A seamless global daily 5 km soil moisture product from 1982 to 2021 using AVHRR satellite data and an attention-based deep learning model

Yufang Zhang[1], Shunlin Liang[2], Han Ma[2], Tao He[3], Feng Tian[3], Guodong Zhang[4], Jianglei Xu[2]

[1]School of Software, Northwestern Polytechnical University, Xi'an, 710072, China
[2]Department of Geography, University of Hong Kong, Hong Kong, 999077, China
[3]Hubei Key Laboratory of Quantitative Remote Sensing of Land and Atmosphere, School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430079, China
[4]Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, 610031, China

Correspondence to: Shunlin Liang (shunlin@hku.hk)

Abstract. Soil moisture (SM) data records longer than 30 years are critical for climate change research and various applications. However, only a few such long-term global SM datasets exist, and they often suffer from large biases, low spatial resolution, or spatiotemporal incompleteness. Here, we generated a consistent and seamless global surface SM product (0–5 cm) spanning 1982 to 2021 using a deep learning (DL) model. The model was trained with the GLASS-MODIS SM product and was designed to integrate four decades of Advanced Very High Resolution Radiometer (AVHRR)-derived albedo and land surface temperature, ERA5-Land SM, as well as terrain and soil texture datasets as input features. Considering the temporal autocorrelation of SM, we explored two types of DL models that are adept at processing sequential data, including three long short-term memory (LSTM)-based models, i.e., the basic LSTM, Bidirectional LSTM (Bi-LSTM), and Attention-based LSTM (AtLSTM), as well as a Transformer model. We also compared the performance of the DL models with the tree-based eXtreme Gradient Boosting (XGBoost) model, known for its high efficiency and accuracy. Our results show that all four DL models outperformed the benchmark XGBoost model, with the AtLSTM model achieving the highest accuracy on the test set, particularly at high SM levels (> 0.4 m^3 m^{-3}). These results suggest that under some challenging conditions, utilizing temporal information as well as adding an attention module can effectively enhance the estimation accuracy of SM. Subsequent analysis of attention weights revealed that the AtLSTM model could automatically learn the necessary temporal information from adjacent positions in the sequence, which is critical for accurate SM estimation. The best-performing AtLSTM model was then adopted to produce a four-decade seamless global SM dataset at 5 km spatial resolution, denoted as the GLASS-AVHRR SM product. Validation of the GLASS-AVHRR SM product using 45 independent International Soil Moisture Network (ISMN) stations prior to 2000 yielded a median correlation coefficient (R) of 0.73 and unbiased RMSE (ubRMSE) of 0.041 m^3 m^{-3}. When validated against SM datasets from three post-2000 field-scale COsmic-ray Soil Moisture Observing System (COSMOS) networks, the median R values ranged from 0.63 to 0.79, and the median ubRMSE values ranged from 0.044 to 0.065 m^3 m^{-3}. Further validation across 22 upscaled 9 km Soil Moisture Active Passive (SMAP) core validation sites indicated that it could well capture the temporal variations in

measured SM and remained unaffected by the large wet biases present in the input ERA5-Land SM product. Moreover, characterized by complete spatial coverage and low biases, this four-decade, 5 km GLASS-AVHRR SM product exhibited high spatial and temporal consistency with the 1 km GLASS-MODIS SM product, and contained much richer spatial details than both the long-term ERA5-Land SM product (0.1°) and European Space Agency Climate Change Initiative combined SM product (0.25°). The annual average GLASS-AVHRR SM dataset from 1982 to 2021 is available at https://doi.org/10.5281/zenodo.14198201 (Zhang et al., 2024b), and the complete product can be freely downloaded from https://glass.hku.hk/archive/SM/AVHRR/.

1 Introduction

Soil moisture (SM) is an essential climate-sensitive variable that exhibits high spatial and temporal variability. It can be measured directly by in situ sensors or indirectly through model simulations or remote sensing techniques (Liang and Wang, 2020). Accurate knowledge of the spatial and temporal distribution of SM can benefit applications across various Earth system domains, including climate, hydrology, and agriculture (Dorigo et al., 2017; Peng et al., 2021a). While local- to regional-scale hydrological and agricultural applications like watershed runoff modeling, evapotranspiration estimation, and crop yield prediction demand SM products with high spatial resolution (≤ 1 km) (Hssaine et al., 2018; Schoener and Stone, 2019; Zhuo et al., 2019), continental- to global-scale climate change-related applications, such as SM trend analyses and drought monitoring, generally require long-term data availability (> 30 years), in addition to moderate spatial resolution and high accuracy (Cheng et al., 2015; Grillakis, 2019).

Long-term point-scale SM can be measured directly by in situ sensors, thus great efforts have been devoted worldwide to deploying and maintaining a series of operational SM networks. In situ SM datasets from some networks were shared by data organizations, which were then processed and released in a harmonized format to the public by the International Soil Moisture Network (ISMN) data repository (Dorigo et al., 2021). Still, these networks are too sparse, unevenly distributed in space, and each covers a different observation period, hindering their use in large-scale applications. Currently, large-scale SM products are typically obtained through model simulations or remote sensing techniques. Driven by long-term forcing variables, land surface models or data assimilation systems can simulate decades of spatiotemporally continuous SM products at the global scale, with an increasingly finer spatial resolution. Several commonly used SM products include those generated by the Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) at 0.5° from 1980 to the present (Gelaro et al., 2017), the Global Land Data Assimilation System version 2 (GLDAS-2) at 1°/0.25° from 1948 to the present (Rodell et al., 2004), and the land component of the fifth generation of European ReAnalysis (ERA5-Land) at 0.1° from 1950 to the present (Muñoz-Sabater et al., 2021). Recently, models that focus on the dynamic simulation of evapotranspiration and SM, such as the fourth generation of the Global Land Evaporation Amsterdam Model (GLEAM4) (0.1°, 1980–2023) and the Simple Terrestrial Hydrosphere model, version 2 (SiTHv2) (0.1°, 1982–2020), have also provided long-term global SM products by integrating multi-source satellite data and hydrometeorological variables (Miralles et al.,

65 2025; Zhang et al., 2024a). Yet, these SM products may suffer from large uncertainties arising from defective forcing data, imperfect model parameterization, and the uneven spatial distribution of input meteorological observations, particularly the limited observational coverage in tropical regions. (Ling et al., 2021; Zhang et al., 2024a).

Alternatively, microwave remote sensing techniques have been utilized for SM retrieval since the 1970s (Schmugge et al., 1974). Various global SM products have been developed from a range of active or passive microwave sensors, such as the

70 advanced scatterometer aboard the Meteorological Operational Satellites (Bartalis et al., 2007), the microwave radiation imager on Fengyun-3 satellites (Kang et al., 2021), and the L-band radiometers on the Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP) satellites (Chan et al., 2018; Entekhabi et al., 2010; Kerr et al., 2012; Li et al., 2022c, b; Wigneron et al., 2021). However, the temporal coverage of these single-sensor SM products is typically short, as constrained by the operational lifespan of the satellites. In this context, the European Space Agency (ESA) Climate

75 Change Initiative (CCI) program released a long-term global SM product spanning the period since 1978, which merged multiple active and passive microwave SM products retrieved from different satellite instruments (Dorigo et al., 2017). Despite being the longest satellite SM dataset currently available, the ESA CCI combined SM product has a relatively low spatial resolution (0.25°) and incomplete spatial coverage, which may restrict its usage in certain applications. According to Zheng et al. (2023), the percentage of missing data in the ESA CCI combined SM product ranges from 21.8% to 94.41% at

80 the daily scale during the period from 2000 to 2020.

In contrast, optical and thermal remote sensing techniques are characterized by long observation period, rich spectral bands, and high spatial resolution, but their relatively low sensitivity to SM poses challenges in deriving long-term global SM product solely from optical and thermal satellite observations. Over the past few decades, optical and thermal datasets have been extensively employed to downscale the coarse-scale microwave or model-simulated SM products. Most of these

85 downscaling studies empirically or physically relate vegetation and temperature parameters to SM conditions based on the universal triangle concept (Gillies and Carlson, 1995; Merlin et al., 2012; Piles et al., 2011). A detailed review of the strengths and limitations of various SM downscaling algorithms can refer to Sabaghy et al. (2018). In recent years, machine learning models have gradually gained popularity in SM estimation and the downscaling of coarse-scale SM products, such as the SMAP, ERA5-Land, and ESA CCI SM products (Cheng et al., 2023; Guevara et al., 2021; Karthikeyan and Mishra,

90 2021; Zhang et al., 2023; Zheng et al., 2023), due to their flexibility to integrate multi-source datasets and ability to implicitly learn the non-linear relationships between SM and its influencing factors. However, the above-mentioned downscaling studies primarily concentrated on enhancing the spatial resolution of SM products, typically through integrating the fine-scale Moderate Resolution Imaging Spectroradiometer (MODIS) datasets, and there is still a lack of focus on developing long-term SM products or utilizing the four-decades Advanced Very High Resolution Radiometer (AVHRR)

95 observations for long-term SM estimation.

Compared with conventional machine learning models, deep learning (DL) models can automatically extract relevant features from raw datasets and learn complex non-linear relationships between variables, without the need for careful feature engineering (LeCun et al., 2015). Recently, significant progress has been made in applying DL techniques to a range of

environmental remote sensing research areas, including land cover mapping (Huang et al., 2018), data fusion and downscaling (Wang et al., 2021), and environmental parameter retrieval (Ma and Liang, 2022; Yuan et al., 2020). In terms of SM retrieval, Fang et al. (2017) first utilized a long short-term memory (LSTM) model to predict spatiotemporally continuous SM over the Continental U.S., with atmospheric forcings, modeled SM, and static attributes employed as input features, and the SMAP SM product serving as the training target. Since then, various DL models have been used in SM estimating (Gao et al., 2022; Sungmin and Orth, 2021), downscaling (Xu et al., 2022; Zhao et al., 2022), forecasting (Fang and Shen, 2020; Li et al., 2022a), and gap-filling researches (Zhang et al., 2022; Zhou et al., 2023). Among them, the most frequently used DL models were the LSTM-based models designed to capture temporal information from sequential data and the convolutional neural network (CNN) based models constructed to extract spatial patterns from grid data, alongside several other models such as the deep neural network and deep belief network. In those studies, input features might include brightness temperature, surface reflectance, meteorological forcings, terrain and soil properties, land cover, precipitation, and land surface temperature (LST), depending on the types of models they aimed to simulate, such as radiative transfer models, downscaling models, or land surface models, while the training target varied from point-scale in situ SM to coarse-scale microwave or simulated SM. Despite the diversity of data sources, research areas, and neural networks, all of those DL models achieved satisfactory performance, demonstrating their good fitting and generalization capabilities, as well as great potential for generating global SM products. Validation of those DL models against the ISMN in situ SM dataset showed that the average correlation coefficient (R) ranged from 0.672 to 0.715, and the unbiased root mean square error (ubRMSE) ranged from 0.041 to 0.061 $m^3 m^{-3}$ (Gao et al., 2022; Xu et al., 2022; Zhang et al., 2022). Nevertheless, there is still a lack of research that utilizes DL models to generate long-term global SM data records, as evident from Table 1. Besides, while Transformer has demonstrated effectiveness in domains like runoff modeling, drought forecasting, and crop mapping (Amanambu et al., 2022; Xu et al., 2020; Yin et al., 2022), its application in SM estimation remains scarce.

Table 1 Main characteristics of currently available long-term (> 30 years) global SM products.

Category	SM products	Spatial resolution	Temporal coverage	Spatial integrity	References
Microwave	ESA CCI	0.25°	1978–2022	incomplete	Dorigo et al. (2017)
Reanalysis	GLDAS-2	1°/0.25°	1948–present	Seamless	Rodell et al. (2004)
	MERRA-2	0.5°	1980–present	Seamless	Gelaro et al. (2017)
	ERA5-Land	0.1°	1950–present	Seamless	Muñoz-Sabater et al. (2021)
Model-simulated	GLEAM4	0.1°	1980–2023	Seamless	Miralles et al. (2025)
	SiTHv2	0.1°	1982–2020	Seamless	Zhang et al. (2024a)
DL-based	GLASS-AVHRR	5 km	1982–2021	Seamless	This study

In this context, we aim to develop a long-term global SM estimation framework based on DL using mainly the long-archived AVHRR satellite observations. Specifically, the AVHRR albedo and LST products from the Global LAnd Surface Satellite (GLASS) product suite, ERA5-Land reanalysis SM product, as well as auxiliary terrain and soil texture datasets are used as

inputs, and the global 1 km GLASS-MODIS SM product (2000–2020) generated by Zhang et al. (2023) is used as the target to train different types of DL models. In particular, three LSTM-based models, i.e., the basic LSTM, Bidirectional LSTM (Bi-LSTM), and Attention-based LSTM (AtLSTM), along with a Transformer model, all of which are adept at processing sequential data, are explored. Then the best-performing model is employed to generate a four-decade (1982–2021) spatiotemporally continuous global surface SM dataset (0–5 cm) at 5 km resolution, denoted as the GLASS-AVHRR SM product. The specific objectives of this study are:

(1) To develop a DL-based global SM estimation model by integrating multi-source datasets and leveraging their complementary strengths, so as to derive a seamless and reliable long-term global SM product;

(2) To compare the performance of different DL models: the basic LSTM, Bi-LSTM, AtLSTM, and Transformer, with the benchmark XGBoost model, and to investigate the effect of input sequence length on model accuracy;

(3) To fully evaluate the accuracy and spatiotemporal consistency of the derived long-term GLASS-AVHRR SM product through validation against in situ SM datasets across different spatial scales and intercomparison with other long-term global SM products.

2 Datasets

The multi-source datasets used in this study to develop the long-term SM estimation model are summarized in Table 2. The input variables were extracted from the GLASS-AVHRR albedo and LST products, the ERA5-Land reanalysis SM product, the Multi-Error-Removed Improved-Terrain (MERIT) DEM, and the SoilGrids datasets, respectively, while the target variable was obtained from the GLASS-MODIS SM product. These input features are widely used in ML and DL-based SM estimation studies. This section also introduces the ISMN, COsmic-ray Soil Moisture Observing System (COSMOS), and SMAP Core Validation Sites (CVSs) in situ SM datasets used for validation, alongside the long-term ESA CCI SM product used for intercomparison.

Table 2 Summary of the multi-source datasets used to develop the long-term SM product.

Dataset	Variable	Temporal resolution	Spatial resolution	Usage	References
GLASS-AVHRR	Albedo	8 d	5 km	input	Qu et al. (2014);Liu et al. (2013)
	LST	daily	5 km	input	Jia (2023)
ERA5-Land	SM	hourly	0.1°	input	Muñoz-Sabater et al. (2021)
MERIT DEM	elevation, slope, aspect	-	90 m	input	Yamazaki et al. (2017)
SoilGrids	clay, sand, silt	-	250 m	input	Poggio et al. (2021)
GLASS-MODIS	SM	daily	1 km	target	Zhang et al. (2023)

2.1 GLASS-AVHRR albedo and LST products

As part of the GLASS product suite, the GLASS-AVHRR albedo and LST products are generated mainly from the long-archived AVHRR satellite observations dating back to the 1980s and are characterized by long-term temporal coverage, spatial continuity, and high accuracy (Liang et al., 2021). In particular, the GLASS-AVHRR albedo product was retrieved through a direct estimation algorithm (Qu et al., 2014) and a spatiotemporal filtering algorithm (Liu et al., 2013). The latest version (V5) of the GLASS-AVHRR albedo product at 5 km spatial resolution can be downloaded from http://www.glass.umd.edu/Albedo/MIX/. Here, the black-sky visible, near-infrared, and shortwave albedo were extracted and used as input variables, with the original 8-day temporal resolution interpolated to daily using linear interpolation to align with the training target. Meanwhile, the global all-sky GLASS-AVHRR LST product was estimated using a surface energy balance-based algorithm (Jia, 2023; Jia et al., 2024), which will be released soon. The daily mean LST at 5 km resolution was also used here as an input variable.

