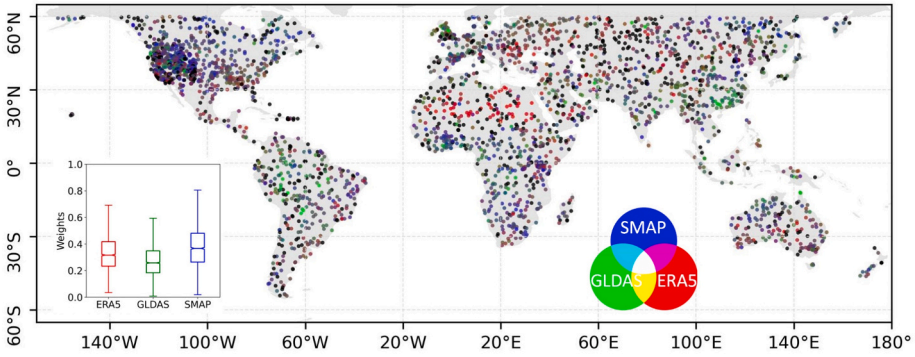

Fig. 4. Spatial distribution of weights estimated for the three source soil moisture products together with the boxplots (bottom left corner) of each product's weights. The color of each point is an RGB fusion based on the weights of the three products. Grey points are these with similar weights for three products.

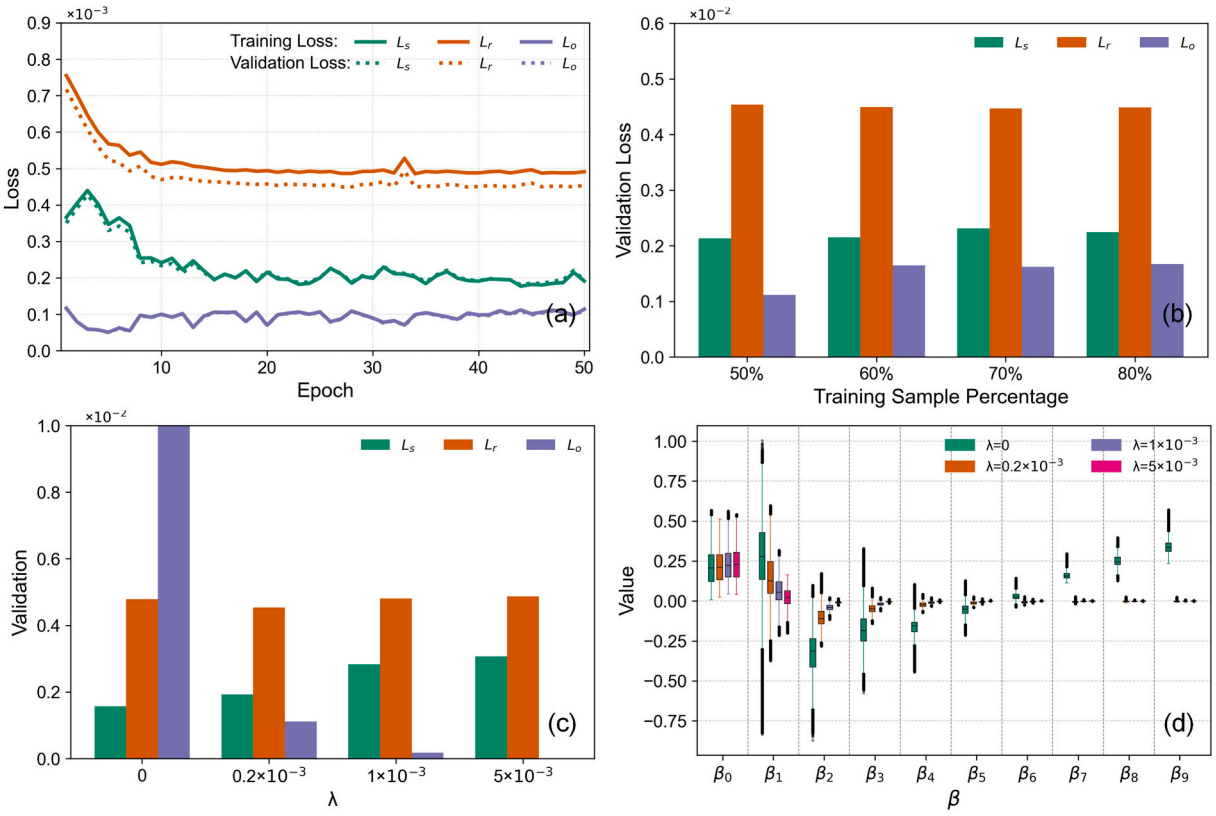

Fig. 5. The training and validation loss with respect to epochs (a), the effect of training sample size on the validation losses (b), the effect of regularization parameter λ on the validation losses (c), and the distribution of betas using different λ (d). L_s, L_r and L_o are the losses of surface soil moisture, root zone soil moisture and profile shapes in Eq. (11).

Fig. 5c and d examine the influence of the regularization parameter λ on the model's validation loss and the structural characteristics of the estimated soil moisture profiles. In the absence of regularization (λ = 0), the model exhibited the greatest flexibility in shaping the soil moisture profiles, resulting in high-frequency fluctuations from the surface to 100 cm depth. This is reflected in the elevated values of higher-order β coefficients (e.g., β_7, β_8, β_9), which dominate the profile representation. Such behavior suggests that without regularization, the model tended to overfit the profile shape, capturing noise or unrealistic variations that are not physically plausible. Introducing a small regularization (e.g., $\lambda = 0.2 \times 10^{-3}$) substantially reduced the profile shape loss (L_o), while only slightly increasing the surface soil moisture loss (L_s). In this case, the higher-order β coefficients were nearly zero, indicating that the estimated profiles were effectively constrained to low-order polynomials (typically third-order) resulting in smoother and more physically consistent moisture gradients. Further increases in λ continue to suppress high-frequency components, but also led to marginal increases in L_s, suggesting a trade-off between smoothness and surface accuracy. Notably, the impact of λ on root zone soil moisture loss (L_r) remained negligible across all tested values, implying that regularization primarily affects the vertical structure rather than the integrated moisture content. Given that real soil moisture profiles rarely exhibit sharp fluctuations within the top 100 cm, a λ value of 0.2×10^{-3} was selected, balancing physical realism with predictive accuracy.

4.3. Inter-comparison with the three sourced products

Soil moisture for any depth within the 0–100 cm range can be calculated using Eq. (2), allowing direct inter-comparison with the three source products. Fig. 6 shows the root mean square difference (RMSD) between the estimated soil moisture and the 8 layers from SMAP L4, ERA5-land, and GLDAS over the validation grid cells. Among these, the estimation achieved the highest agreement ($R = 0.949$, RMSD = 0.035 m^3/m^3, bias = 0.0247 m^3/m^3) with the SMAP L4 0–5 cm value. This superior consistency was mainly due to the larger weight assigned to the SMAP L4 0–5 cm soil moisture during training as shown in Eq. (11), which constrained the model to better fit SMAP L4 0–5 cm with an improved consistency on the testing set. Spatially, the estimations had widespread low RMSD around the world with respect to SMAP L4 0–5 cm. At similar depths of near-surface soil moisture, the estimations achieved moderate and consistent RMSD values of 0.059 m^3/m^3 and 0.060 m^3/m^3 with ERA5 0–7 cm and GLDAS 0–10 cm respectively, both with small negative biases. However, distinct spatial patterns of RMSD were observed between ERA5 and GLDAS. The RMSD of ERA5 0–7 cm was generally larger than 0.12 m^3/m^3 across much of Asia and had relatively higher values in South America, whereas GLDAS 0–10 cm had large RMSD in Russia and northwestern North America.