2.2 ERA5-Land SM product

The ERA5-Land is a state-of-the-art long-term reanalysis dataset that includes multiple variables related to water and energy cycles spanning from 1950 to the present (Muñoz-Sabater et al., 2021). It offers seamless global coverage with an hourly temporal resolution and 0.1° spatial resolution. Previous validation studies show that although it typically exhibited high temporal correlations with in situ SM datasets, it often suffered from large biases (Gao et al., 2022; Xing et al., 2023; Zheng et al., 2022). Here, the first-layer (0–7 cm) ERA5-Land SM product was downloaded from https://cds.climate.copernicus.eu/. The daily mean SM was then calculated and up-sampled to 5 km through bilinear interpolation before being used as an input variable for the model to provide SM background information. Moreover, the ERA5-Land SM product was also validated against in situ SM datasets and intercompared with the generated GLASS-AVHRR SM product.

2.3 Terrain and soil texture datasets

Topography and soil properties are the main factors that affect the spatial distribution of SM at fine scales. Here, we used the MERIT DEM (http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/), a high accuracy DEM generated by integrating multiple spaceborne DEMs (Yamazaki et al., 2017). This dataset covers 90°N–60°S over land at a resolution of 90 m and shows significant improvement in flat regions compared to previous spaceborne DEMs. After downloading the MERIT DEM, it was then used to derive elevation, slope, and aspect. Meanwhile, we also used the 250-m SoilGrids product (https://www.isric.org/explore/soilgrids), a high-resolution soil property dataset generated from global soil profiles and environmental variables using machine learning models (Poggio et al., 2021). Specifically, the mean sand, silt, and clay content of the top soil layer (0–5 cm) were extracted from the SoilGrids product. All of these terrain and soil texture variables were resampled to 5 km before being used as inputs to the SM estimation model.

2.4 GLASS-MODIS SM product

The training target used in this study was the global 1 km spatiotemporally continuous GLASS-MODIS surface SM product (0–5 cm), which was generated using an XGBoost machine learning model that integrated the GLASS-MODIS albedo, LST, and leaf area index (LAI) products with multi-source datasets. In situ SM from the representative ISMN stations distributed globally was utilized by the XGBoost model as training target (Zhang et al., 2023). This product exhibits high spatial and temporal consistency with both the ESA CCI and SMAP/Sentinel-1 L2 Radiometer/Radar SM products, while maintaining a more complete spatial coverage. The daily GLASS-MODIS SM product from 2000 to 2020 is freely available at https://glass.hku.hk/archive/SM/MODIS/. Here, we derived training samples from the 5 km resampled GLASS-MODIS SM product rather than directly using in situ SM as training target, as the global SM product could provide a much richer and representative training set than the sparse ISMN SM dataset.

2.5 In situ SM datasets

After generating the GLASS-AVHRR SM product using the developed DL model, three types of in situ SM datasets at different spatial scales were adopted to evaluate its accuracy and consistency. The characteristics of these in situ SM datasets are listed in Table 3, and the spatial distribution of the corresponding SM stations is shown in Fig. A1. The first type is the point-scale ISMN SM dataset (Dorigo et al., 2021), providing a valuable reference for validating gridded SM products, despite the relatively poor spatial representativeness of some SM stations. There were 1672 ISMN stations available for validation during Period I (2000–2018). Among them, 715 spatially representative stations were selected using the triple collocation method, as described in detail in Zhang et al. (2023). Although SM datasets from these representative stations were previously used as target to train the GLASS-MODIS SM estimation model, making them only partially independent, they can be used here to assess the consistency between the GLASS-AVHRR and GLASS-MODIS SM products. Moreover, the 45 fully independent ISMN stations from Period II (1982–1999) can be used to evaluate the accuracy of the GLASS-AVHRR SM product during the earlier years. The daily mean SM was calculated by averaging the hourly SM measurements at the top soil layer (0–5 cm) obtained from https://ismn.earth/, considering only those flagged as "G" for good quality.

The second type is the COsmic-ray Soil Moisture Observing System (COSMOS) SM dataset, which includes area-averaged SM measurements at the field scale from three COSMOS networks: COSMOS (Zreda et al., 2012), COSMOS-UK (Cooper et al., 2021), and COSMOS-Europe (Bogena et al., 2022). The COSMOS sensors detect low-energy cosmic-ray neutrons above the ground, which can be converted to SM within a footprint radius of 130 to 240 m and a penetration depth of up to 83 cm, depending on factors such as air humidity, SM, and vegetation (Köhli et al., 2015). Although data from the COSMOS and COSMOS-UK networks had been integrated into the ISMN database, they were excluded from the training dataset of the GLASS-MODIS SM estimation model because their observation depths exceeded the 5 cm threshold. Recently, data from the COSMOS-Europe network have been released and can be accessed at https://doi.org/10.34731/x9s3-kr48. Collectively, these post-2000 SM datasets can serve as an independent source for validating the GLASS-AVHRR SM

product at an intermediate scale. After filtering based on the quality flags and aligning with the GLASS-AVHRR SM product, there were 102 COSMOS, 45 COSMOS-UK, and 51 COSMOS-Europe stations available for validation. The distribution of sensing depths for each station across the three COSMOS networks is presented in Fig. A2. While COSMOS sensors measure SM at relatively deeper layers, they have been used to validate microwave and modelled surface SM products and show good correlations with them (Montzka et al., 2017; Peng et al., 2021b).

The third type is the SMAP/In situ core validation site (CVS) match-up dataset, which contains the up-scaled in situ SM measurements derived from multiple quality-controlled stations that have been aligned with SMAP SM products (Colliander et al., 2017). A total of 22 globally distributed CVSs were matched with the SMAP-Sentinel L2 SM product gridded at 9 km resolution (SMAPL2SMSP9km). This independent 9 km SMAP CVS in situ dataset can be used to validate the GLASS-AVHRR SM product with reduced impact of scale difference. It covers the period from 2015 to the present and can be downloaded from https://nsidc.org/data/nsidc-0712/versions/1.

Table 3 Characteristics of three types of in situ SM datasets used in this study at different spatial scales.

Dataset	Group of stations	No. of stations	Spatial scale	Sensing depth	Time period	References
ISMN	All ISMN (Period I)	1672	Point-scale	0–5 cm	2000–2018	Dorigo et al. (2021)
	Representative ISMN (Period I)	715			2000–2018	
	ISMN (Period II)	45			1982–1999	
COSMOS	COSMOS	102	130–240 m	15–83 cm	2008–2018	Zreda et al. (2012)
	COSMOS-UK	45			2013–2018	Cooper et al. (2021)
	COSMOS-Europe	51			2011–2018	Bogena et al. (2022)
CVS	SMAP CVS	22	9 km	0–5 cm	2015–2021	Colliander et al. (2017)

2.6 ESA CCI SM product

The European Space Agency (ESA) launched the Climate Change Initiative (CCI) SM project to develop the ESA CCI SM dataset, a global daily multi-decadal dataset aimed at supporting climate research (Dorigo et al., 2017). This dataset merged multiple microwave SM products into active-only, passive-only, and combined active-passive products, respectively. Here, we used the ESA CCI SM v7.1 combined product at a resolution of 0.25° (https://climate.esa.int/en/projects/soil-moisture/data/), which covered the period 1978–2021. Despite being the most widely used long-term satellite SM product, it suffers from spatial incompleteness due to the lack of satellite observations in the earlier years, the observation gaps in satellite orbits, and the physical limitations of microwave observations for SM retrieval over densely vegetated areas (Dorigo et al., 2017). In this study, the spatial consistency between the ESA CCI combined SM product and our GLASS-AVHRR product was investigated.

3 Methods

Figure 1 shows the flowchart of the proposed long-term global GLASS-AVHRR SM estimation framework, which consists of three main parts: data preprocessing and training samples preparation, model training and performance comparison, and generation and evaluation of the GLASS-AVHRR SM product.

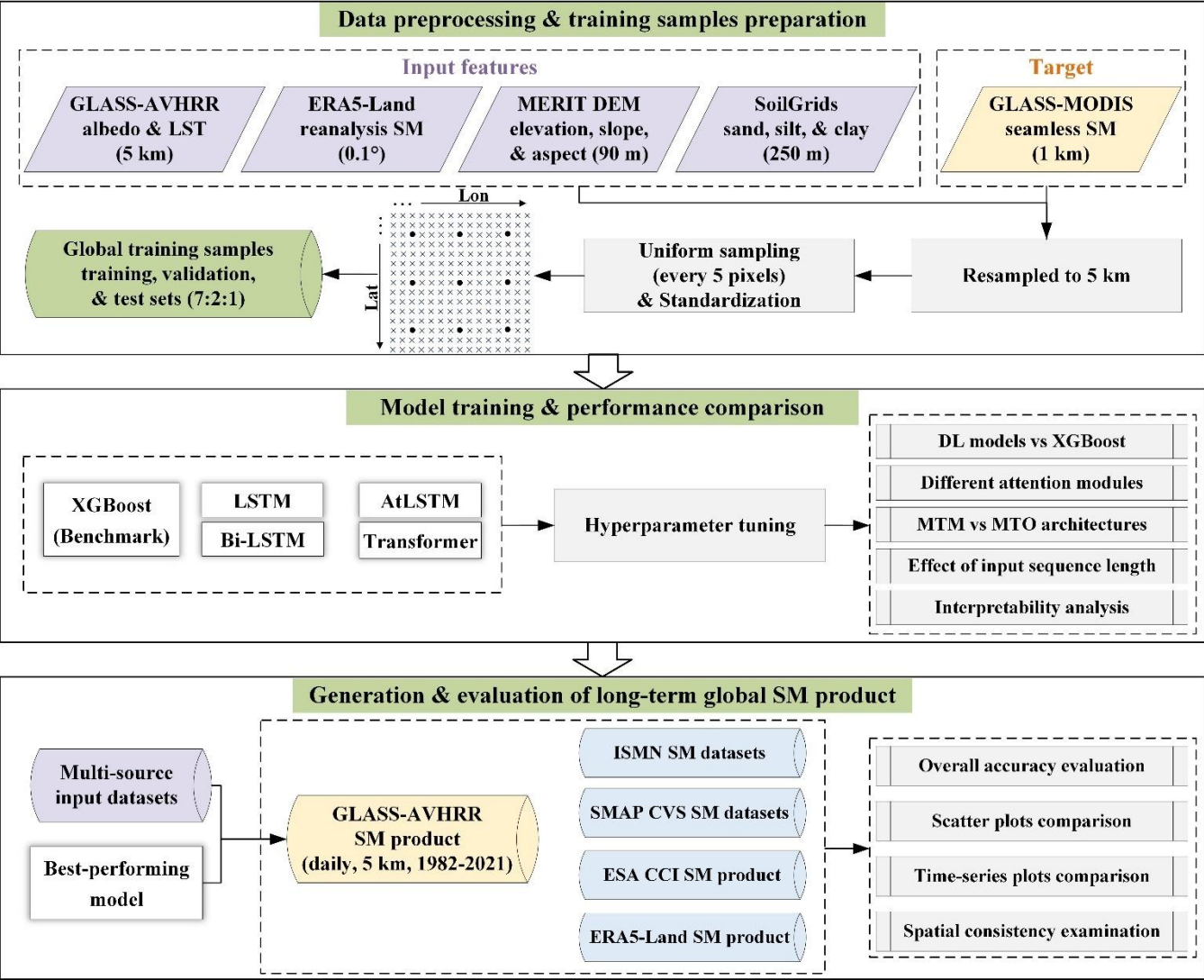



Figure 1 Flowchart of the proposed long-term global GLASS-AVHRR SM estimation framework

3.1 Training samples

The global GLASS-MODIS SM product resampled at 5 km was used as the training target of the long-term SM estimation model, from which a large number of representative and evenly distributed training samples could be obtained. Considering

that the size of training samples would be too large if all the pixels were included, these samples were selected at 25 km (5 pixels) intervals along both the longitude and latitude, and a total of 135,360 pixels were chosen after excluding those with a large proportion of missing values. Based on the geographic coordinates of these pixels, the values corresponding to each input feature as well as the target SM for the years 2005, 2010, and 2015 were extracted, which collectively formed the time-series training samples. While the three years were selected to represent different periods within the available time span (2000–2020), this selection may introduce some uncertainty, as climate and environmental conditions can vary annually, and extreme weather or climate events in certain years may affect the representativeness of variables such as LST and SM. Nevertheless, this approach was adopted to control the sample size while ensuring the representativeness of samples across different years. These samples were then randomly divided into training, validation, and test datasets in the ratio of 7:2:1 based on their locations, ensuring spatial independence with distances between any two samples exceeding 25 km, thereby minimizing the influence of spatial autocorrelation. While the training and validation datasets were used to train and tune the hyperparameters of the models, the accuracy of the models was evaluated on the test dataset. Figure 2 clearly illustrates the process of constructing time-series input samples for the DL models. Note that, the input features need to be scaled before training a DL model, which helps to speed up the convergence process, avoids bias towards larger-scale features, and improves the model stability. Here, each input feature was standardized by subtracting the mean and then dividing by the standard deviation, whereas for the target SM, no further processing is needed as it is by definition scaled.

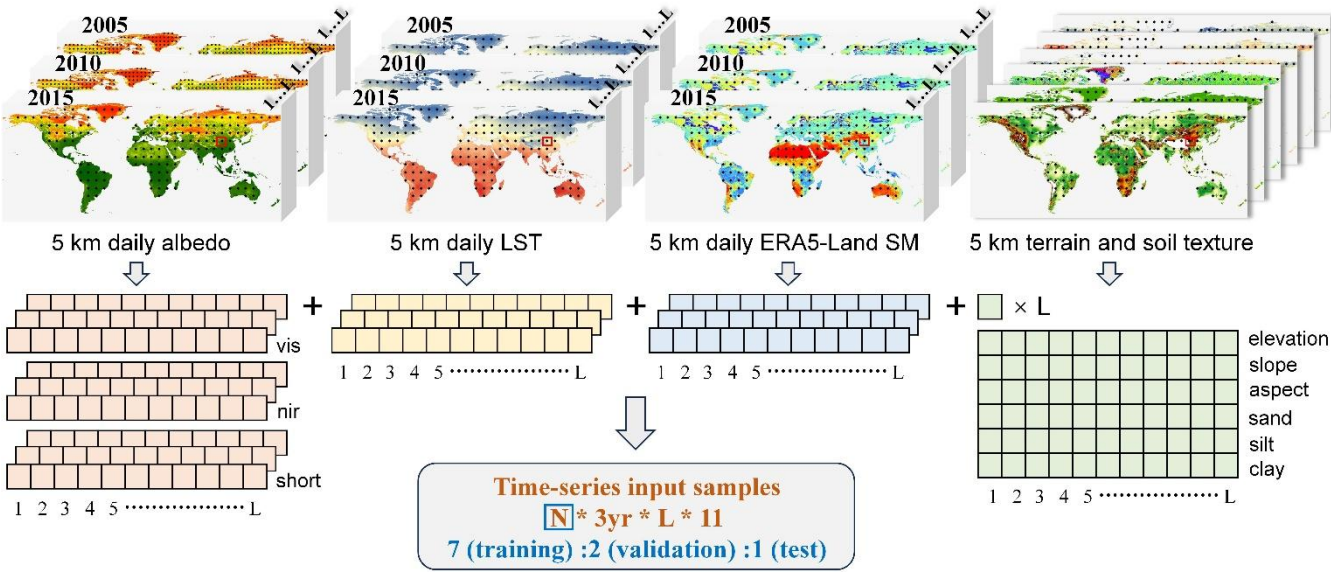



Figure 2 Schematic diagram illustrating the construction of time-series input samples for the DL models. N denotes the total number of samples, and L represents the sequence length, a hyperparameter that needs to be tuned.

3.2 Benchmark model

When generating the global 1 km GLASS-MODIS SM product, an XGBoost model was employed to integrate the multi-source datasets because of its good performance and high training and predicting speed. Here, we used the XGBoost model as a benchmark and compared its performance with the DL models (LSTM-based and Transformer) to analyze whether the DL models exhibit an advantage over this widely used machine learning model in SM estimation. The XGBoost model (Chen and Guestrin, 2016) is a type of gradient boosting model, in which multiple trees are iteratively constructed through correcting the prediction residuals of the preceding trees. The schematic diagram of the XGBoost model is shown in Fig. 3e, where predictions from multiple trees are combined to make the final SM prediction. The key hyperparameters were configured as follows: n_estimators = 1000, learning_rate = 0.1, and max_depth = 8. The time-series training samples constructed above were put together to train the XGBoost model, and the overall accuracy achieved by the XGBoost model on the test dataset was then compared with that of the DL models as a benchmark.

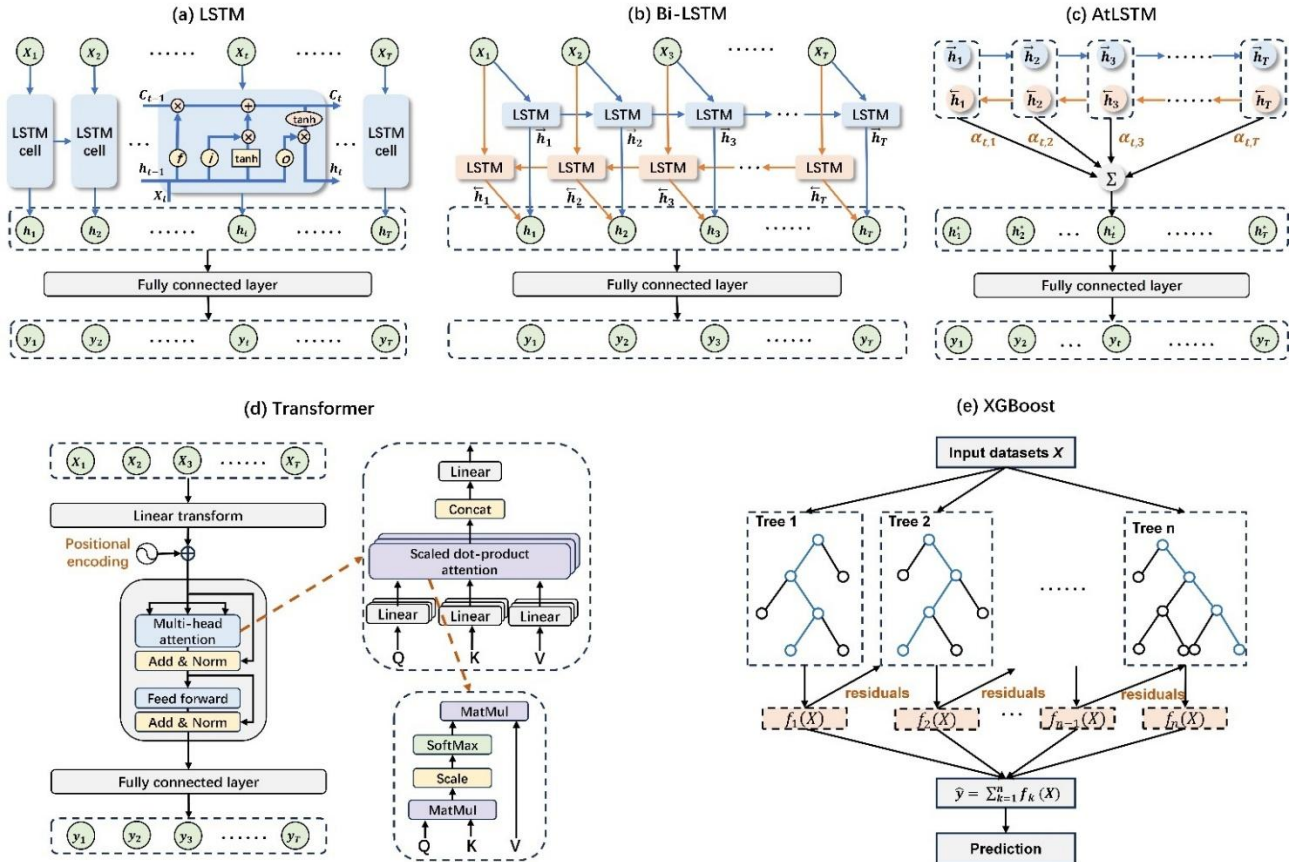



Figure 3 Schematic diagrams of the five models used in this study: (a) LSTM, (b) Bi-LSTM, (c) AtLSTM, (d) Transformer, and (e) XGBoost. In subplots (a-d), x_t, y_t, and h_t represent the input datasets, SM prediction, and hidden state output by the models at time step t, respectively.

3.3 Long short-term memory-based models

The LSTM network (Hochreiter and Schmidhuber, 1997) is a special type of recurrent neural network (RNN) designed to

275   solve the problems of gradient vanishing and exploding when training long sequences. The basic LSTM network introduces the memory cell, which is a special type of hidden state that shares the same shape as the hidden state but is designed to record long-term information. Each recurrent unit within the LSTM has three distinct gates, i.e., the forget gate, input gate, and output gate, as illustrated in Fig. 3a. The formulas used to calculate the three gates (f_t, i_t, o_t), cell state (c_t), and hidden state (h_t) are given below:

280    $$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3}$$

$$c_t = f_t * c_{t-1} + i_t * \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$h_t = o_t * \tanh(c_t) \tag{5}$$

285   where x_t represents the input datasets at time step t, h_{t-1} is the hidden state at the previous time step; f_t, i_t, and o_t are all calculated as linear functions of x_t and h_{t-1} with different weights and biases, and are then rescaled using a non-linear sigmoid (σ) function. The σ function acts as the gating function for the three gates with an output ranging between 0 and 1, thereby determining which portion of the information passes through the gates. Both the σ and tanh functions add non-linearity to the LSTM network. The Bidirectional LSTM (Bi-LSTM) extends the LSTM network by incorporating both

290   forward and backward LSTM units within a single layer, allowing the model to capture contextual information from both directions before concatenating their outputs. As displayed in Fig. 3b, the Bi-LSTM model can learn bidirectional (preceding and following) information at each time step.

The LSTM network has different architectures, including many-to-one (MTO) and many-to-many (MTM). In research areas like crop mapping and runoff prediction, the MTO architecture is primarily adopted, which uses inputs from multiple time

295   steps to output estimates for a single time step. Alternatively, we adopt the MTM architecture, which takes time-series inputs and outputs SM estimates for all time steps simultaneously by feeding the hidden states from all time steps into a fully connected layer. We also conducted an experiment to compare the estimation accuracy of these two architectures.

In addition to the basic LSTM and Bi-LSTM networks introduced above, an attention module was added to the Bi-LSTM network, referred to as the AtLSTM network, to explore if the estimation accuracy of SM could be further improved. The

300   AtLSTM network was constructed based on Bahdanau et al. (2016) and Xu et al. (2020), and adapted here for the MTM architecture. As illustrated in Fig. 3c, the attention module generates the attention weights (α), which are then multiplied with the hidden states (h) to get the weighted hidden states (h^*). The α and h^* can be calculated as follows:

$$e_t = W_a \cdot h_t + b_a \tag{6}$$

$$\alpha_{t,i} = softmax(e_{t,i}) = \frac{exp(e_{t,i})}{\sum_{j=1}^{T} exp(e_{t,j})} \tag{7}$$

305 $$h_t^* = \sum_{i=1}^{T} \alpha_{t,i} * h_i \qquad (8)$$

where W_a and b_a denote the learnable parameters that map the hidden states h into a weight matrix e, and T is the sequence length of the input features. The weight matrix (with the shape of $T \times T$) is then rescaled by a softmax function to obtain the attention weights for each hidden state, which range between 0 and 1 and sum to 1. The weighted hidden states h^* are then fed into a fully connected layer to estimate the target variable. Intuitively, higher attention weights indicate that the 310 corresponding hidden states have a greater influence on the estimation of SM at a specific time step.

In this study, the LSTM-based models were implemented using the open-source PyTorch 2.0 framework. The mean square error (MSE) between was used as the loss function, and the Adam optimizer was adopted to update the learnable parameters of the models. Several key hyperparameters were tuned, including the hidden size, number of epochs, and learning rate (Zhang et al., 2021). For each model, the hidden size was determined after testing values of 64, 128, 256, and 512; the 315 number of epochs after testing 20, 50, 100, and 200; and the learning rate after testing 0.1, 0.01, 0.001, and 0.0001. The final settings of the major hyperparameters for the three LSTM-based models are listed in Table 4.

Table 4 Key hyperparameters configured for the DL models used in this study.

Hyperparameters	LSTM	Bi-LSTM	AtLSTM	Transformer
Hidden size	256	256	256	64
Number of heads	/	/	/	4
Number of epochs	100	100	200	100
Number of layers	1	1	1	1
Batch size	100	100	100	100
Learning rate	1e-3	1e-3	1e-4	1e-3
Sequence length	425	425	425	365

3.4 Transformer

The Transformer network is a DL architecture based entirely on attention mechanisms, dropping the recurrent structure to 320 avoid the constraint of sequential calculation. After being proposed by Vaswani et al. (2017), Transformer has soon become the state-of-the-art model for natural language processing and has also been applied successfully to areas like computer vision (Dosovitskiy et al., 2021) and time series analysis (Wen et al., 2023). Its core component is the multi-head self-attention layers, which can relate any two positions in a sequence. More specifically, multi-head attention involves applying the attention function to multiple sets of key, value, and query vectors in parallel, thus enabling the model to focus on 325 different parts of the input sequence simultaneously. Unlike the attention function used in the AtLSTM model (Eq. (6-7)), Transformer uses the scaled dot-product attention α, which can be calculated as follows:

$$\alpha (Q, K, V) = softmax \left(\frac{QK^T}{\sqrt{d_k}} \right) V \qquad (9)$$

where Q, K, and V refer to the query, key, and value vectors, respectively, which are derived by multiplying the embedded input sequence with the corresponding learnable projection matrix; d_k is the dimension of the key and query vectors.

330 Additionally, with the help of a positional encoding function, the Transformer network can retain some ordinal information for elements in the input sequence. A detailed description of Transformer and the multi-head self-attention mechanism can be found in Vaswani et al. (2017). Compared with recurrent or convolutional neural networks, the Transformer network can efficiently parallelize much larger amounts of computation and capture long-range dependencies in the input sequence more easily. Here, we only used the encoder portion of the original Transformer network to map the input features into hidden

335 representations, which were then fed into a fully connected layer to output the time-series SM estimates (Fig. 3d). The same training samples, optimizer, and loss function used for the LSTM-based models were employed to train the Transformer network, with the settings of its hyperparameters also listed in Table 4. Notably, the number of heads is a unique hyperparameter of Transformer that refers to the number of parallel self-attention layers of the encoder.

3.5 Evaluation of the models and GLASS-AVHRR SM product

340 After training the benchmark XGBoost model and the four DL models described above using the same training samples distributed worldwide, their performance on the test set was then compared from multiple perspectives, including comparisons between the DL models and XGBoost model, between the DL models with different attention mechanisms, and between the DL models with MTM or MTO architectures. Moreover, the effect of input sequence length on model accuracy was investigated using the LSTM-based models, and a preliminary interpretability analysis was performed through

345 visualizing the attention weights of both the AtLSTM and Transformer models. Then, the best-performing model, along with the multi-source input datasets, was employed to generate the global daily GLASS-AVHRR SM product at 5 km resolution from 1982 to 2021. To fully assess the derived long-term SM product, different SM datasets and evaluation strategies were combined, including overall accuracy evaluation, scatter plots analysis, time-series plots comparison, and spatial consistency examination. Specifically, the accuracy of this product was first evaluated against the point-scale ISMN, field-scale

350 COSMOS, and upscaled 9 km SMAP CVS in situ SM datasets, respectively. Then, the GLASS-AVHRR SM product was intercompared with the GLASS-MODIS SM product and two widely used long-term global SM products, namely the ERA5-Land and ESA CCI, to investigate their spatial consistency.

4 Results

4.1 Comparison of model performance

355 Table 5 lists the performance metrics achieved by the benchmark tree-based XGBoost model and four DL models on the training set, validation set, and two types of test sets, respectively. The XGBoost model achieved similar overall accuracy across the training, validation, and test sets, with a coefficient of determination (R^2) of 0.984 and RMSE of 0.012 m^3 m^{-3} on the training set, and an R^2 of 0.982 and RMSE of 0.013 m^3 m^{-3} on both the validation and test sets, indicating a low tendency

for overfitting. The fairly high overall accuracy of the benchmark XGBoost model may be attributed to the large number of training samples, specifically 135,360 pixels per day over 3 years, evenly distributed across the globe on a daily basis. To evaluate the impact of sample size on model performance, we conducted an experiment by reducing the number of training samples. When the sample size was reduced by a factor of 100, the accuracy of the XGBoost model dropped considerably, with an R^2 of 0.96 and RMSE of 0.017 m^3 m^{-3} on the test set. This highlights the importance of having sufficient samples to achieve high accuracy with XGBoost and indicates the advantage of using the GLASS-MODIS SM product as training target, which can provide much richer samples than the sparse in situ ISMN SM dataset. Meanwhile, Table 5 also shows that the accuracy of the XGBoost model decreases drastically on the test set with SM observations exceeding 0.4 m^3 m^{-3}, yielding an R^2 of 0.413 and RMSE of 0.022 m^3 m^{-3}, likely due to the relatively smaller portion of samples at high SM levels.

In comparison, the LSTM model developed using time-series training samples performed slightly better than the XGBoost model, with the R^2 on the test set increasing to 0.983, and when the Bi-LSTM model was employed, the overall accuracy on the test set was further improved, with the R^2 increasing to 0.985 and RMSE decreasing to 0.012 m^3 m^{-3}. Although the increase in the overall accuracy might not be significant, the Bi-LSTM model exhibited significant improvement over the XGBoost model at high SM levels, achieving an R^2 of 0.482 and RMSE of 0.020 m^3 m^{-3} on the test set for observations exceeding 0.4 m^3 m^{-3}. As also can be seen from the density scatter plots in Fig. 4, the majority of samples had SM values below 0.4 m^3 m^{-3} (indicated by the red dots), where all models achieved high prediction accuracy. However, on the relatively infrequent samples with high SM values, where the XGBoost model tended to yield lower estimates, both the LSTM and Bi-LSTM models provided more accurate estimates. Given the temporal autocorrelation of SM, these results suggest that learning both forward and backward temporal information from the time-series training samples enhances the ability of DL models to estimate SM more accurately, especially at high SM levels with sparser samples.

Table 5 Performance metrics of the benchmark XGBoost model and four DL models on the training set, validation set, and two types of test sets, respectively.