At intermediate depths, the estimations maintained comparable correlation (R = ~0.79) with ERA5 7–28 cm and GLDAS 10–40 cm. The median RMSD values (~0.047 m^3/m^3) were substantially lower than those of their near-surface layers (0.060 m^3/m^3). This difference can be attributed to the fact that near-surface soil moisture was strongly influenced by the SMAP L4 0–5 cm soil moisture, which does not provide detailed information for intermediate depths. Consequently, soil moisture at these depths was primarily determined by ERA5 7–28 cm and GLDAS 10–40 cm. Both layers achieved relatively small RMSD in eastern North America and western Europe, while areas with higher RMSD were scattered across the globe without a clear spatial pattern. For the deepest layers, the estimations achieved a consistent median RMSD of ~0.043 m^3/m^3 with GLDAS 40–100 cm and ERA5 28–100 cm. The spatial RMSD pattern of each layer was generally consistent with its corresponding intermediate depth.

SMAP L4 provides a single root zone layer spanning 0–1 m, which had widespread low RMSD over regions with low-to-moderate

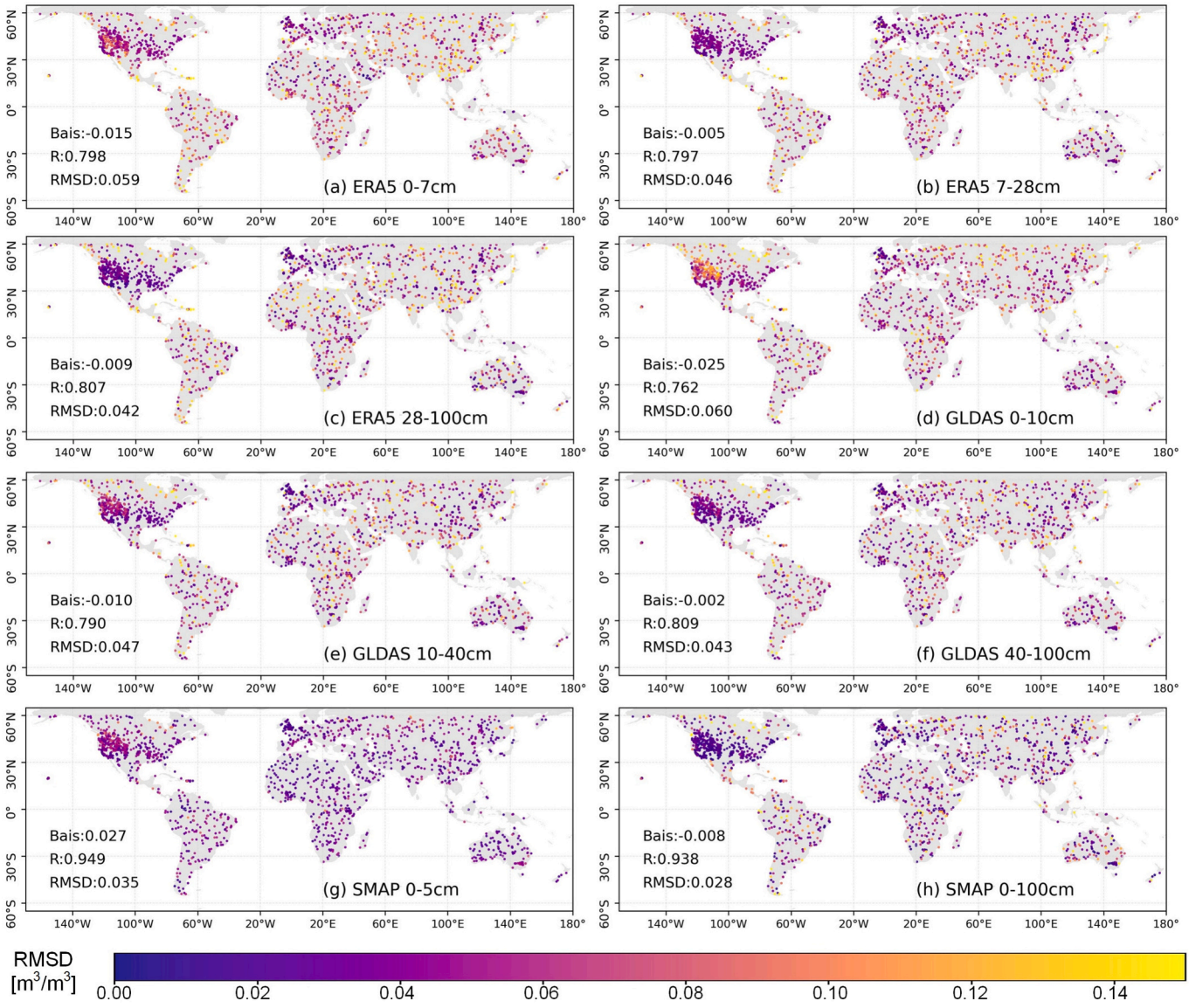

Fig. 6. Root mean square difference (RMSD) between the estimated soil moisture and the 8 layers from the three soil moisture products involved in the fusion. (a)-(h) are the results for ERA5 0–7 cm, ERA5 7–28 cm, ERA5 28–100 cm, GLDAS 0–10 cm, GLDAS 10–40 cm, GLDAS 40–100 cm, SMAP 0–5 cm and SMAP 0–100 cm, respectively. The median bias, R and RMSD of all grid cells is also provided at the bottom left corner of each panel.

vegetation such as the Great Plains, European lowlands, and eastern China, while relatively higher RMSD was observed in dense tropical forests and snow-affected latitudes. These discrepancies may result from greater uncertainties in SMAP L4 retrievals under dense canopy and frozen conditions. The SMAP L4 0–1 m layer achieved a small RMSD of 0.028 m^3/m^3 and a high correlation of 0.938 with the estimated soil moisture, being substantially better than the ERA5 and GLDAS layers. This advantage is partly related to differences in layer thickness. According to Eq. (4) and Eq. (5), the weight of each ERA5 or GLDAS layer was normalized by the total thickness of 1 m. Consequently, the SMAP L4 0–1 m layer received a larger weight than individual ERA5 or GLDAS layers, which partially explains why RMSD decreased from 0.059 m^3/m^3 to 0.042 m^3/m^3 across the three ERA5 layers.

4.4. Evaluation using ground measurements

Fig. 7 presents violin plots illustrating model performance across the 9 soil depth intervals. Each plot depicts the distribution of performance metrics across core grid cells, with black dots representing individual grid cell values. The depth intervals were defined based on the availability of core grid cells, which decreased with increasing depth. Bias values were generally centered around zero across most depths (Fig. 7a), indicating small systematic overestimation or underestimation. Shallower depths (<10 cm) show narrow distributions with low variance, while deeper layers such as 60–80 cm and 100 cm show broader spreads and slightly increased bias variability.

The R values were highest in the upper soil layers (Fig. 7b), with median values close to 0.8 in the 0–5 cm and 5–10 cm intervals, indicating strong agreement between estimations and observations. No clear depth-dependent trend in R was observed, suggesting balanced predictive skill across different depths. However, the spread of R values increased with depth, reflecting reduced model consistency in deeper layers. This pattern was also observed in RMSE. The lowest median RMSE values, around 0.05 m^3/m^3, were observed at 15 cm and 50 cm, although these depths included fewer than 10 grid cells. Across the top 20 cm and intermediate layers such as 40–45 cm and 50 cm, RMSE values remained generally consistent, with medians below 0.06 m^3/m^3. Estimation errors were substantially reduced after removing grid cell-specific bias, with all median ubRMSE values falling below 0.05 m^3/m^3. A clear decreasing trend in ubRMSE with depth indicates that grid-

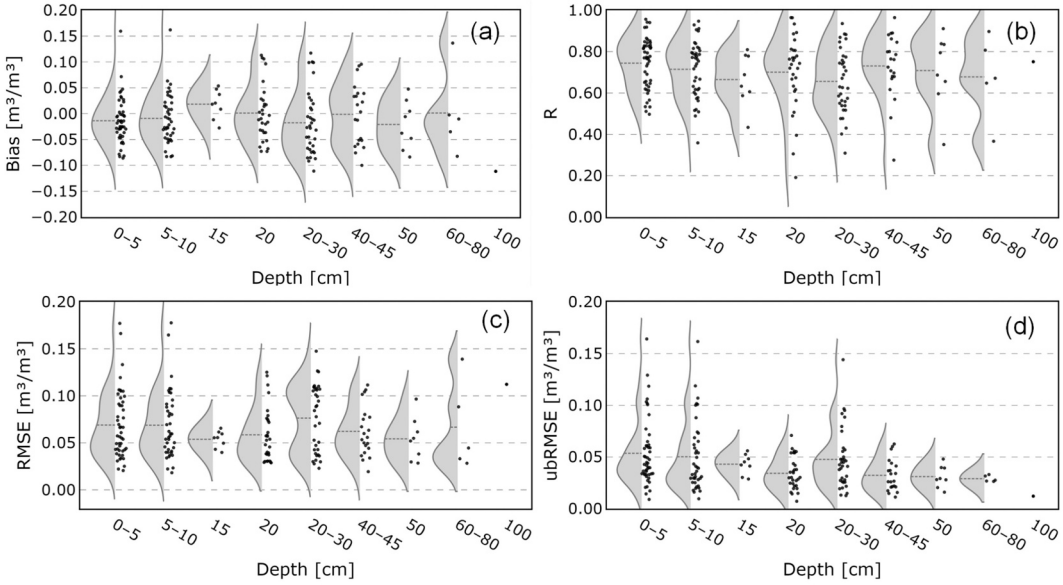

Fig. 7. Performance on the core grid cells. Violin plot is for 9 depths/depth groups. Each black point represents a core grid cell, with the number of available core grid cells reducing as the depth increased.

specific bias contributed more substantially to retrieval errors in deeper layers.

Fig. 8 presents the evaluation results of the proposed method across 12 selected networks containing core grid cells, as identified in Fig, 1. Among the three North American networks, the method achieved the lowest errors and highest correlations on Fort Cobb, with RMSE values below 0.04 m^3/m^3 for the upper two layers. On the SoilSCAPE network, which included the largest number of core grid cells, the method maintained consistent accuracy across depths from 0 to 50 cm, with RMSE values remaining below 0.06 m^3/m^3. In contrast, higher RMSEs (0.07 to 0.08 m^3/m^3) were observed on SNOTEL, likely due to the complex terrain characteristics of its grid cells.

Across the three European networks, the method generally performed well, particularly on HOAL and IMA_CAN1. Satisfactory accuracy was also achieved on RAAM, where a decreasing ubRMSE trend was observed from 0.051 m^3/m^3 at 5 cm to 0.028 m^3/m^3 at 80 cm. However, substantial depth-dependent bias was present, with RMSE at 80 cm largely driven by bias, suggesting that the estimated profiles for RAAM did not fully capture vertical soil moisture variation.

A similar pattern was observed for SONTE-China in Asia, where absolute bias increased from 0.02 m^3/m^3 at 5 cm to 0.075 m^3/m^3 at 20 cm before decreasing to 0.058 m^3/m^3 at 40 cm, resulting in a depth-dependent RMSE. In contrast, relatively stable and satisfactory accuracy across depths were achieved on the other two Asian networks. For Oceania networks, increasing RMSE with depth was observed on MySMNet and OZNet, primarily driven by growing bias. However, PTSMN achieved the lowest overall RMSE (~0.03 m^3/m^3) and the highest correlation (~0.89) among the Oceania networks, indicating comparatively better performance in this region.

Fig. 9 shows the spatial distribution of R between estimated soil moisture and in-situ measurements, including the extra grid cells for an extended evaluation. Overall, the model achieved acceptable performance across most regions, with a median R of 0.729 at the top 5 cm layer (Fig. 9a). Large uncertainties were limited to a few grid cells, primarily located in Hawaii, central United States, Puerto Rico, South Korea, and the Qinghai-Tibet Plateau. Similar results were observed for the 5–10 cm interval (Fig. 9b), with a median R of 0.723. Although only a small number of grid cells were located south of 10°N latitude, they all achieved R values above 0.75 in the top 10 cm. The lowest median RMSE of 0.073 m^3/m^3 was observed at 15 cm; however, these grid cells were mostly located in the United Kingdom and may not reflect global performance. In contrast, results at 20 cm depth were notably poorer, with many grid cells showing R values <0.2, particularly in the western United States, Alaska, and China.

At the intermediate depths (Fig. 9e and f), grid cells were more scattered globally. European sites generally showed better performance, while those in western China had the lowest R values. Among the deeper layers (Fig. 8g, h, and i), the 60–80 cm interval included only 89 grid cells, with relatively better performance observed in Europe. Grid cells at 50 cm and 100 cm depth were mainly clustered in the CONUS, where R values at 50 cm were substantially higher than those at 100 cm, with the median value dropping from 0.617 to 0.406. Fig. 9j presents violin plots of R for each depth or depth interval, along with a fitted polynomial curve for the median R values. A clear decreasing trend is observed, with performance degrading more rapidly in deeper layers than in shallow ones.

The performance of the proposed DeepProfile model was further investigated across different land cover types (Fig. 10a) and soil textures (Fig. 10b). The performance of the proposed DeepProfile model was further investigated across different land cover types (Fig. 10a) and soil textures (Fig. 10b). Land cover data was extracted from the MCD12C1 V006 (Friedl and Sulla-Menashe, 2015), while the soil texture was extracted from the Harmonized World Soil Database v2.0. Categories or depths with fewer than 20 grid cells were excluded to ensure robustness. Consistent with general trends (Fig. 9j), the median R for all land cover types decreased as soil depth increased. However, this performance deterioration was less pronounced in forest areas compared to other types. In the near-surface layers (0–5 cm and 5–10 cm), cropland achieved slightly higher median R values compared to grassland and forest areas, though a Student's *t*-test indicated that this difference was statistically significant ($p < 0.05$) only between grassland and cropland at the 5–10 cm depth. For the deeper layers (50 cm and 100 cm), although median R values exhibited larger variations across land cover types, the differences were generally statistically insignificant. However, this lack of significance can be partly attributed to the limited sample size (number of available grid cells) at these depths. Similarly, performance differences among soil textures were found to be statistically insignificant across all depths. Sandy soils had slightly lower median *R* values in the 0–5 cm and 5–10 cm layers but achieved higher correlations at the 20 cm depth. Clay loam maintained relatively stable median correlations at intermediate depths (≤ 50 cm) but had a wide spread of outliers at 100 cm, likely reflecting the heterogeneity of moisture retention