Model	Training set		Validation set		Test set		Test set (> 0.4 m^3 m^{-3})	
	R^2	RMSE (m^3 m^{-3})	R^2	RMSE (m^3 m^{-3})	R^2	RMSE (m^3 m^{-3})	R^2	RMSE (m^3 m^{-3})
XGBoost	0.984	0.012	0.982	0.013	0.982	0.013	0.413	0.022
LSTM	0.986	0.012	0.983	0.013	0.983	0.013	0.424	0.021
Bi-LSTM	0.988	0.011	0.984	0.012	0.985	0.012	0.482	0.020
AtLSTM	0.990	0.010	0.986	0.011	0.987	0.011	0.621	0.016
Transformer	0.990	0.010	0.984	0.012	0.985	0.012	0.460	0.021

Then, after adding the attention module into the Bi-LSTM model, the derived AtLSTM model achieved the best performance, with an R^2 of 0.987 and RMSE of 0.011 m^3 m^{-3} on the test set. In contrast, despite that the Transformer model also incorporated an attention module, its accuracy was slightly lower than that of the AtLSTM model on the test set and significantly lower on samples with high SM levels (> 0.4 m^3 m^{-3}) in our experiments. As mentioned above, the main

advantage of the Transformer model is its ability to capture long-range dependencies and handle long sequences effectively. However, soil moisture often exhibits high temporal variability, meaning it can change rapidly due to factors such as rainfall and evaporation. In this context, short-term adjacent temporal information can be critical for accurate SM estimation. The slightly better performance of the AtLSTM model compared with the Transformer model may be attributed to its superior ability to capture these short-term adjacent dependencies, which are critical for modeling the nuances in rapidly changing SM levels. This will be further investigated through the analysis of attentional weights below. Additionally, a feature importance analysis was conducted for the best-performing AtLSTM model, as shown in Fig. A3. Specifically, the gradients of the model's output with respect to each input feature were computed on the test set, and the absolute values of these gradients were then averaged across all samples and time steps. Input features with larger average gradients are considered to exert a more significant influence on the model's predictions. The results indicate that elevation, black-sky visible albedo, ERA5-Land reanalysis SM, and slope are the most influential features for the AtLSTM model. In particular, although elevation is a static variable, it plays a critical role in shaping the spatial distribution of SM by influencing precipitation, temperature, vegetation type, and evaporation processes. Its impact on the spatial variability of SM tends to be more stable and consistent over time. In contrast, the contributions of dynamic input features such as ERA5-Land SM may fluctuate across time and space and can be diminished by inherent uncertainties and biases in the input data. Moreover, their importance may also be influenced by correlations with other input features.

While the numerical differences in overall accuracy among all these models may not seem remarkable, a more intuitive comparison can be drawn from their density scatter plots. As shown in Fig. 4, on the majority of samples, both the best-performing AtLSTM model and benchmark XGBoost model can achieve high prediction accuracy, resulting in a relatively small difference in their overall performance on the test set. However, there remains a small portion of samples that are more challenging to predict, on which the SM estimates from the AtLSTM model are much closer to the 1:1 line compared with the XGBoost model. Furthermore, the AtLSTM model significantly improves upon the tendency of the XGBoost model to produce lower estimates at high SM levels, achieving an R^2 of 0.621 and RMSE of 0.016 $m^3\ m^{-3}$ on the test set for observations exceeding 0.4 $m^3\ m^{-3}$. Overall, while both the XGBoost model and the four DL models can achieve high SM estimation accuracy, the AtLSTM model yields the highest accuracy among them and performs well across different SM levels with a low tendency for overfitting. This suggests that utilizing bidirectional temporal information from the input sequence as well as adding an attention module are both effective in further improving the estimation accuracy of SM.

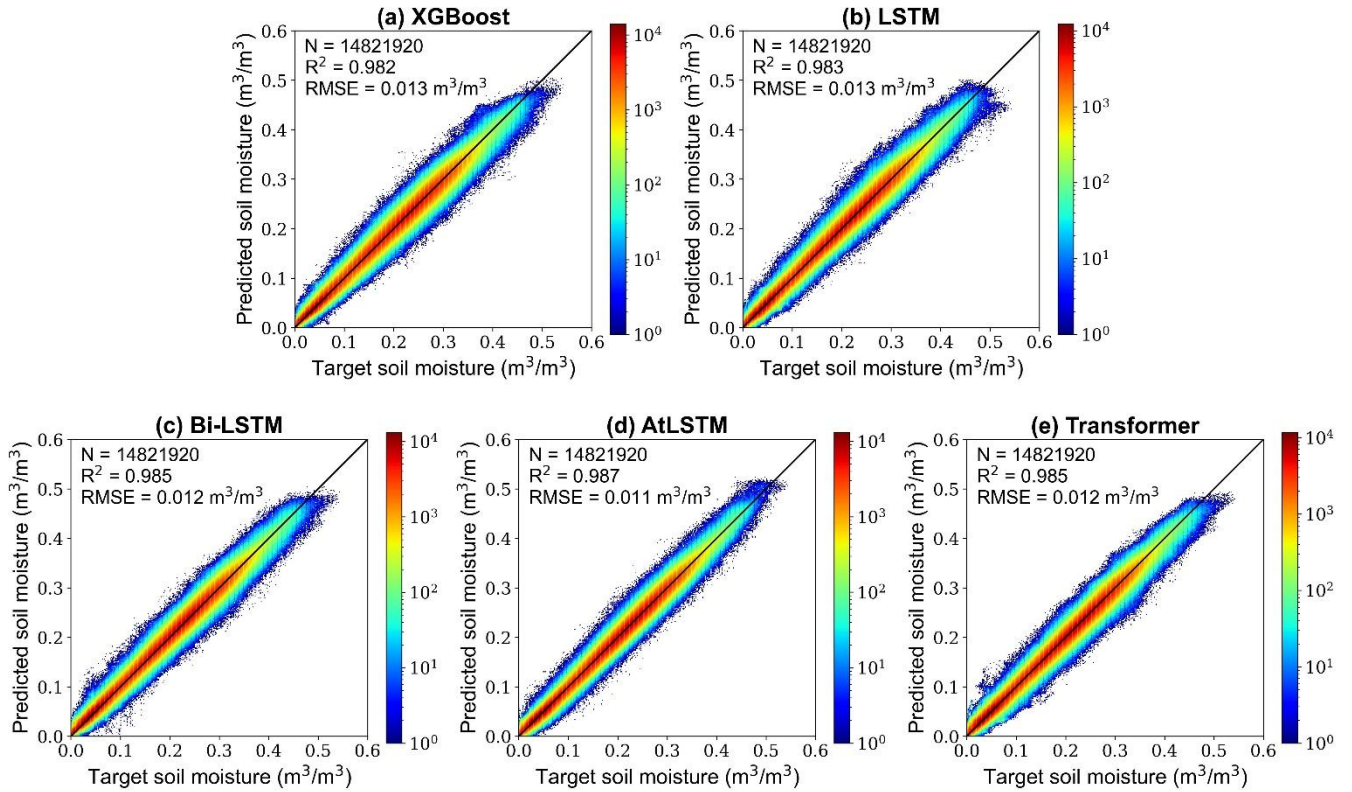



Figure 4 Scatter plots between target SM and predicted SM for the (a) XGBoost, (b) LSTM, (c) Bi-LSTM, (d) AtLSTM and (e) Transformer models on the test set. The colors of the dots indicate different probability densities, and the black line represents the 1:1 line.

As mentioned in Sect. 3.3, We chose to use the MTM architecture when developing the DL models to output time-series SM estimates at once. Here, to compare the accuracy of the MTM architecture with the more commonly used MTO architecture, as well as to investigate the effect of input sequence length on model accuracy, we calculated performance metrics for the LSTM models utilizing these two different architectures under varying lengths of input sequences. Specifically, both types of models were trained using input features from a given date (e.g., the first day of 2015) and n days (0-29) prior to that date, respectively, and the accuracy of the models was then evaluated on the test set for that given date. To save training time, the number of epochs for these LSTM models was set to 20. It can be seen from the R^2 and RMSE curves in Fig. 5a that, as the length of input sequence increased, the accuracy of the LSTM model with the MTO architecture also increased, and then the accuracy leveled off at a sequence length of about 10 days. This indicates that while accounting for temporal information can be beneficial for current SM estimation, only the most recent input sequences have a remarkable effect on the model's accuracy. In comparison, the LSTM model with the MTM architecture, which can output a sequence of SM estimates simultaneously, achieved similar accuracy to that of the MTO architecture, and its R^2 and RMSE curves stabilized at a

sequence length of about 5 days. This demonstrates the feasibility of adopting the MTM architecture in the LSTM model, which not only reduces considerable production time but also maintains the estimation accuracy.

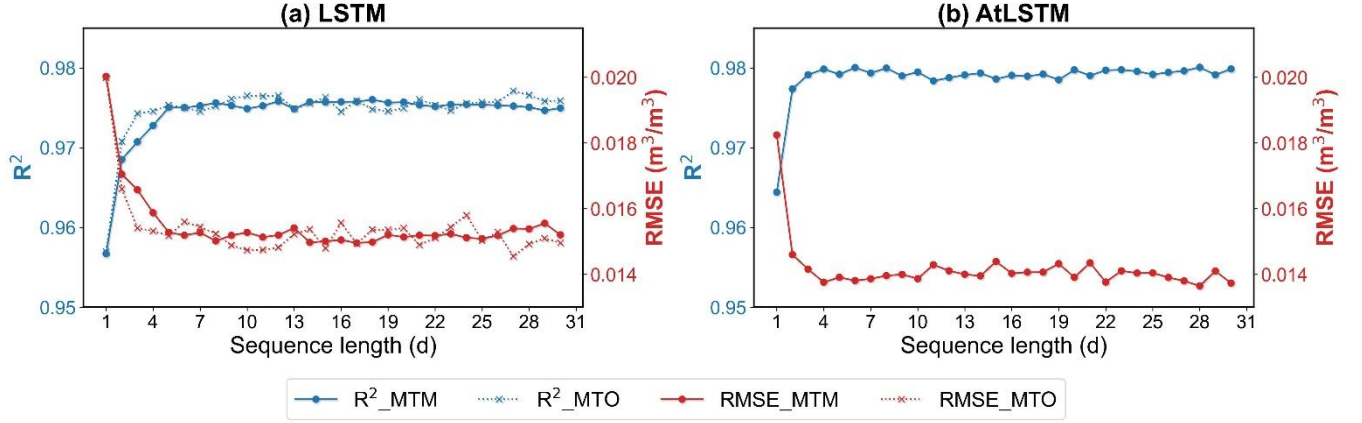



Figure 5 Performance metrics of the (a) LSTM models with two different types of architectures (MTO and MTM) and (b) AtLSTM model with the MTM architecture trained using varying lengths of input sequences on the test set. The blue and red curves represent the R^2 and RMSE curves, respectively.

Moreover, we also investigated the effect of input sequence length on the overall accuracy of the AtLSTM model with the MTM architecture, and the performance metrics were calculated here based on SM estimates over the entire time series instead of on a given date. To save training time and accounting for the smaller learning rate used for the AtLSTM model (Table 4), the number of epochs was set to 50. As displayed in Fig. 5b, the overall accuracy of the AtLSTM model increased sharply as the length of input sequence increased, and then the accuracy plateaued at a sequence length of about 4 days. The more rapid stabilization of the AtLSTM model's accuracy may be attributed to the incorporation of the Bi-LSTM module in the model, which can utilize both forward and backward temporal information. In addition, it seems that when the input sequence is long enough, the model can automatically learn the necessary temporal information to accurately estimate SM at each position in the sequence. However, it should be noted that at the beginning or end of the sequence, the model's accuracy tends to decrease as only forward or backward information can be utilized, which is a common issue encountered by the LSTM-based models with the MTM architecture. Therefore, to facilitate the production process, the sequence length of the LSTM-based models was finally set to 425 days, and both the first 30 and last 30 values were discarded (a rather sufficient number) after the model output the time-series SM estimates, so that an entire year's SM estimates could be obtained in a single run. Note that, during both the training and production phases, the first and last 30 days of each 425-day sequence were padded with actual data from adjacent years to ensure consistency.

Although data-driven DL models are commonly perceived as "black boxes", there are many techniques that can be employed to increase the interpretability of DL models. In the case of attention-based deep neural networks, this can be achieved by analyzing the distribution of attention weights. In a long sequence, perhaps only a portion of the information is critical to the model prediction at a given time step, and the attention mechanism enables the model to focus on these critical positions. In

particular, the attention module of the AtLSTM model can dynamically adjust the weights of the hidden states output by the model at each time step. Figure 6a illustrates the distribution of the averaged attentions weights calculated using the best-performing AtLSTM model on the test set (40,608 samples). To show more detail, only the attention weights of 30 consecutive days selected from the entire sequence (425 days) are displayed here, and attention weights less than 0.0001 are masked out. It is observed that, for the hidden state at each time step in the sequence (vertical axis), the largest attention weight was located approximately 3 days around that time step (horizontal axis). This indicates that when the attention module of the AtLSTM model learns to readjust the hidden states, it primarily utilizes the temporal information from adjacent positions in the sequence.

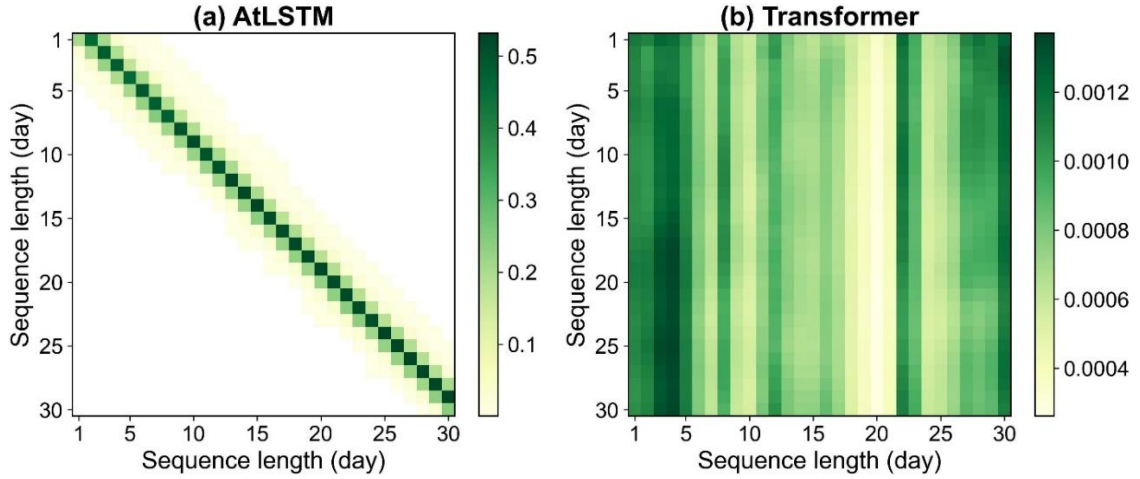



Figure 6 Heatmaps of the averaged attention weights calculated using the (a) AtLSTM and (b) Transformer model on the test set (40,608 samples). Only the attention weights of 30 consecutive days selected from the entire sequence are displayed here for illustration.

In contrast, as a core component of the Transformer model, the multi-head self-attention layers can capture various aspects of relationships between different positions within a sequence, and the attention weights generated by these layers are then directly applied to the embedded input sequence. Figure 6b shows, as a comparison, the distribution of attention weights calculated by averaging the outputs from the four attention heads of the Transformer model. The attention weights heatmap of the Transformer model is quite different from that of the AtLSTM model, with the weight at each position being much smaller and dispersed. This is likely because the self-attention module can relate any two positions in the sequence, and inputs from more distance positions may contribute more to the model output at the current time step. In addition, for each time step in the sequence (vertical axis), there were some common positions (horizontal axis) with larger weights that were more important for model prediction. Despite the distinct attention mechanisms employed by these two DL models, both of them achieved high SM estimation accuracy. Given that SM is temporally autocorrelated and highly variable over time, the slightly better performance of the AtLSTM compared to the Transformer model may be attributed to that it extracts temporal information mainly from adjacent positions in the sequence, rather than from more distance ones, for SM estimation.

4.2 Validation of the GLASS-AVHRR SM product

After generating the GLASS-AVHRR SM product using the best-performing AtLSTM model with the MTM architecture, permanent snow and ice as well as water bodies were masked out with the help of the MODIS land cover type product (MCD12C1) (Friedl and Sulla-Menashe, 2022). The derived SM product was then evaluated against three types of in situ SM datasets at different spatial scales. The first type is the point-scale ISMN SM dataset, which is distributed globally and covers a wide range of land cover types. There were 1672 ISMN stations and 715 spatially representative stations available for validation during Period I (2000–2018). The distribution of validation metrics achieved by the GLASS-AVHRR SM product on these partially independent ISMN stations during Period I, grouped by all stations and representative stations, is presented in Fig. 7, alongside those of the GLASS-MODIS and ERA5-Land SM products for comparison. The GLASS-AVHRR SM product achieved comparable performance to that of the GLASS-MODIS SM product across all ISMN stations and representative stations during Period I. In addition, both GLASS SM products performed significantly better at the representative stations. This demonstrates the high level of consistency in accuracy between the two GLASS SM products. Note that the validation metrics for the GLASS-MODIS product were derived using a site-independent cross-validation method, which was designed to accurately reflect the product's performance over unknown areas. Given the consistency in the distribution of validation metrics between the GLASS-AVHRR and GLASS-MODIS SM products, the accuracy achieved by the GLASS-AVHRR product at these partially independent stations should also approach its true accuracy. In contrast, although the ERA5-Land SM product achieved a similar distribution of R to the two GLASS SM products across all ISMN stations and representative stations, it exhibited much larger biases and ubRMSE values.