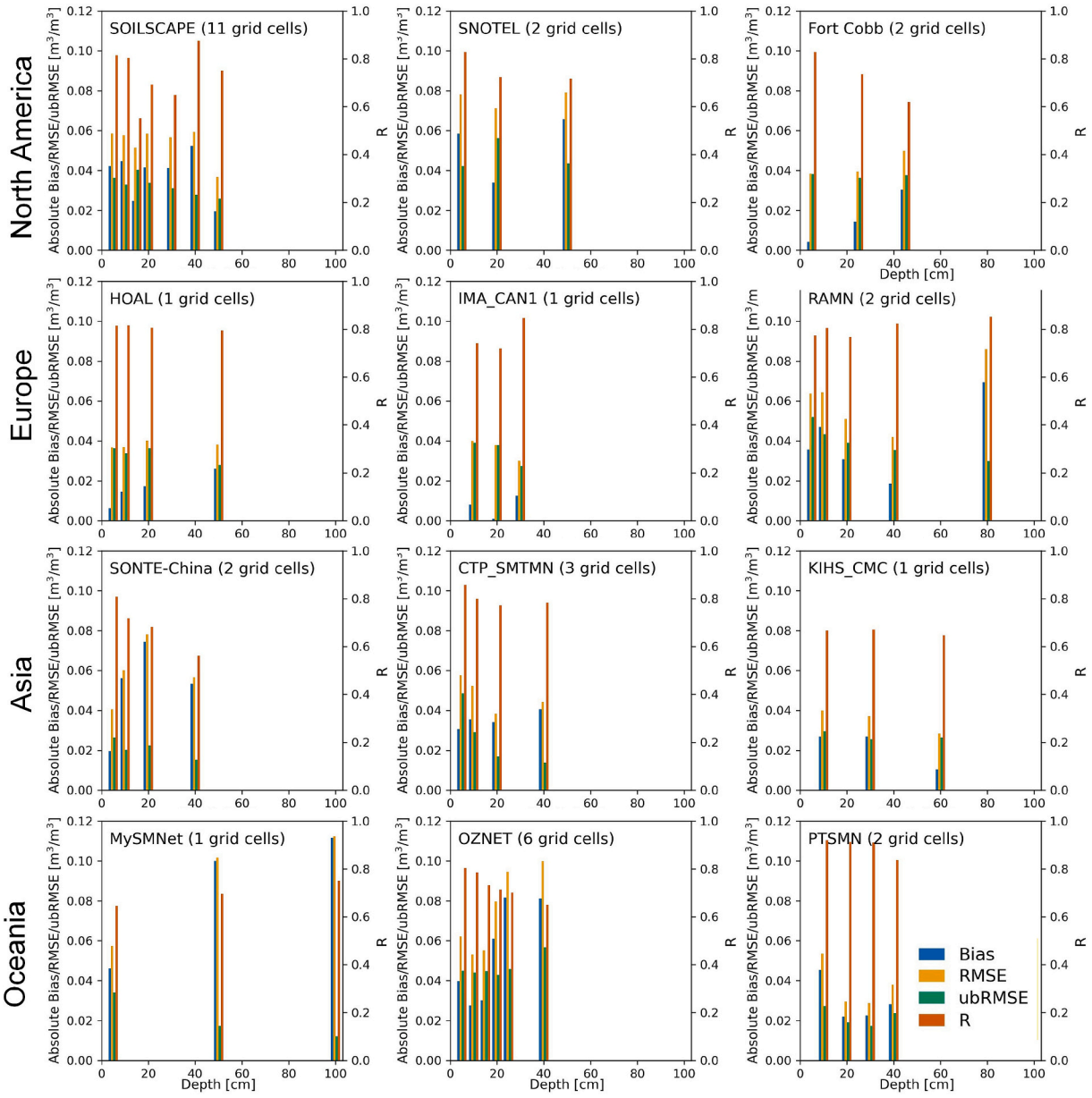

Fig. 8. Performance on the twelve selected networks containing core grid cells, as identified in Fig. 1. Three networks were selected from each of the four major regions: North America, Europe, Asia, and Oceania.

characteristic of finer-textured soils.

The effect of the TC status was investigated in Fig. 10c. In this analysis, grid cells were categorized based on whether the source products satisfied the statistical assumptions required for TC. The validation revealed that the "TC valid" subset consistently outperformed the "TC failed" and "All" stations groups across all depths, maintaining higher median R values and narrower distributions. Even at the 100 cm depth, where overall correlations are lowest, the "TC valid" locations maintained a median R noticeably higher than the "TC failed" group. This persistent performance gap indicates that locations where source products violate TC assumptions tended to yield lower agreement with in-situ observations.

5. Discussion

The proposed inversion fusion strategy introduced a paradigm shift in root zone soil moisture estimation by reframing the fusion problem as a supervised learning task with multiple products as learning targets rather than inputs. This inverse fusion strategy removed the need for spatial and temporal harmonization of heterogeneous datasets, enabling the model to learn from partially overlapping products with differing layer systems. Soil moisture profiles are represented analytically using flexible polynomial functions, enabling smooth interpolation across arbitrary depths and the direct extraction of layer-specific values for diverse applications. The DeepProfile architecture further strengthened this capability by incorporating temporal dynamics and auxiliary spatial features through a modular deep learning design with shortcut connections. The proposed inversion fusion addresses key limitations of traditional fusion methods, including rigid input requirements, oversimplified profile structures, and the need for gap-filling. It also partially mitigates the uncertainty of triple collocation by using TC-derived weights for the sole purpose of guiding loss balancing, rather than directly determining outputs. However, as shown in Fig. 5, the assumption of mutual independence among the three source products was violated in approximately 22% of grid cells globally. This violation coincided with elevated estimation uncertainty and notably lower

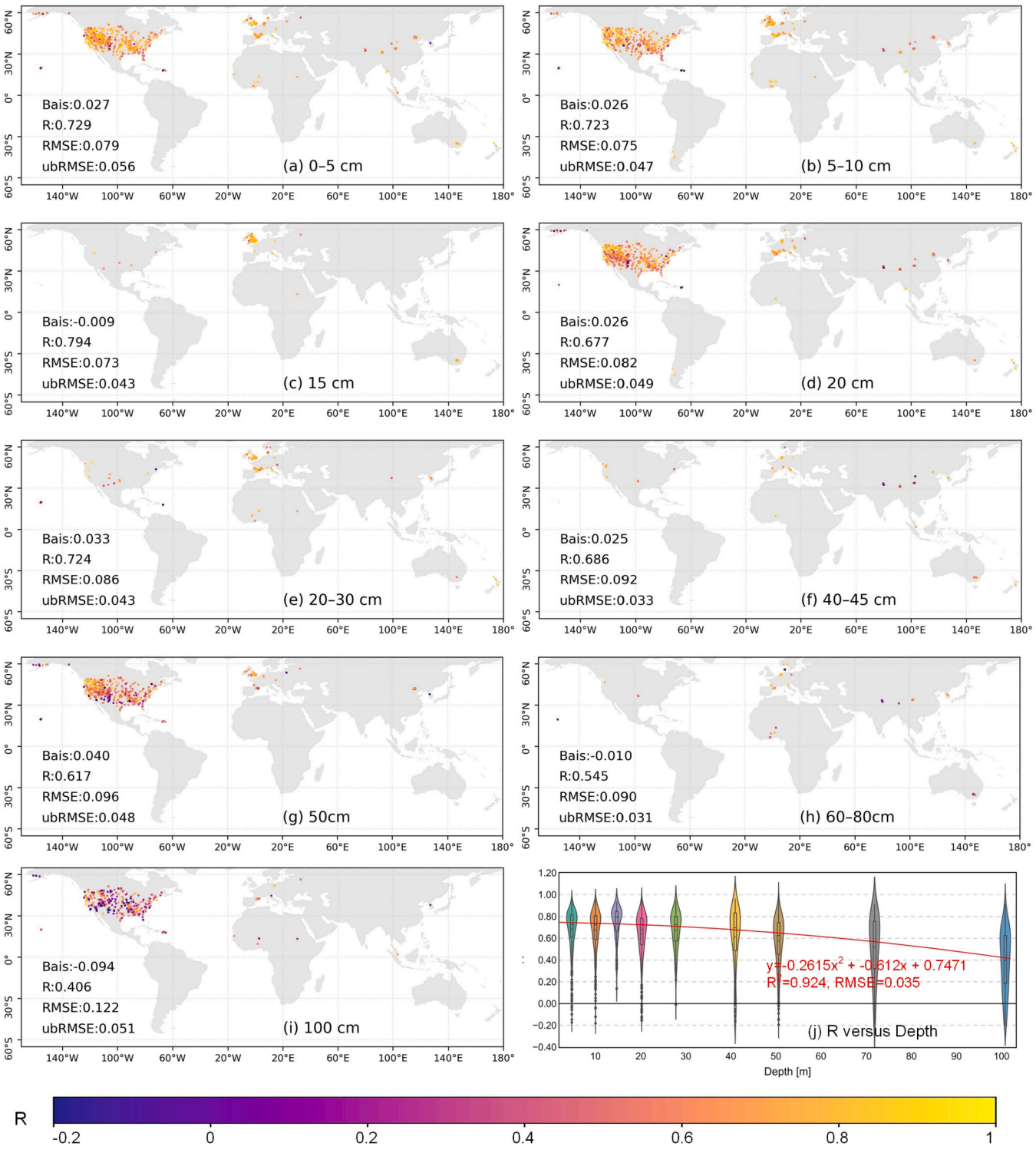

Fig. 9. Correlation coefficient between estimated and in-situ measurements of soil moisture. (a)-(i) are the results for 0–5 cm, 5–10 cm, 15 cm, 20 cm, 20–30 cm, 40–45 cm, 50 cm, 60–80 cm and 100 cm, respectively. The median bias, R, RMSD and ubRMSE of all sites is also provided at the bottom left corner of each figure. (j) is the relationship between R and average depth of each group, with the red line being the fitted line for the median values of each violin plot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

accuracy in those regions (Fig. 10c). The diminished performance in these TC-failed areas is likely attributable to two primary factors: first, the potential unsuitability of applying global median weights to regions with distinct local error structures; and second, the fact that the intrinsic accuracy of the source products themselves was significantly lower in these locations compared to other areas (Fig. A1). To address these limitations, future studies could explore learned weighting strategies that leverage in-situ measurements and transfer learning. However, adopting such approaches would represent a fundamental shift from the unsupervised fusion framework presented in this study toward supervised calibration.