To conduct a more independent evaluation of the GLASS-AVHRR SM product, the ISMN SM dataset from Period II (1982–1999) was also collected. After excluding stations that overlapped with the 715 representative stations from Period I, only 45 independent stations remained for evaluation during Period II. The observations at these stations were also quite limited, hence the validation metrics derived from them may not provide a comprehensive assessment. Nevertheless, it can be seen from Fig. 7 that the GLASS-AVHRR product achieved rather high accuracy at these stations, with a median R of 0.73 and a median ubRMSE of 0.041 m^3 m^{-3}. Likewise, while the ERA5-Land SM product exhibited a similar distribution of R to the GLASS-AVHRR product at these stations, it achieved much larger biases and ubRMSE values. The second type of in situ SM dataset comprises field-scale measurements from three COSMOS networks: COSMOS, COSMOS-UK, and COSMOS-Europe, which can provide an independent evaluation of the GLASS-AVHRR SM product at an intermediate scale. As shown in Fig. 7, the GLASS-AVHRR, GLASS-MODIS, and ERA5-Land SM products all achieved good performance across the three COSMOS networks. The two GLASS SM products showed comparable overall accuracy across these networks, although some site-specific discrepancies were observed, which are likely due to differences in the satellite remote sensing inputs and spatial resolution. Yet, their accuracies varied considerably across these networks, with the median R ranging from 0.63 to 0.79 and the median ubRMSE ranging from 0.044 to 0.065 m^3 m^{-3} for the GLASS-AVHRR product. This variability may be attributed to the different footprint radii of COSMOS sensors, which result in varying degree of

510   spatial representativeness and spatial mismatches with gridded SM products. These factors can introduce uncertainty into the validation results, particularly affecting the bias and ubRMSE metrics. The biases of the GLASS-AVHRR SM product on the COSMOS-UK network were much larger than those on the other two COSMOS networks, with the median bias reaching -0.09 m^3 m^{-3}. This is likely due to the greater sensing depth of the COSMOS-UK network, which has a median depth of 30 cm, compared to 21 cm and 22 cm for the COSMOS and COSMOS-Europe networks, respectively (Fig. A2). Moreover,

515   both the GLASS-AVHRR and ERA5-Land SM products exhibited larger ubRMSE values on the COSMOS-UK network. This may be related to the increased uncertainty of COSMOS measurements in organic soils or humid regions, which are prevalent in the UK, as also reported by Zheng et al. (2024). Meanwhile, although the first-layer (0–7 cm) ERA5-Land SM product was used here for evaluation, it still exhibited large wet biases across these COSMOS networks, further suggesting its extensive overestimation issue.

520

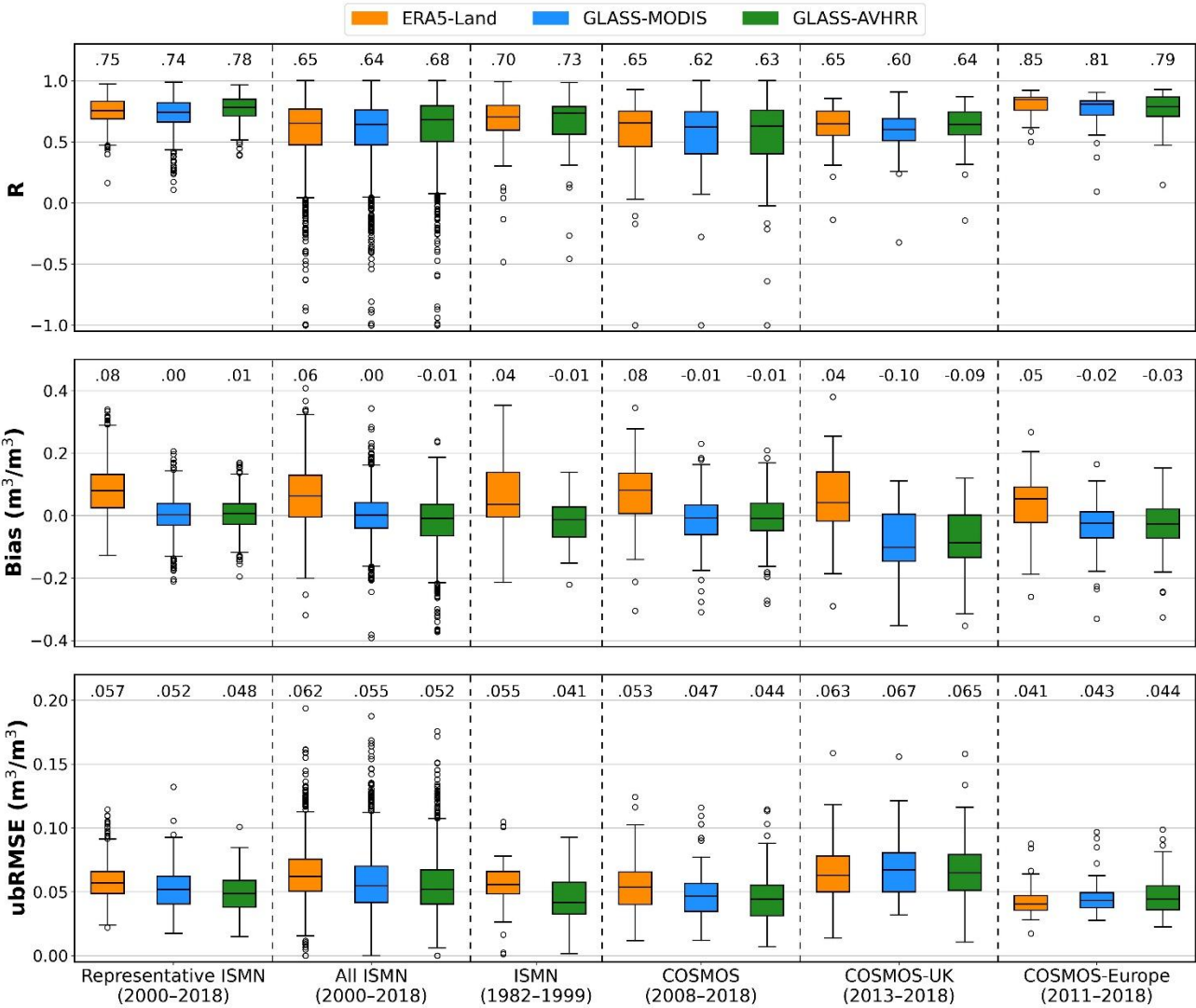



Figure 7 Boxplots of R, bias, and ubRMSE for the GLASS-AVHRR SM product across different groups of ISMN stations and three field-scale COSMOS networks, in comparison with the GLASS-MODIS and ERA5-Land SM products. The number above each box represents the median value of the metrics across all stations within each network.

Despite the high accuracy achieved when validating the GLASS-AVHRR SM product using both the point-scale ISMN and field-scale COSMOS in situ SM datasets, the validation results were inevitably affected by the scale differences between these datasets. Therefore, the upscaled 9 km SMAP CVS in situ SM dataset from 22 different locations was also utilized to validate the GLASS-AVHRR SM product from 2015 to 2021 as a complement. Specifically, the mean SM values of the 5 km GLASS-AVHRR SM product within a 2 × 2 window corresponding to each 9 km SMAP CVS grid were first calculated, and then the validation metrics for the GLASS-AVHRR SM product were estimated at each CVS, as listed in Table 6. As a comparison, validation metrics for the ERA5-Land SM product (~ 9 km horizontal resolution) were also calculated at each CVS and presented in the table.

Table 6 Validation metrics for the GLASS-AVHRR and ERA5-Land SM products at 22 upscaled 9 km SMAP core validation sites, with the best-performing metrics highlighted in bold.

Site	GLASS-AVHRR				ERA5-Land				LC	No.
	R	bias ($m^3\ m^{-3}$)	RMSE ($m^3\ m^{-3}$)	ubRMSE ($m^3\ m^{-3}$)	R	Bias ($m^3\ m^{-3}$)	RMSE ($m^3\ m^{-3}$)	ubRMSE ($m^3\ m^{-3}$)		
HOBE	0.61	-0.07	0.100	0.069	**0.63**	**-0.02**	**0.069**	**0.066**	Croplands	252
Kenaston1	**0.76**	-0.07	0.078	**0.036**	0.72	**0.02**	**0.051**	0.048	Croplands	87
Kenaston2	**0.80**	-0.08	0.084	**0.035**	0.77	**0.01**	**0.046**	0.045		87
Carman	**0.71**	**0.01**	**0.042**	**0.042**	0.61	0.10	0.115	0.053	Croplands	145
South Fork	0.61	**0.00**	**0.062**	0.062	**0.67**	0.07	0.096	**0.060**	Croplands	179
St. Josephs	0.71	-0.07	0.077	0.037	**0.75**	**0.05**	**0.063**	0.035	Croplands	115
REMEDHUS1	**0.87**	**0.05**	**0.051**	**0.022**	0.86	0.16	0.172	0.071	Croplands	557
REMEDHUS2	**0.86**	**-0.04**	**0.050**	**0.034**	0.84	0.09	0.101	0.046		540
Valencia	0.54	**-0.01**	**0.047**	**0.045**	**0.59**	0.08	0.111	0.078	Savannas	107
Tonzi Ranch	**0.95**	**0.00**	**0.030**	**0.030**	0.94	0.09	0.097	0.045	Savannas	79
Fort Cobb	0.81	**0.01**	**0.034**	**0.034**	**0.83**	0.08	0.085	0.040	Grasslands	248
Little Washita	**0.78**	**0.01**	**0.039**	**0.038**	0.77	0.05	0.071	0.049	Grasslands	225
Walnut Gulch1	**0.71**	0.01	**0.030**	**0.027**	0.69	0.01	0.062	0.061	Shrublands	159
Walnut Gulch2	**0.74**	0.04	**0.042**	**0.021**	0.71	0.11	0.126	0.062		189
Little River	0.36	**0.00**	**0.043**	0.043	**0.76**	0.22	0.225	**0.040**	Cropland/ Natural mosaic	84
TxSON1	0.87	**0.00**	**0.024**	**0.024**	**0.88**	0.09	0.100	0.040	Grasslands	55

TxSON2	0.90	**0.02**	**0.028**	**0.023**	**0.91**	0.07	0.076	0.038		103
Niger	**0.73**	**0.00**	**0.018**	**0.017**	0.69	0.04	0.061	0.046	Grasslands	138
Benin	**0.91**	**0.04**	**0.052**	**0.037**	0.88	0.22	0.228	0.062	Savannas	217
Monte Buey	**0.78**	-0.07	0.081	**0.035**	0.74	**0.01**	**0.053**	0.052	Croplands	120
Yanco1	**0.92**	**-0.02**	**0.049**	**0.043**	0.87	0.04	0.064	0.050	Croplands	121
Yanco2	**0.90**	**0.00**	**0.035**	**0.035**	0.86	0.09	0.095	0.041	Grasslands	117
Average	0.77	**-0.01**	**0.050**	**0.037**	0.77	0.08	0.099	0.053	/	/
All	**0.82**	**-0.01**	**0.054**	**0.054**	0.65	0.09	0.119	0.083	/	3924

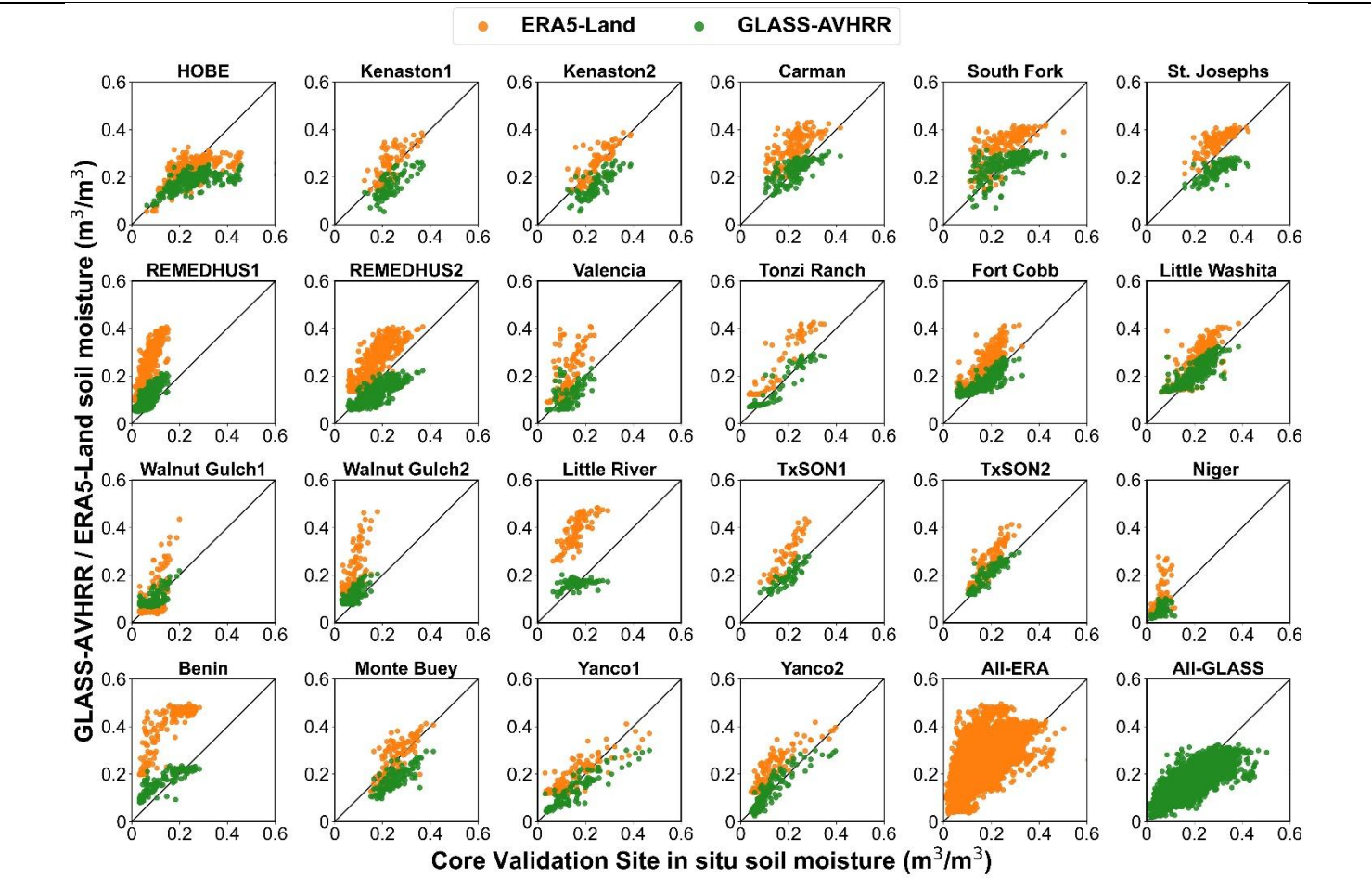



Figure 8 Scatter plots between the upscaled in situ SM and the corresponding estimated SM from the GLASS-AVHRR or ERA5-Land product at each SMAP core validation site.