The DeepProfile model was trained using a globally distributed

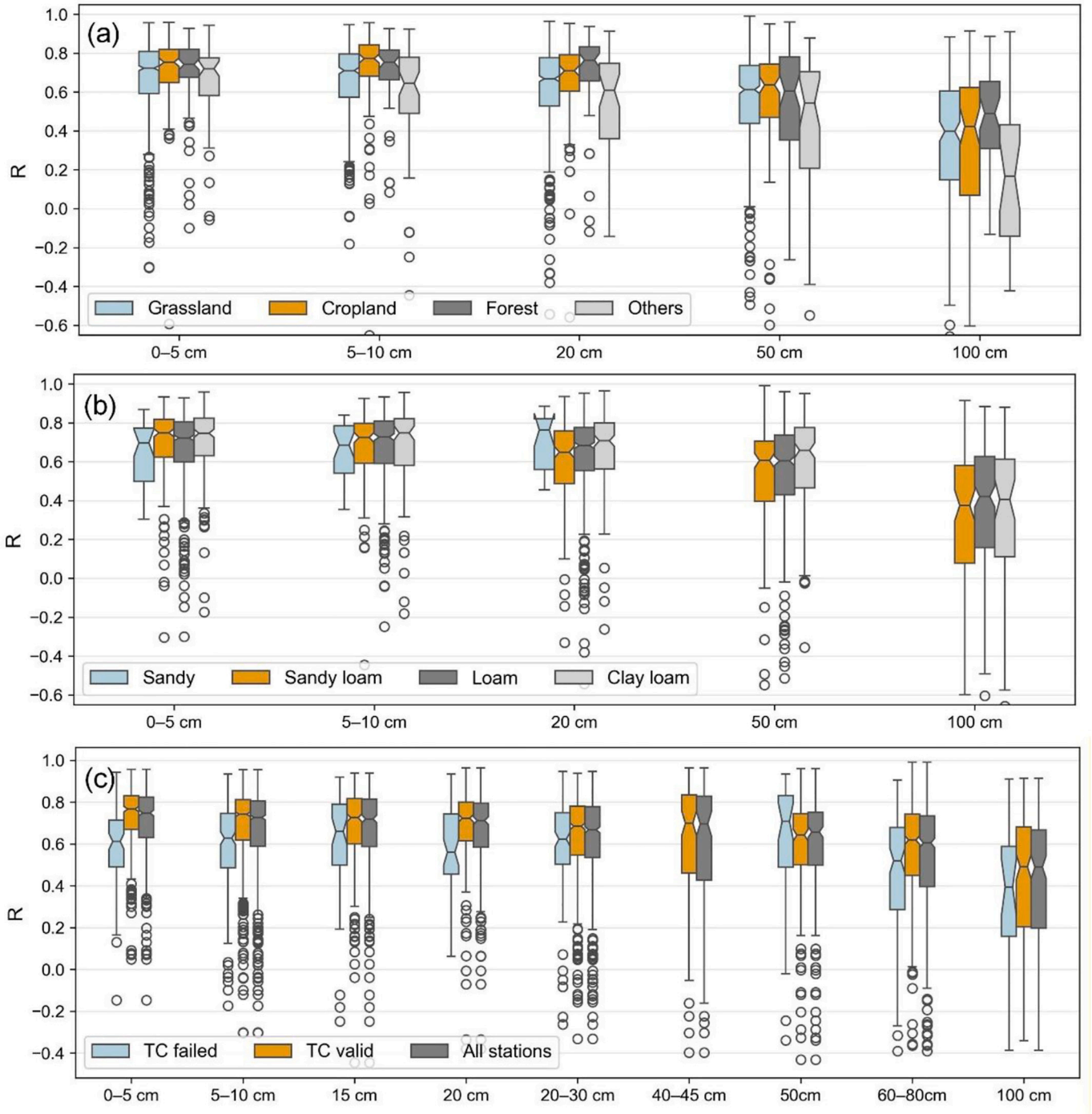

Fig. 10. Performance of the proposed DeepProfile (a) under different landcover types, (b) under different soil types and (c) for grid cells where triple collocation failed, where it was valid, and for all grid cells combined. Depth and/or categories with fewer than 20 grid cells were excluded.

dataset, and analysis of the validation fusion loss revealed that increasing the number of training samples contributed negligibly (Fig. 4b). This suggested that the DeepProfile model had already captured the dominant patterns present in the fusion targets of ERA5, GLDAS, and SMAP L4. However, it is important to note that the validation dataset used to evaluate fusion loss consisted of unseen samples from these same products, rather than independent ground truth observations. Consequently, the fusion loss reflects internal consistency with the source products rather than true physical accuracy. This highlights a critical limitation, that the fusion targets themselves may contain biases or structural uncertainties that were not revealed through product-based validation. It is for this reason that this study focused on establishing the theoretical and operational validity of the inverse fusion framework, rather than releasing a global soil moisture product. The current outputs primarily reflect the consensus of the parent products. To produce a definitive, application-ready global dataset requires the additional integration of in-situ measurements to correct for the inherent biases of these parent models. Consequently, further training or fine-tuning using in-situ soil moisture measurements as the reference truth, particularly those collected from core validation sites with dense and reliable observations, could enhance the physical realism and improve the relevance of the estimated soil moisture profiles for applications. This opens a promising direction for transfer learning, where the pretrained DeepProfile model can be adapted to specific regions or

networks using ground measurements to refine its predictive accuracy (Zhu et al., 2025a).

Following established approaches for root zone soil moisture estimation from remote sensing (Brakhasi et al., 2024; Sadeghi et al., 2016), polynomial functions were adopted in this study to approximate the complex vertical variations of soil moisture profiles. Although the DeepProfile model supports high-order polynomial representations, the maximum degree was limited to 9 to balance expressiveness and computational stability. The effective order of the fitted profiles was further constrained by an order-aware regularization term (λ), which penalizes higher-order coefficients to suppress overfitting and oscillatory artifacts. Empirical analysis showed that increasing the freedom of higher-order coefficients yielded only marginal improvements in validation losses (Fig. 4c and d), indicating limited benefit from higher-order terms. Accordingly, a relatively small regularization coefficient ($\lambda = 0.2 \times 10^{-3}$) was selected to preserve essential nonlinear patterns while maintaining physical plausibility. The resulting soil moisture profiles were typically governed by second- or third-order polynomials, being consistent with remote sensing based studies (Brakhasi et al., 2024; Sadeghi et al., 2016).