At most of the CVSs, the GLASS-AVHRR SM product achieved similar R values to the ERA5-Land SM product, except at the Little River site, where the R value for the GLASS-AVHRR product was significantly lower. This is probably because the land cover type at this site is "Cropland or Natural mosaic", making the upscaled in situ SM measurements less

representative and the validation results at this site less reliable. Meanwhile, while the GLASS-AVHRR SM product exhibited notable dry biases only at a few CVSs, the ERA5-Land SM product showed large wet biases at most of the CVSs, as also reported in detail by Lal et al. (2022). The varying degrees of bias in these two SM products can be more intuitively observed through their scatter plots against the upscaled in situ SM at each CVS (Fig. 8). As one of the main inputs for generating the GLASS-AVHRR SM product, the ERA5-Land reanalysis SM exhibited notable wet biases at almost all CVSs, especially at REMEDHUS1, Little River, and Benin, which were largely corrected by the GLASS-AVHRR product, with the data points on the scatter plots being much closer to the 1:1 line. This can be attributed to the use of the GLASS-MODIS SM product as training target, although it may have also contributed to the slight dry bias in the GLASS-AVHRR SM product, given that optical and thermal satellite SM estimates typically represent a shallower depth than in situ SM datasets. In addition, at the CVS where the ERA5-Land product exhibited a large wet bias, the RMSE and ubRMSE values of the GLASS-AVHRR product were often much lower than those of the ERA5-Land product. The average R and ubRMSE values achieved by the GLASS-AVHRR SM product at 22 CVSs were 0.77 and 0.037 m^3 m^{-3}, respectively, similar to those reported for the 9 km SMAP-Sentinel L2 SM product, which were 0.79 and 0.035 m^3 m^{-3}, respectively (Das et al., 2020). When combining all the CVS in situ SM measurements, an overall R of 0.82 and ubRMSE of 0.054 m^3 m^{-3} were obtained by the GLASS-AVHRR SM product, showing significant improvement over the ERA5-Land SM product, which had values of 0.65 and 0.083 m^3 m^{-3}, respectively. This is also evident from the more concentrated scatter points of the GLASS-AVHRR SM product displayed in Fig. 8.

To intuitively examine the ability of the GLASS-AVHRR SM product to capture temporal variations in measured SM and its temporal consistency with the GLASS-MODIS product, time-series curves for the GLASS-AVHRR (aggregated at 10 km), GLASS-MODIS (aggregated at 9 km), and in situ SM (upscaled at 9 km) at six CVSs with different land cover types were plotted, with the ERA5-Land SM product (~ 9 km horizontal resolution) also included for reference (Fig. 9). Through extending the GLASS-MODIS SM product from 2000 back to 1982, the GLASS-AVHRR SM product attained complete temporal coverage from 1982 to 2021, and a high degree of temporal consistency between these two products could be observed from the time-series plots. Despite that the ERA5-Land SM product also had long-term temporal coverage, it exhibited large wet biases when compared with the upscaled in situ SM at all six CVSs, whereas both the GLASS-MODIS and GLASS-AVHRR SM products aligned more closely with the dynamic ranges of measured SM. As mentioned above, the GLASS-AVHRR SM product exhibited notable dry biases at a few CVSs. However, as can be seen from the time-series curves at REMEDHUS2 (Fig. 9a) and Yanco1 (Fig. 9f), suspicious abrupt rises in measured SM, as well as temporary spikes in SM (possibly caused by irrigation), might also have partially contributed to these dry biases. Overall, the GLASS-AVHRR SM product could well capture the temporal variations in measured SM at these CVSs, except for the Little River site (Fig. 9d) where the land cover type is "Cropland or Natural mosaic". Measured SM at this site did not show a clear seasonal pattern as at the other sites, and there was less consistency between the two GLASS SM products, likely due to the stronger spatial heterogeneity of this site. Besides, at the Walnut Gulch1 site (Fig. 9c) where the dominant land cover type is

"shrubland", while the GLASS-AVHRR product captured high SM values well, it slightly overestimated when the measure SM approached zero.

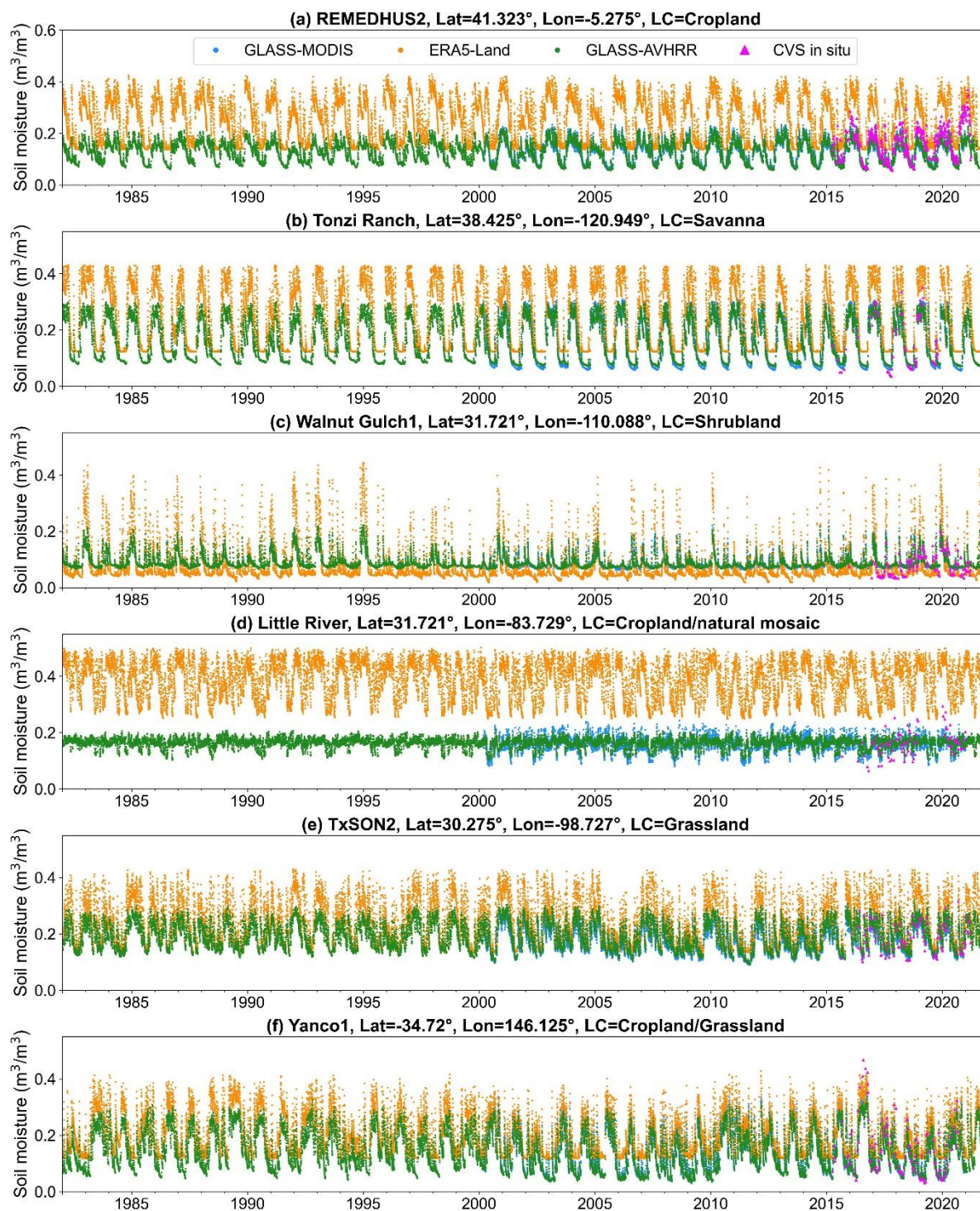



Figure 9 Time-series plots of the GLASS-AVHRR (aggregated at 10 km), GLASS-MODIS (aggregated at 9 km), ERA5-Land (~ 9 km horizontal resolution), and in situ SM (upscaled at 9 km) at six CVSs with different land cover types for the period 1982–2021.

4.3 Spatial consistency with global SM products

580   To further investigate the spatial consistency between the GLASS-AVHRR and GLASS-MODIS SM products, as well as with two widely used long-term global SM products, mean SM maps of the GLASS-AVHRR, GLASS-MODIS, ESA CCI, and ERA5-Land products were plotted for January and July of 2016, respectively (Fig. 10). It can be seen that the GLASS-AVHRR SM product had the most complete spatial coverage among these products, after masking out permanent snow and ice and water bodies (Fig. 10g-h). Despite the spatiotemporal continuity of the ERA5-Land reanalysis SM product, it yielded

585 negative SM values close to zero in parts of the northern Africa, especially in July, which were masked out here (Fig. 10c-d). The ESA CCI combined SM product exhibited substantial spatial gaps above 30° N in January, in addition to the persistent absence of valid estimates in some densely vegetated regions (e.g., the Congo River and Amazon River basins), due to the attenuation of microwave signals in these areas (Fig. 10a-b) (Dorigo et al., 2017). Meanwhile, because of the lack of GLASS-MODIS albedo products at high latitudes during the cold season, GLASS-MODIS SM estimates were unavailable at

590 high latitudes (above 60° N) in January (Fig. 10e). Nevertheless, this does not affect the complete spatial coverage of the GLASS-AVHRR SM product, although it should still be used with caution in areas covered by seasonal snow and ice. In this regard, the performance of the GLASS-AVHRR SM product during the winter season (December–February) was evaluated using ISMN stations located above 30°N latitude, retaining only those with more than 100 matched records. The product achieved a median R value of 0.69 at 374 representative ISMN stations during Period I (2000–2018) and 0.63 at 19 stations

595 during Period II (1982–1999). Therefore, despite its relatively lower accuracy in winter, the GLASS-AVHRR SM product still can provide valuable estimates and serve as a useful complement to the ESA CCI SM product.

    In terms of the spatial distribution patterns of SM, the GLASS-AVHRR and GLASS-MODIS SM products showed a high degree of consistency, which further demonstrates the effectiveness of the developed DL model. In general, both GLASS SM products were slightly drier than the ESA CCI combined SM product, probably because optical and thermal satellite SM

600 estimates typically represent a shallower depth compared to microwave SM products. In contrast, the ERA5-Land SM product was much wetter than the other three SM products, especially in regions with high SM levels. While the three satellite SM products generally ranged between 0 and 0.5 m^3 m^{-3}, the ERA5-Land reanalysis SM product showed a range of 0–0.7 m^3 m^{-3}, indicating a clear tendency for overestimation. Although varying degrees of biases existed among the four global SM products, similar spatial patterns could be observed in all of them, characterized by higher SM values in the

605 eastern United States, northern South America, central Africa, and southern Asia, and lower SM values in the western USA, Middle East, northern and southern Africa, and Australia. Moreover, July was slightly drier than January in all four SM products, particularly in regions such as the western USA, eastern South America, and central Asia.

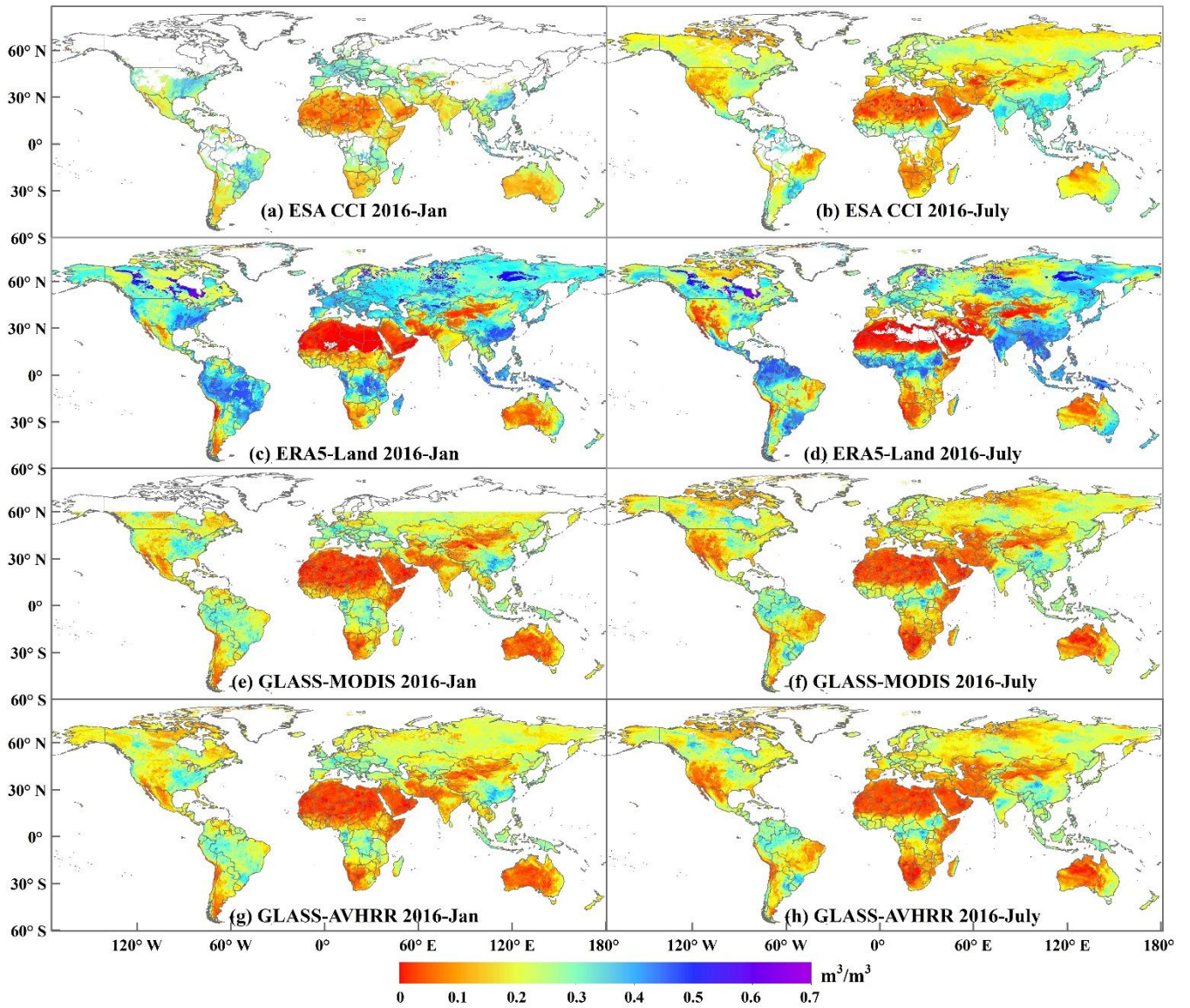



Figure 10 Mean global SM maps of the (a–b) 0.25° ESA CCI combined, (c–d) 0.1° ERA5-Land, (e–f) 1 km GLASS-MODIS, and (g–h) 5 km GLASS-AVHRR SM product in January and July of 2016. Permanent snow and ice as well as water bodies have been masked out using the MODIS land cover product (MCD12C1), while SM values in northern high latitudes in January should be interpreted with caution due to the widespread presence of permafrost, snow, and ice.

Figure 11 presents a zoomed-in comparison between the four SM products across the Tibetan Plateau in July 2016. The Tibetan Plateau, located in Central Asia, is the highest and most extensive plateau in the world, with an average elevation exceeding 4,000 meters. Its climate is extreme and varied, featuring significant seasonal and interannual variations. The unique topographic and climatic characteristics of the Tibetan Plateau make it one of the hotspots for global climate change

research. As can be observed from Fig. 11, all of the SM products show similar spatial distribution patterns: lower SM levels in the western and northern parts of the plateau, where rainfall is scarce and vegetation is sparse, and higher SM levels in the eastern and southern regions, where rainfall is more abundant and vegetation is denser. The GLASS-AVHRR SM product also exhibited high spatial consistency with the GLASS-MODIS SM product over the Tibetan Plateau, indicating that the adopted DL model effectively learned spatial features from the target SM product without introducing significant biases. Compared to the other three products, the ERA5-land SM product was much wetter in the southern part of the plateau, and the large positive bias in the ERA5-land reanalysis SM over the Tibetan Plateau was also reported in a previous study (Xing et al., 2021). Notably, there were many small patches with abrupt SM changes in the ERA5-land product (Fig. 11c), which were markedly improved in both the GLASS-AVHRR and GLASS-MODIS SM products. Moreover, compared to the ERA5-land and ESA CCI SM products at coarser resolutions, the GLASS-AVHRR SM product contained much richer spatial details and could well capture the distribution patterns of topography and vegetation.