Inter-comparison between the estimated soil moisture profiles and the eight layers from SMAP, ERA5, and GLDAS products demonstrates the effectiveness of the fusion strategy in capturing depth-resolved soil moisture dynamics (Fig. 5). The highest agreement was observed for SMAP 0–5 cm, which achieved a strong median correlation ($R = 0.949$) and low median RMSD (0.035 m^3/m^3). This performance is driven by the independent surface loss term (L_{surf}), which explicitly anchors the upper boundary of the estimated profile to the SMAP L4 near-surface soil moisture. By enforcing consistency with this satellite-derived signal, the framework effectively functions as a learning-based "assimilation" system, propagating the high-fidelity remote sensing observation into the deeper root-zone. Notably, the performance over forest was comparable to that of other land cover types, with no statistically significant differences observed (Fig. 10). The model maintained high correlations even in densely vegetated regions, such as parts of Malaysia and Ghana (Fig. 9a). While direct L-band retrievals typically degrade under dense canopy, such improvement suggests that the fusion framework can leverage the complementary stability of the model-based targets (ERA5 and GLDAS) to partially mitigate potential retrieval artifacts while preserving the temporal dynamics of the satellite signal. The SMAP 0–100 cm layer achieved the lowest RMSD overall (0.028 m^3/m^3), benefiting from its integrated depth and higher relative weight. Considering the low RMSD of the SMAP L4, the trained DeepProfile model may be interpreted as a vertical resolution enhancement method for SMAP L4. Beyond merely improving vertical granularity, DeepProfile demonstrated superior performance compared to the three individual source products, their simple arithmetic average, and standard TC-based fusion, across both the near-surface layer and the entire root zone. However, the difference is generally marginally between the proposed method and the TC-based method (~ 0.02 in R). Since the proposed method is a supervised learning method, it's reasonable that the performance is broadly consistent with its "teacher". It is important to acknowledge that the reference "truth" for the 0–100 cm layer was aggregated from discrete measurements at limited depths. The reliability of this evaluation thus is constrained by the uncertainty inherent in estimating a continuous vertical profile from sparse in-situ observations. Similarly, the different depths of the three near-surface soil moisture bring in unknown uncertainty in this comparison. Consequently, this comparison was provided in Appendix (Fig. A1).

Evaluation against in-situ soil moisture measurements confirmed the reliability of the DeepProfile model across a range of depths and geographic regions. Strong agreement was observed in the upper soil layers, particularly within the top 10 cm where median correlation values exceeded 0.72 and median RMSE remained below 0.06 m^3/m^3. This reflects the model's ability to capture surface dynamics, likely aided by the strong constraint from SMAP surface observations. Although no consistent depth-dependent trend in correlation was observed across core grid cells, deeper layers exhibited greater variability and bias, indicating reduced consistency in vertical profile estimation. This pattern became more pronounced with the inclusion of extra grid cells, revealing a clear global decline in correlation with depth. The reduced accuracy at deeper layers is primarily attributed to the coarser vertical resolution of the source products and weaker constraints during training. Incorporating more detailed vertical soil moisture information from advanced data assimilation systems and/or a further transfer learning using in-situ measurements of deep layers may further help to improve deep-layer estimates. Network-specific analysis further showed that depth-dependent bias is a major contributor to error at greater depths (Fig. 7). Moreover, spatial evaluation across both the core and extra grid cells highlighted regional discrepancies, with notably lower accuracy in high-latitude and mountainous regions such as Alaska, the Qinghai-Tibet Plateau, and Hawaii.

The proposed method is a data driven solution and thus does not require the complex scattering modeling and the cumbersome inversion process of remote sensing data. The critical limitation of shallow sensing depth of remote sensing data was mitigated by using the sensed soil moisture as a surface constraint. Apart from the first continuous solution for RZSM, the proposed method provides a valuable opportunity to partly address the two major challenges in direct RZSM inversion from remote sensing data. The estimated soil moisture profiles can serve as *prior* information in the inversion process, either as an initial guess of the profile shape or as bounds on soil moisture at different depths. Such prior knowledge has been widely recognized as essential for addressing the ill-posed nature of near-surface soil moisture (Zhu et al., 2022; Zhu et al., 2023). More importantly, the continuous soil moisture profile enables further investigation into the varying sensing depths of the commonly used C, L and P band radar and radiometer (Shen et al., 2020), allowing a unique opportunity to advance root zone scattering and emission models.

6. Conclusion

This study introduced an inverse fusion strategy for estimating the root zone soil moisture profile by leveraging the complementary information embedded in three widely used products; SMAP L4, ERA5-Land, and GLDAS. By treating these products as learning targets rather than inputs, the proposed DeepProfile model effectively learned spatial-temporal patterns of soil moisture, and produced continuous analytical profiles from 0 to 100 cm. The use of polynomial functions, guided by order-aware regularization, enabled flexible and physically consistent profile estimation across diverse regions and depth systems. Location-specific weights derived from triple collocation allowed the DeepProfile model to adaptively balance the influence of each product, while the inclusion of SMAP surface observations provided a strong constraint on the upper boundary. Evaluation against in-situ measurements confirmed the model's reliability, particularly in the top 50 cm, and revealed that depth-dependent bias and terrain complexity remain key challenges in deeper layers. The inverse fusion strategy and the DeepProfile model not only improved vertical resolution for SMAP L4 but also offers a scalable foundation for future integration of additional products and ground observations through transfer learning. These findings highlight its potential to support more accurate and application-ready soil moisture datasets for hydrological forecasting, agricultural management, and climate adaptation.

CRediT authorship contribution statement

Liujun Zhu: Conceptualization. **Yi Tan:** Conceptualization. **Shanshui Yuan:** Conceptualization. **Junliang Jin:** Conceptualization. **Zhengyang Tang:** Conceptualization. **Jeffrey P. Walker:** Conceptualization.

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 Natural Science Foundation of China (42371369, 52525902, 52279018 and 52121006) and the Basic Research Project of Jiangsu Province (BK20250192). The authors express sincere thanks to the data providers listed in Table 1 and the International Soil Moisture Network for data sharing.

Appendix A. Appendix

The performance of the proposed DeepProfile model was evaluated against five benchmark datasets for both the near-surface and root-zone layers. For the near-surface layer, in-situ measurements at depths ≤5 cm served as the ground truth. The comparison included ERA5 0–7 cm, SMAP 0–5 cm, GLADS 0–10 cm, the simple arithmetic average of these three products, their TC-based fusion, and the DeepProfile estimates for 0–5 cm. For the root zone (0–100 cm), in-situ soil moisture measurements from varying depths were aggregated into a single 0–100 cm value using the midpoint boundary method. This method assigns a representative soil layer thickness to each sensor based on the midpoints between adjacent sensor depths. Specifically, for sensors at depths d_1, d_2, …, d_n, the first sensor represents the layer from 0 cm to $(d_1 + d_2) / 2$. The last sensor represents the layer from $(d_{n\text{-}1} + d_n) / 2$ to 100 cm. Intermediate sensors cover the interval between the midpoints of their respective upper and lower neighbors. The depth-weighted average is thus calculated as

$$\theta_{0-100} = \frac{\sum_{i=1}^{n} (\theta_i \cdot L_i)}{100} \tag{A1}$$

where θ_i is the measured soil moisture at depth i, and L_i is the layer thickness determined by the midpoint method. As shown in Fig. A1, DeepProfile demonstrated superior robustness compared to the baselines. In both the surface and root-zone layers, DeepProfile achieved median R values that matched or exceeded those of the best-performing individual source (SMAP) and the conventional fusion methods (Average and TC).