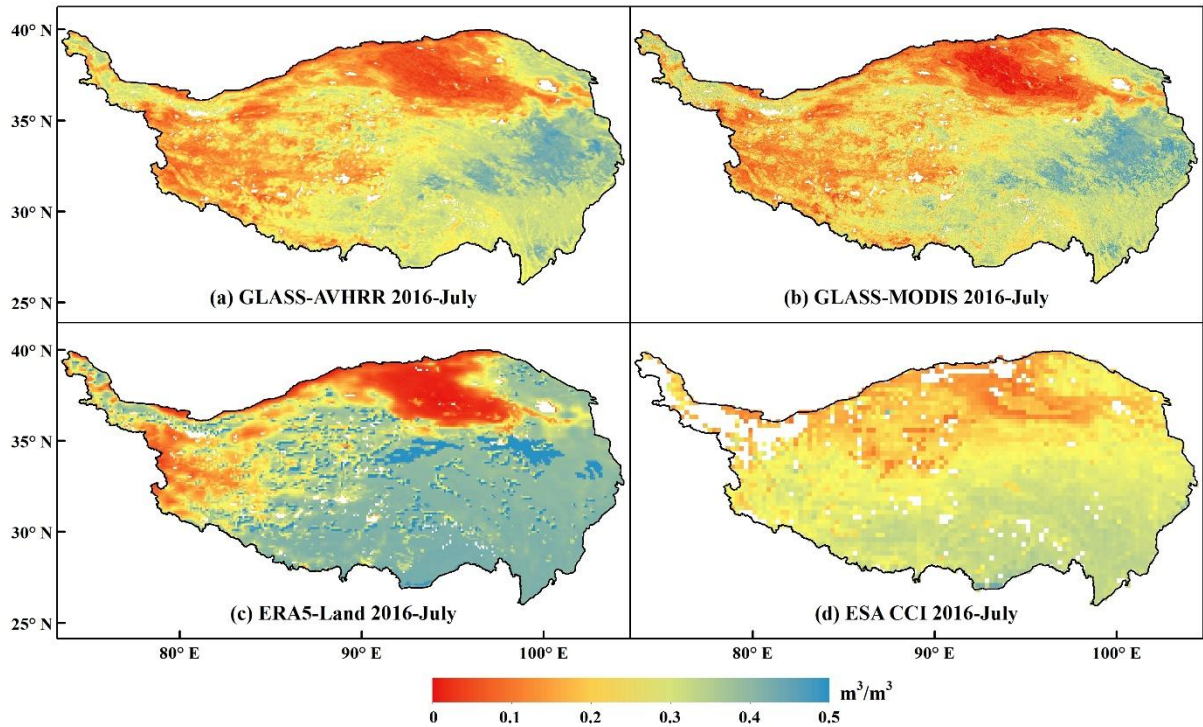



Figure 11 Zoomed-in comparison of the (a) 5 km GLASS-AVHRR, (b) 1 km GLASS-MODIS, (c) 0.1° ERA5-Land, and (d) 0.25° ESA CCI combined SM products across the Tibetan Plateau in July 2016. Permanent snow and ice as well as water bodies have been masked out using the MODIS land cover product (MCD12C1).

5 Discussion

This study aimed to develop a long-term global SM estimation framework using DL models, so as to derive a temporally consistent SM product with reliable accuracy over the last four decades. Therefore, we mainly explored two types of widely used DL models that are adept at processing sequential data: the LSTM-based models and Transformer. While LSTM has been utilized to retrieve SM since 2017 (Fang et al., 2017), the state-of-the-art Transformer model is still rarely used for SM estimation. Specifically, the accuracy of these DL models was compared from multiple perspectives, such as comparisons between the DL models and benchmark tree-based XGBoost model, between models with different attention mechanisms or different application architectures. Results showed that the Attention-based LSTM (AtLSTM) model achieved the best performance on the test set, and the MTM architecture could output a sequence of SM estimates simultaneously while maintaining similar accuracy to that of the MTO architecture. Note that, Transformer was reported to outperform LSTM-based models in several hydrological applications due to its ability to better handle long sequences and relate any two positions in the sequence (Amanambu et al., 2022; Yin et al., 2022). Meanwhile, according to Xu et al. (2021), Transformer achieved similar accuracy to the AtLSTM model in multi-temporal crop mapping tasks. However, Zeng et al. (2023) found that a simple linear model can outperform Transformer in long-term time-series forecasting tasks, and ascribed this to the temporal information loss associated with the self-attention mechanism. Therefore, the superiority of Transformer for time-series forecasting or estimating remains a topic of ongoing debate (Amanambu et al., 2022; Xu et al., 2021; Yin et al., 2022; Zeng et al., 2023). In our study, the accuracy of the Transformer model was slightly lower than that of the AtLSTM model, particularly for samples with high SM levels (> 0.4 m^3 m^{-3}). Given the high temporal variability of SM and the relatively short temporal length of SM memory, which typically ranges from 5 to 40 days and diminishes with increasing time lags (Orth and Seneviratne, 2012), this result may be attributed to the superior ability of the AtLSTM model to capture short-term adjacent dependencies. Yet, additional experiments with diverse training datasets are necessary to confirm the general applicability of this result.

We also investigated the effect of input sequence length on model accuracy, and it was found that the overall accuracy of the AtLSTM model with the MTM architecture leveled off at a sequence length of about 4 days. Subsequent analysis of the distribution of attention weights indicated that the model could automatically learn the necessary temporal information from adjacent positions in the sequence to accurately estimate SM. Despite that the overall accuracy of the LSTM-based models with the MTM architecture would converge as long as the length of the input sequence is sufficiently long, the models' accuracy is typically lower at the beginning or end of the sequence, and the affected estimates need to be identified and removed. In contrast, most of the current LSTM or Transformer application architecture is MTO and their accuracy remains unaffected at both ends of the sequence. But it is still necessary to identify the optimal sequence length during the training process to improve model efficiency, as the amount of input data would increase substantially with increasing sequence length. Here, we mainly explored the ability of the LSTM-based models and Transformer to capture temporal information from time-series input datasets for SM estimation. Future research could consider incorporating spatial patterns by

combining the AtLSTM or Transformer models with CNNs, or adapting the network of Transformer to improve its applicability for time-series estimating tasks. Moreover, different input features and data sources can also be integrated to investigate whether the estimation accuracy of SM can be further improved.

To examine the accuracy and consistency of the generated four-decade global daily GLASS-AVHRR SM product, different strategies were combined to fully evaluate it, including the validation against in situ SM datasets from point-scale ISMN stations, field-scale COSMOS networks, and upscaled 9 km SMAP CVSs, respectively, as well as the intercomparison with two widely used long-term global SM products. However, the evaluation of the GLASS-AVHRR SM product is still subject to certain limitations. The ISMN in situ SM dataset prior to 2000 is relatively scarce, with only 45 independent stations available for evaluation during this period, and large scale difference exists between this point-scale SM dataset and the 5 km GLASS-AVHRR SM product. The COSMOS sensors generally have varying footprint radii and sensing depths, and their measurements tend to exhibit higher uncertainties in organic soils or humid regions, which can lead to spatial and vertical representativeness issues. Additionally, there is only a limited number of upscaled SMAP CVSs, and the data collected may also contain errors caused by varying degrees of spatial representativeness.

Although validation results demonstrated that the GLASS-AVHRR SM product achieved high accuracy across different spatial scales, its performance was inevitably influenced by the GLASS-MODIS SM product, which served as the training target for the SM estimation model. Meanwhile, as a data-driven product, the quality of the GLASS-AVHRR SM product largely depends on the selected input features, their accuracy and consistency, and the representativeness of the training data. Potential uncertainties may arise from biases or errors in the satellite and reanalysis inputs. In particular, the reduced model accuracy observed in the high SM range is likely due to the inherent imbalance in the numerical distribution of SM samples and increased uncertainty in the accuracy of input features under wet surface conditions. In terms of feature selection, due to constraints such as record length, spatiotemporal completeness, and accuracy requirements, some informative but less consistently available variables may have been excluded, further contributing to the uncertainties in the final SM product. Moreover, as the ERA5-Land reanalysis SM was used as one of the input features, the generated product cannot be considered entirely independent. Future research could explore developing a fully independent, long-term, and seamless global SM product with sufficiently reliable accuracy.

Nevertheless, intercomparison with the long-term ERA5-Land and ESA CCI combined SM products showed that the derived GLASS-AVHRR SM product achieved the most complete spatial coverage, contained much richer spatial details, and remained unaffected by the large wet biases present in the input ERA5-Land SM product. While cumulative distribution function (CDF)-based methods can also be used for bias correction, they typically adjust statistical distributions locally, which limits their spatial generalization capability, particularly in regions lacking in situ SM data. In addition, they often overlook the temporal dependencies and nonlinear dynamics inherent in SM time series. Therefore, both the proposed DL-based SM estimation farmwork and the derived long-term global SM product present clear value. It should be noted that the ESA CCI combined SM product was generated by synthesizing SM products retrieved from multiple microwave sensors using different algorithms. This approach was necessary because no single microwave sensor covered the sufficiently long

700 time period (> 30 years) required for a climate data record, but it also inevitably led to variations in the product's accuracy over time and space (Dorigo et al., 2012). In contrast, the GLASS-AVHRR SM product was estimated using mainly the seamless GLASS-AVHRR albedo and LST products retrieved from the long-archived AVHRR satellite observations spanning four decades, which ensured its spatial and temporal completeness and consistency. Moreover, although microwave sensors are more sensitive to SM, their signal are significantly attenuated in densely vegetated areas, resulting in

705 persistent data gaps in the ESA CCI product. While the GLASS-AVHRR SM product is less accurate in these regions (with a median R of 0.57 at 20 COSMOS forest stations), it can provide a valuable complement to microwave SM products. In future research, greater efforts should be devoted to both the development and validation of long-term SM climate data records, and it is also crucial to assess the long-term trends in these SM datasets.

6 Data availability

710 The seamless global 5 km SM product (GLASS-AVHRR SM) at daily scale from 1982 to 2021 is freely accessible at https://glass.hku.hk/archive/SM/AVHRR/. Additionally, the annual average GLASS-AVHRR SM dataset was also generated, which can be downloaded from https://doi.org/10.5281/zenodo.14198201 (Zhang et al., 2024b). Note that this product represents the volumetric water content in the uppermost soil layer (0–5 cm), with areas of permanent snow and ice and water bodies masked. A scale factor of 1000 was applied, with missing values filled with -9999.

7 Conclusions

715

A four-decade (1982–2021) seamless global surface SM product (0–5 cm) at 5 km resolution was derived here, denoted as the GLASS-AVHRR SM product. This product was estimated using mainly the long-archived AVHRR satellite observations and multi-source datasets based on DL. Specifically, a large number of evenly distributed training samples extracted from the global 1 km daily GLASS-MODIS SM product were used as target to train three LSTM-based models (LSTM, Bi-LSTM,

720 and AtLSTM) and a Transformer model, with an XGBoost model employed as the benchmark. After identifying the AtLSTM as the best-performing model, it was ultimately adopted to generate the long-term GLASS-AVHRR SM product, which was then fully evaluated for reliability and consistency. The main results are summarized as follows:

(1) Evaluation of the models on the test set showed that all four DL models outperformed the benchmark XGBoost model, particularly at high SM levels (> 0.4 m^3 m^{-3}). Notably, the AtLSTM model achieved the best performance,

725 with an R^2 of 0.987 and RMSE of 0.011 m^3 m^{-3}, and its SM estimates were much closer to 1:1 line than those from other models. These results indicate that utilizing bidirectional temporal information from the input sequence as well as adding an attention module are both effective in improving the estimation accuracy of SM. Meanwhile, The MTM architecture adopted in this study achieved similar accuracy to that of the MTO architecture, while being able to output a sequence of SM estimates simultaneously and reduce considerable production time.

(2) The AtLSTM model with the MTM architecture was then employed to investigate the effect of input sequence length on model accuracy, and it was found that the overall accuracy of the model leveled off at a sequence length of about 4 days. Further analysis of attention weights revealed that the AtLSTM model with the MTM architecture could automatically learn the necessary information from adjacent positions in the sequence to accurately estimate SM at each position. In contrast, the temporal information learned by the self-attention module of the Transformer model was more dispersed distributed, and the slightly lower accuracy of the Transformer model than the AtLSTM model might be attributed to the typically high temporal variability of SM and that short-term adjacent temporal information played a more critical role in the accurate estimation of SM.

(3) The derived GLASS-AVHRR SM product was first evaluated using 45 independent point-scale ISMN stations prior to 2000, resulting in a median R of 0.73 and ubRMSE of 0.041 m^3 m^{-3}. Then, the product was validated against SM datasets from three post-2000 field-scale COSMOS networks, with median R values ranging from 0.63 to 0.79 and median ubRMSE values between 0.044 and 0.065 m^3 m^{-3}. Validation of the GLASS-AVHRR SM product at 22 upscaled 9 km SMAP CVSs yielded an overall R of 0.82 and ubRMSE of 0.054 m^3 m^{-3}. Whereas the ERA5-Land SM product had large wet biases at most of the CVSs, the GLASS-AVHRR SM product basically corrected these biases. Moreover, the time-series plots at six CVSs further demonstrated that the GLASS-AVHRR SM product could well capture the temporal variations in measured SM and showed a high degree of temporal consistency with the GLASS-MODIS SM product.

(4) Finally, the GLASS-AVHRR SM product was intercompared with two widely used long-term global SM products to investigate their spatial consistency. With the most complete spatial coverage, the GLASS-AVHRR SM product was slightly drier than the ESA CCI combined SM product, possibly due to the shallower depth it represents, whereas the ERA5-Land SM product exhibited a clear tendency for overestimation. While similar spatial patterns of SM could be observed in all of these products, the GLASS-AVHRR SM product contained much richer spatial details than the two long-term SM products at coarser resolutions.

Our study demonstrates the feasibility of utilizing the attention-based DL model and AVHRR satellite observations to generate long-term global SM product. The derived GLASS-AVHRR SM product has the advantages of long-term coverage, spatial and temporal integrity, reliable accuracy and consistency. As a reliable extension of the GLASS-MODIS SM product and a valuable complement to microwave SM products, this four-decade global SM product will be beneficial for a range of large-scale climate change-related research. Future studies could combine other DL models or integrate different data sources to further improve the quality of the long-term SM product.