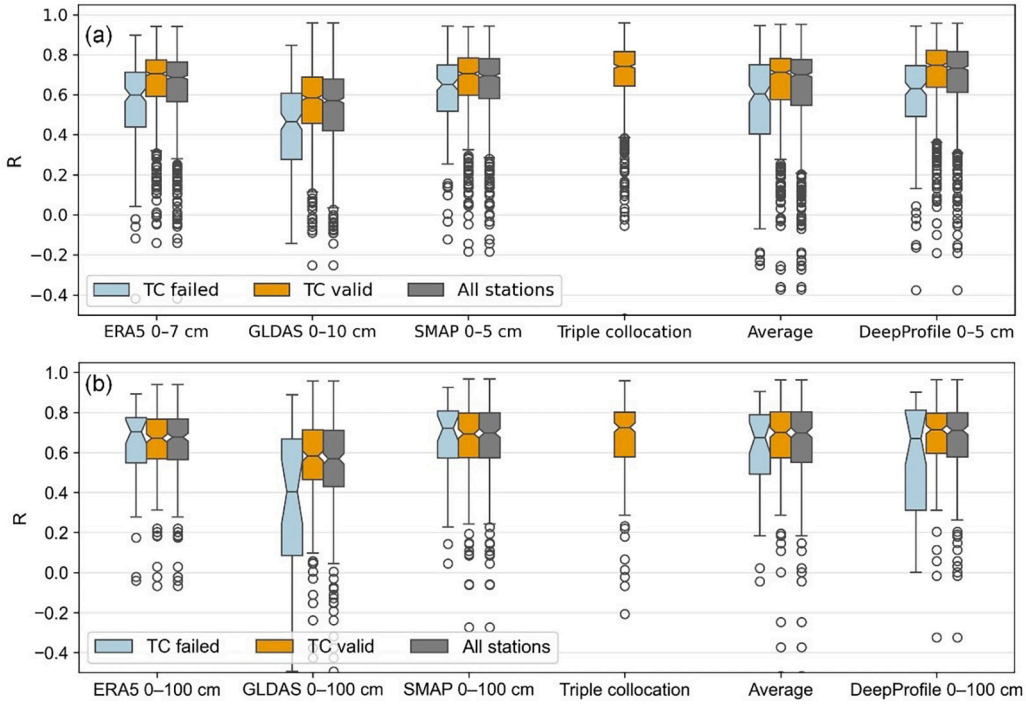

Fig. A1. Performance of the three source products, their average, and the fusion product derived using the triple collocation method at (a) the surface layer and (b) the 0–100 cm depth. In-situ measurements at ≤5 cm were used as the reference in (a), while integrated 0–100 cm in-situ values served as the reference in (b). Correction coefficients (R) were calculated separately for grid cells where triple collocation failed, where it was valid, and for all grid cells combined.

Table A1
The geophysical variables of SPL4SMGP used as the input.

Classification	Geophysical variables	Physical meanings
Meteorological Forcing (11 variables)	surface_pressure	Surface pressure
	height_lowatmmodlay	Height of the lowest atmospheric model layer
	temp_lowatmmodlay	Temperature of the lowest atmospheric model layer
	specific_humidity_lowatmmodlay	Specific humidity of the lowest atmospheric model layer
	windspeed_lowatmmodlay	Wind speed of the lowest atmospheric model layer
	radiation_shortwave_downward_flux	Downward shortwave radiation flux
	radiation_longwave_absorbed_flux	Absorbed longwave radiation flux
	net_downward_shortwave_flux	Net downward shortwave radiation (surface absorbed)
	net_downward_longwave_flux	Net downward longwave radiation (surface absorbed)
	precipitation_total_surface_flux	Total precipitation flux
	snowfall_surface_flux	Snowfall flux
Surface State Variables (17 variables)	sm_profile	Profile soil moisture
	sm_surface_wetness	Surface soil moisture wetness index
	sm_rootzone_wetness	Root zone soil moisture wetness index
	sm_profile_wetness	Profile soil moisture wetness index
	surface_temp	Land surface skin temperature
	soil_temp_layer1	Soil temperature of layer 1
	soil_temp_layer2	Soil temperature of layer 2
	soil_temp_layer3	Soil temperature of layer 3
	soil_temp_layer4	Soil temperature of layer 4
	soil_temp_layer5	Soil temperature of layer 5
	soil_temp_layer6	Soil temperature of layer 6
	snow_mass	Snow water equivalent
	snow_depth	Snow depth
	land_fraction_saturated	Fraction of saturated land
	land_fraction_unsaturated	Fraction of unsaturated land
	land_fraction_wilting	Fraction of land at wilting point
	land_fraction_snow_covered	Fraction of snow covered land
	vegetation_greenness_fraction	Greenness vegetation fraction (GVF)
	leaf_area_index	Leaf area index (LAI)
	mwrtm_vegopacity	Microwave vegetation opacity (τ)
Auxiliary Quantities (11 variables)	land_evapotranspiration_flux	Latent heat flux / Evapotranspiration
	overland_runoff_flux	Overland runoff flux
	baseflow_flux	Baseflow flux (subsurface runoff)
	snow_melt_flux	Snowmelt flux
	soil_water_infiltration_flux	Soil water infiltration flux
	heat_flux_sensible	Sensible heat flux
	heat_flux_latent	Latent heat flux
	heat_flux_ground	Ground heat flux

Data availability

Data will be made available on request.

References

Al-Yaari, A., Dayau, S., Chipeaux, C., Aluome, C., Kruszewski, A., Loustau, D., Wigneron, J.-P., 2018. The AQUI soil moisture network for satellite microwave remote sensing validation in South-Western France. Remote Sens. 10, 1839.

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

Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V.K., Tuller, M., 2021. Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. Remote Sens. Environ. 260, 112434.

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.

Bengough, A., 2012. Water dynamics of the root zone: rhizosphere biophysics and its control on soil hydrology. Vadose Zone J. 11 (vzj2011), 0111.

Benninga, H.-J.F., Carranza, C.D., Pezij, M., van Santen, P., van der Ploeg, M.J., Augustijn, D.C., van der Velde, R., 2018. The Raam regional soil moisture monitoring network in the Netherlands. Earth Syst. Sci. Data 10, 61–79.

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.

Brakhasi, F., Walker, J.P., Judge, J., Liu, P.-W., Shen, X., Ye, N., Wu, X., Yeo, I.-Y., Boopathi, N., Kim, E., 2023. A comparison of passive microwave emission models for estimating brightness temperature at l-and P-bands under bare and vegetated soil conditions. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 17, 2570–2585.

Brakhasi, F., Walker, J.P., Judge, J., Liu, P.-W., Shen, X., Ye, N., Wu, X., Yeo, I.-Y., Kim, E., Kerr, Y., 2024. Soil moisture profile estimation under bare and vegetated soils using combined l-band and p-band radiometer observations: an incoherent modeling approach. Remote Sens. Environ. 307, 114148.

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., 2011. Soil moisture estimation through ASCAT and AMSR-E sensors: an intercomparison and validation study across Europe. Remote Sens. Environ. 115, 3390–3408.

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.

Chai, L., Zhu, Z., Liu, S., Xu, Z., Jin, R., Li, X., Kang, J., Che, T., Zhang, Y., Zhang, J., 2024. QLB-NET: a dense soil moisture and freeze–thaw monitoring network in the qinghai lake basin on the Qinghai–Tibetan Plateau. Bull. Am. Meteorol. Soc. 105, E584–E604.

Colliander, A., Al Jassar, H., Asanuma, J., Berg, A., Bongiovanni, T., Bosch, D., Caldwell, T., Caylor, K., Cosh, M., Dorigo, W., 2017. SMAP/In Situ Core Validation Site Land Surface Parameters Match-up Data, Version 1. NASA National Snow and Ice Data Center (Distributed Active Archive Center (DAAC) data set, DXAVIXLY18KM).

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

Cooper, H.M., Bennett, E., Blake, J., Blyth, E., Boorman, D., Cooper, E., Evans, J., Fry, M., Jenkins, A., Morrison, R., 2021. COSMOS-UK: national soil moisture and hydrometeorology data for environmental science research. Earth Syst. Sci. Data 13, 1737–1757.

Cosh, M.H., Jackson, T.J., Starks, P., Heathman, G., 2006. Temporal stability of surface soil moisture in the little Washita river watershed and its applications in satellite soil moisture product validation. J. Hydrol. 323, 168–177.

Das, N.N., Mohanty, B.P., 2006. Root zone soil moisture assessment using remote sensing and vadose zone modeling. Vadose Zone J. 5, 296–307.
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. Discuss. 1–83.
Fan, L., Xing, Z., De Lannoy, G., Frappart, F., Peng, J., Zeng, J., Li, X., Yang, K., Zhao, T., Shi, J., 2022. Evaluation of satellite and reanalysis estimates of surface and root-zone soil moisture in croplands of Jiangsu province, China. Remote Sens. Environ. 282, 113283.
Friedl, M., Sulla-Menashe, D., 2015. MCD12C1 MODIS/Terra+ Aqua Land Cover Type Yearly L3 Global 0.05 Deg CMG V006. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MCD12C1.006.
Fuchsberger, J., Kirchengast, G., Kabas, T., 2021. WegenerNet high-resolution weather and climate data from 2007 to 2020. Earth Syst. Sci. Data 13, 1307–1334.
Fuentes, I., Padarian, J., Vervoort, R.W., 2022. Towards near real-time national-scale soil water content monitoring using data fusion as a downscaling alternative. J. Hydrol. 609, 127705.
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.
Hajdu, I., Yule, I., White, M., 2022. The Patitapu soil moisture network (PTSMN) dataset and its deployment in new Zealands hill country. Agric. Water Manag. 274, 107915.
Hamman, J.J., Nijssen, B., Bohn, T.J., Gergel, D.R., Mao, Y., 2018. The variable infiltration capacity model version 5 (VIC-5): infrastructure improvements for new applications and reproducibility. Geosci. Model Dev. 11, 3481–3496.
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049.
Hirschi, M., Mueller, B., Dorigo, W., Seneviratne, S.I., 2014. Using remotely sensed soil moisture for land–atmosphere coupling diagnostics: the role of surface vs. root-zone soil moisture variability. Remote Sens. Environ. 154, 246–252.
Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J.E., Weinberger, K.Q., 2017a. Snapshot ensembles: train 1, get m for free. arXiv 1–14 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 Recognition. IEEE 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.
Jiang, L., Wang, J., Cui, H., Wang, G., Zhao, T., Zhao, S., Chai, L., Liu, X., Yang, J., 2020. In situ soil moisture and temperature network in Genhe watershed and Saihanba area in China. Data Brief 31, 105693.
Kang, C.S., Kanniah, K.D., Kerr, Y.H., Cracknell, A.P., 2016. Analysis of in-situ soil moisture data and validation of SMOS soil moisture products at selected agricultural sites over a tropical region. Int. J. Remote Sens. 37, 3636–3654.
Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv 1–15 preprint arXiv: 1412.6980.
Li, B., Wang, C., Gu, X., Zhou, X., Ma, M., Li, L., Feng, Z., Ding, T., Li, X., Jiang, T., 2022. Accuracy calibration and evaluation of capacitance-based soil moisture sensors for a variety of soil properties. Agric. Water Manag. 273, 107913.
Li, M., Lang, R., Cosh, M., 2024. P-band and l-band radiometry retrieval of soil moisture and temperature profiles. IEEE Trans. Geosci. Remote Sens. 62, 1–15.
Liu, E., Zhu, Y., Calvet, J.-C., Lü, H., Bonan, B., Zheng, J., Gou, Q., Wang, X., Ding, Z., Xu, H., 2024. Evaluation of root zone soil moisture products over the Huai river basin. Hydrol. Earth Syst. Sci. 28, 2375–2400.
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.
Nguyen, H.H., Kim, H., Choi, M., 2017. Evaluation of the soil water content using cosmic-ray neutron probe in a heterogeneous monsoon climate-dominated region. Adv. Water Resour. 108, 125–138.
Njoku, E.G., Kong, J.A., 1977. Theory for passive microwave remote sensing of near-surface soil moisture. J. Geophys. Res. 82, 3108–3118.
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.
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.
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. 252, 112162.
Peng, C., Zeng, J., Chen, K.-S., Ma, H., Letu, H., Zhang, X., Shi, P., Bi, H., 2024. Spatial representativeness of soil moisture stations and its influential factors at a global scale. IEEE Trans. Geosci. Remote Sens. 63, 4402915.
Reichle, R.H., Liu, Q., Koster, R.D., Crow, W.T., De Lannoy, G.J., Kimball, J.S., Ardizzone, J.V., Bosch, D., Colliander, A., Cosh, M., 2019. Version 4 of the SMAP level-4 soil moisture algorithm and data product. J. Adv. Model. Earth Syst. 11, 3106–3130.
Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., 2004. The global land data assimilation system. Bull. Am. Meteorol. Soc. 85, 381–394.
Sadeghi, M., Tabatabaeenejad, A., Tuller, M., Moghaddam, M., Jones, S.B., 2016. Advancing NASAs AirMOSS p-band radar root zone soil moisture retrieval algorithm via incorporation of Richards equation. Remote Sens. 9, 17.
Sahaar, S.A., Niemann, J.D., 2024. Estimating rootzone soil moisture by fusing multiple remote sensing products with machine learning. Remote Sens. 16, 3699.
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).
Shen, X., Walker, J.P., Ye, N., Wu, X., Boopathi, N., Yeo, I.-Y., Zhang, L., Zhu, L., 2020. Soil moisture retrieval depth of p-and L-band radiometry: predictions and observations. IEEE Trans. Geosci. Remote Sens. 59 (8), 6814–6822.
Shen, X., Fan, L., Zuo, T., Cui, T., Wu, J., Ye, N., Brakhasi, F., Wu, X., Zhu, L., Wigneron, J.-P., 2024. P-band radiometry for enhanced vegetation optical depth (VOD) and soil moisture retrieval in dense crop canopies. Remote Sens. Environ. 313, 114353.
Shrestha, P., Kurtz, W., Vogel, G., Schulz, J.P., Sulis, M., Hendricks Franssen, H.J., Kollet, S., Simmer, C., 2018. Connection between root zone soil moisture and surface energy flux partitioning using modeling, observations, and data assimilation for a temperate grassland site in Germany. J. Geophys. Res. Biogeosci. 123, 2839–2862.
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.
Stocker, B.D., Tumber-Dávila, S.J., Konings, A.G., Anderson, M.C., Hain, C., Jackson, R. B., 2023. Global patterns of water storage in the rooting zones of vegetation. Nat. Geosci. 16, 250–256.
Stoffelen, A., 1998. Toward the true near-surface wind speed: error modeling and calibration using triple collocation. J. Geophys. Res. Oceans 103, 7755–7766.
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.
Tabatabaeenejad, A., Burgin, M., Duan, X., Moghaddam, M., 2015. P-band radar retrieval of subsurface soil moisture profile as a second-order polynomial: first AirMOSS results. IEEE Trans. Geosci. Remote Sens. 53, 645–658.
Tian, J., Zhang, Y., 2025. Error estimation and data fusion of root zone soil moisture products over China based on the three corned hat method. Glob. Planet. Chang. 251, 104797.
van der Velde, R., Benninga, H.-J.F., Retsios, B., Vermunt, P.C., Salama, M.S., 2023. Twelve years of profile soil moisture and temperature measurements in Twente, the Netherlands. Earth Syst. Sci. Data 15, 1889–1910.
Vereecken, H., Huisman, J., Bogena, H., Vanderborght, J., Vrugt, J., Hopmans, J., 2008. On the value of soil moisture measurements in vadose zone hydrology: a review. Water Resour. Res. 44.
Xu, L., Chen, N., Zhang, X., Moradkhani, H., Zhang, C., Hu, C., 2021. In-situ and triple-collocation based evaluations of eight global root zone soil moisture products. Remote Sens. Environ. 254, 112248.
Yang, K., 2013. A multi-scale soil moisture and freeze-thaw monitoring network on the Tibetan plateau and its applications. In: AGU Fall Meeting Abstracts. H31F-1237.
Zhang, C., Long, D., 2021. Estimating spatially explicit irrigation water use based on remotely sensed evapotranspiration and modeled root zone soil moisture. Water Resour. Res. 57, e2021WR031382.
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.
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.
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., Cai, Q., Jin, J., Yuan, S., Shen, X., Walker, J.P., 2025a. Multi-scale domain adaptation for high-resolution soil moisture retrieval from synthetic aperture radar in data-scarce regions. J. Hydrol. 657, 133073.
Zhu, L., Dai, J., Jin, J., Yuan, S., Xiong, Z., Walker, J.P., 2025b. Are the current expectations for SAR remote sensing of soil moisture using machine learning over-optimistic? IEEE Trans. Geosci. Remote Sens. 63, 4501815.
Zreda, M., Desilets, D., Ferré, T., Scott, R.L., 2008. Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. Geophys. Res. Lett. 35.