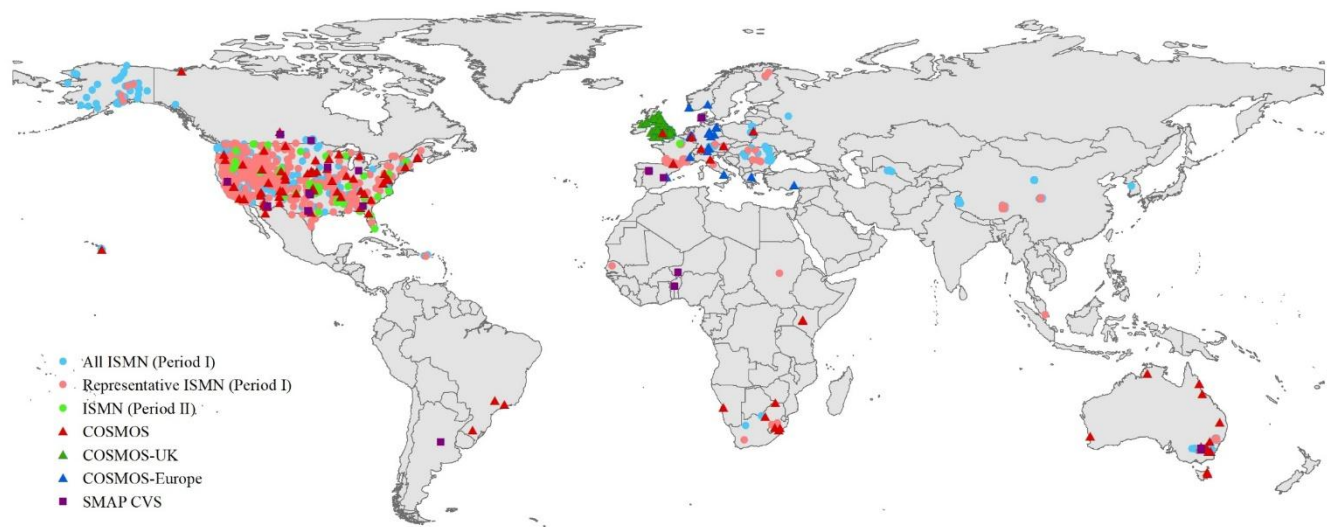
Appendix A: Supplementary figures

760

Figure A1 The spatial distribution of SM stations for each in situ SM dataset used in this study. Period I refers to 2000–2018 and Period II refers to 1982–1999

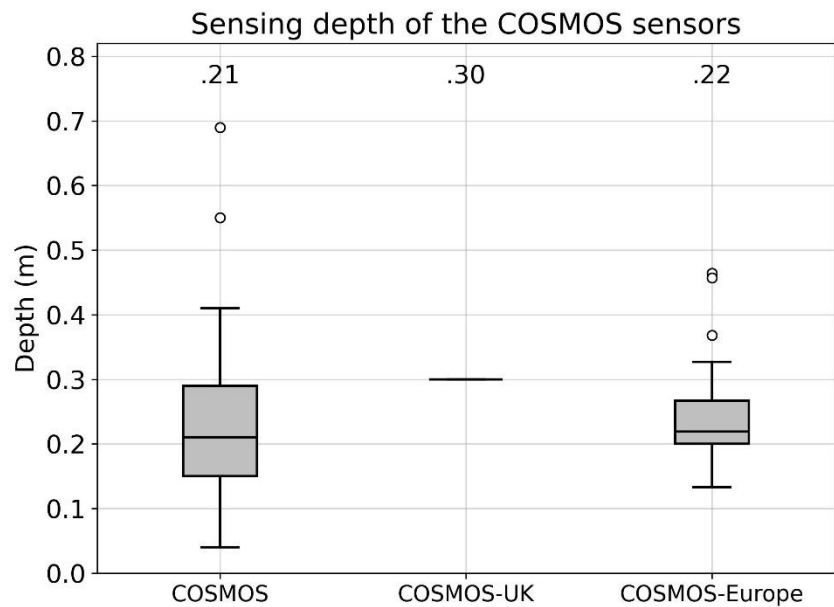



Figure A2 Boxplots of sensing depths across the three COSMOS networks used for validation

765

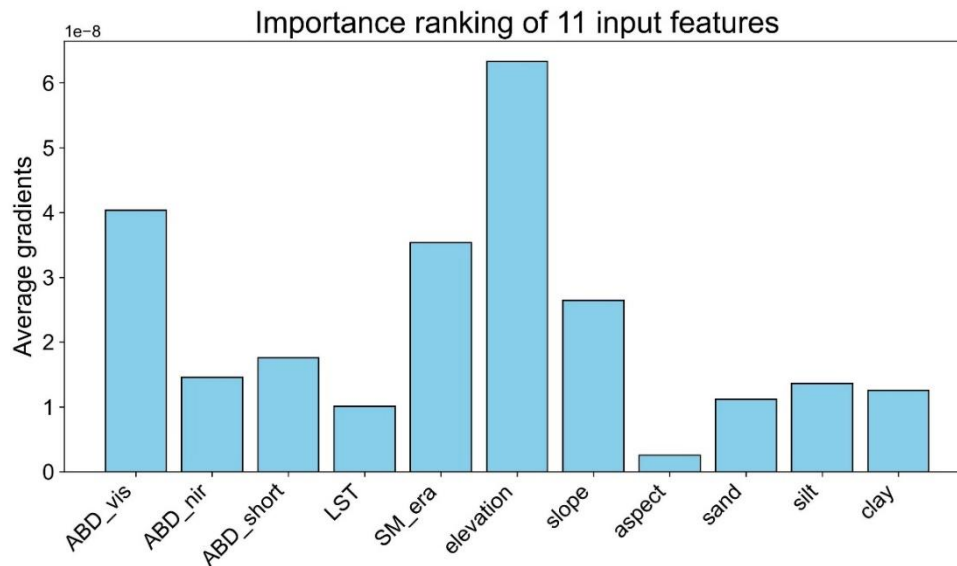



Figure A3 Importance ranking of 11 input features for the AtLSTM model based on gradients analysis

Author contributions

SL and YZ developed the methodology and designed the experiments. YZ, HM, JX, and GZ collected and processed the data. YZ carried out the experiments. TH and FT provided guidance on data analysis and experimental design refinements. YZ
770   prepared the manuscript with contributions from all co-authors.

Competing interests

Some authors are members of the editorial board of Earth System Science Data.

Acknowledgements

This study was supported by the Open Research Program of the International Research Center of Big Data for Sustainable
775   Development Goals (NO. CBAS2022ORP01), the National Key Research and Development Program of China (2023YFF1303702), the Fundamental Research Funds for the Central Universities, the National Key Research and Development Program of China (NO. 2016YFA0600103), and the National Natural Science Foundation of China (NO. 42090011). We also acknowledge the data support from "National Earth System Science Data Center, National Science & Technology Infrastructure of China (http://www.geodata.cn)".

References

Amanambu, A. C., Mossa, J., and Chen, Y.-H.: Hydrological Drought Forecasting Using a Deep Transformer Model, Water, 14, 3611, https://doi.org/10.3390/w14223611, 2022.

Bahdanau, D., Cho, K., and Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate, arXiv [preprint], arXiv:1409.0473, 2016.

Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J., and Anderson, C.: Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT), Geophys. Res. Lett., 34, L20401, https://doi.org/https://doi.org/10.1029/2007GL031088, 2007.

Bogena, H. R., Schrön, M., Jakobi, J., Ney, P., Zacharias, S., Andreasen, M., Baatz, R., Boorman, D., Duygu, M. B., Eguibar-Galán, M. A., Fersch, B., Franke, T., Geris, J., González Sanchis, M., Kerr, Y., Korf, T., Mengistu, Z., Mialon, A., Nasta, P., Nitychoruk, J., Pisinaras, V., Rasche, D., Rosolem, R., Said, H., Schattan, P., Zreda, M., Achleitner, S., Albentosa-Hernández, E., Akyürek, Z., Blume, T., del Campo, A., Canone, D., Dimitrova-Petrova, K., Evans, J. G., Ferraris, S., Frances, F., Gisolo, D., Güntner, A., Herrmann, F., Iwema, J., Jensen, K. H., Kunstmann, H., Lidón, A., Looms, M. C., Oswald, S., Panagopoulos, A., Patil, A., Power, D., Rebmann, C., Romano, N., Scheiffele, L., Seneviratne, S., Weltin, G., and Vereecken, H.: COSMOS-Europe: a European network of cosmic-ray neutron soil moisture sensors, Earth Syst. Sci. Data, 14, 1125–1151, https://doi.org/10.5194/essd-14-1125-2022, 2022.

Chan, S. K., Bindlish, R., O'Neill, P., Jackson, T., Njoku, E., Dunbar, S., Chaubell, J., Piepmeier, J., Yueh, S., Entekhabi, D., Colliander, A., Chen, F., Cosh, M. H., Caldwell, T., Walker, J., Berg, A., McNairn, H., Thibeault, M., Martínez-Fernández, J., Uldall, F., Seyfried, M., Bosch, D., Starks, P., Holifield Collins, C., Prueger, J., van der Velde, R., Asanuma, J., Palecki, M., Small, E. E., Zreda, M., Calvet, J., Crow, W. T., and Kerr, Y.: Development and assessment of the SMAP enhanced passive soil moisture product, Remote Sens. Environ., 204, 931–941, https://doi.org/https://doi.org/10.1016/j.rse.2017.08.025, 2018.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Cheng, F., Zhang, Z., Zhuang, H., Han, J., Luo, Y., Cao, J., Zhang, L., Zhang, J., Xu, J., and Tao, F.: ChinaCropSM1 km: a fine 1 km daily soil moisture dataset for dryland wheat and maize across China during 1993-2018, Earth Syst. Sci. Data, 15, 395–409, https://doi.org/10.5194/essd-15-395-2023, 2023.

Cheng, S., Guan, X., Huang, J., Ji, F., and Guo, R.: Long-term trend and variability of soil moisture over East Asia, J. Geophys. Res. Atmos., 120, 8658–8670, https://doi.org/https://doi.org/10.1002/2015JD023206, 2015.

Colliander, A., Asanuma, J., Berg, A., Bongiovanni, T., Bosch, D., Caldwell, T., Holifield-Collin, C., and Jensen, K.: SMAP/In Situ Core Validation Site Land Surface Parameters Match-Up Data (NSIDC-0712, Version 1), NASA National Snow and Ice Data Center Distributed Active Archive Center [data set], https://doi.org/10.5067/DXAVIXLY18KM, 2017.

Cooper, H. M., Bennett, E., Blake, J., Blyth, E., Boorman, D., Cooper, E., Evans, J., Fry, M., Jenkins, A., Morrison, R., Rylett, D., Stanley, S., Szczykulska, M., Trill, E., Antoniou, V., Askquith-Ellis, A., Ball, L., Brooks, M., Clarke, M. A., Cowan, N., Cumming, A., Farrand, P., Hitt, O., Lord, W., Scarlett, P., Swain, O., Thornton, J., Warwick, A., and Winterbourn, B.: COSMOS-UK: national soil moisture and hydrometeorology data for environmental science research, Earth Syst. Sci. Data, 13, 1737–1757, https://doi.org/10.5194/essd-13-1737-2021, 2021.

Das, N. N., Entekhabi, D., Dunbar, S., Kim, S., Yueh, S., Colliander, A., Jackson, T. J., O'Neill, P. E., Cosh, M., Caldwell, T., Walker, J., Berg, A., Rowlandson, T., Martínez-Fernández, J., González-Zamora, Á., Starks, P., Holifield-Collins, C., Prueger, J., and Lopez-Baeza, E.: Assessment Report for the L2_SM_SP Version-3 Release Data Products, SMAP Project, JPL D-56549, Jet Propulsion Laboratory, Pasadena, CA, 2020.

Dorigo, W., De Jeu, R., Chung, D., Parinussa, R., Liu, Y., Wagner, W., and Fernández-Prieto, D.: Evaluating global trends (1988-2010) in harmonized multi-satellite surface soil moisture, Geophys. Res. Lett., 39, L18405, https://doi.org/10.1029/2012GL052988, 2012.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, Remote Sens. Environ., 203, 185–215, https://doi.org/10.1016/J.RSE.2017.07.001, 2017.

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W., Xaver, A., Annor, F., Ardö, J., Baldocchi, D., Bitelli, M., Blöschl, G., Bogena, H., Brocca, L., Calvet, J.-C., Camarero, J. J., Capello, G., Choi, M., Cosh, M. C., van de Giesen, N., Hajdu, I., Ikonen, J., Jensen, K. H., Kanniah, K. D., de Kat, I., Kirchengast, G., Kumar Rai, P., Kyrouac, J., Larson, K., Liu, S., Loew, A., Moghaddam, M., Mart\'\inez Fernández, J., Mattar Bader, C., Morbidelli, R., Musial, J. P., Osenga, E., Palecki, M. A., Pellarin, T., Petropoulos, G. P., Pfeil, I., Powers, J., Robock, A., Rüdiger, C., Rummel, U., Strobel, M., Su, Z., Sullivan, R., Tagesson, T., Varlagin, A., Vreugdenhil, M., Walker, J., Wen, J., Wenger, F., Wigneron, J. P., Woods, M., Yang, K., Zeng, Y., Zhang, X., Zreda, M., Dietrich, S., Gruber, A., van Oevelen, P., Wagner, W., Scipal, K., Drusch, M., and Sabia, R.: The International Soil Moisture Network: serving Earth system science for over a decade, Hydrol. Earth Syst. Sci., 25, 5749–5804, https://doi.org/10.5194/hess-25-5749-2021, 2021.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, arXiv [preprint], arXiv:2010.11929v2, 2021.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Tsang, L., and Van Zyl, J.: The Soil Moisture Active Passive (SMAP) Mission, Proc. IEEE, 98, 704–716, https://doi.org/10.1109/JPROC.2010.2043918, 2010.

Fang, K. and Shen, C.: Near-Real-Time Forecast of Satellite-Based Soil Moisture Using Long Short-Term Memory with an Adaptive Data Integration Kernel, J. Hydrometeorol., 21, 399–413, https://doi.org/10.1175/JHM-D-19-0169.1, 2020.

Fang, K., Shen, C., Kifer, D., and Yang, X.: Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental U.S. Using a Deep Learning Neural Network, Geophys. Res. Lett., 44, 11030–11039, https://doi.org/https://doi.org/10.1002/2017GL075619, 2017.

Friedl, M. and Sulla-Menashe, D.: MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V061, NASA EOSDIS Land Processes Distributed Active Archive Center [data set], https://doi.org/10.5067/MODIS/MCD12C1.061, 2022.

Gao, L., Gao, Q., Zhang, H., Li, X., Chaubell, M. J., Ebtehaj, A., Shen, L., and Wigneron, J.-P.: A deep neural network based SMAP soil moisture product, Remote Sens. Environ., 277, 113059, https://doi.org/https://doi.org/10.1016/j.rse.2022.113059, 2022.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Clim., 30, 5419–5454, https://doi.org/https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gillies, R. R. and Carlson, T. N.: Thermal Remote Sensing of Surface Soil Water Content with Partial Vegetation Cover for Incorporation into Climate Models, J. Appl. Meteorol. Climatol., 34, 745–756, https://doi.org/https://doi.org/10.1175/1520-0450(1995)034<0745:TRSOSS>2.0.CO;2, 1995.

Grillakis, M. G.: Increase in severe and extreme soil moisture droughts for Europe under climate change, Sci. Total Environ., 660, 1245–1255, https://doi.org/https://doi.org/10.1016/j.scitotenv.2019.01.001, 2019.

Guevara, M., Taufer, M., and Vargas, R.: Gap-free global annual soil moisture: 15 km grids for 1991–2018, Earth Syst. Sci. Data, 13, 1711–1735, https://doi.org/10.5194/essd-13-1711-2021, 2021.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Comput., 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Hssaine, B. A., Merlin, O., Rafi, Z., Ezzahar, J., Jarlan, L., Khabba, S., and Er-Raki, S.: Calibrating an evapotranspiration model using radiometric surface temperature, vegetation cover fraction and near-surface soil moisture data, Agric. For. Meteorol., 256–257, 104–115, https://doi.org/https://doi.org/10.1016/j.agrformet.2018.02.033, 2018.

Huang, B., Zhao, B., and Song, Y.: Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery, Remote Sens. Environ., 214, 73–86, https://doi.org/https://doi.org/10.1016/j.rse.2018.04.050, 2018.

Jia, A.: Estimation and Spatiotemporal Analysis of All-Sky Land Surface Temperature from Multiple Satellite Data, Ph.D. thesis, University of Maryland, College Park, USA, 2023.

Jia, A., Liang, S., Wang, D., Mallick, K., Zhou, S., Hu, T., and Xu, S.: Advances in Methodology and Generation of All-Weather Land Surface Temperature Products From Polar-Orbiting and Geostationary Satellites: A comprehensive review,

880   IEEE Geosci. Remote Sens. Mag., 12, 218–260, https://doi.org/10.1109/MGRS.2024.3421268, 2024.

Kang, C. S., Zhao, T., Shi, J., Cosh, M. H., Chen, Y., Starks, P. J., Collins, C. H., Wu, S., Sun, R., and Zheng, J.: Global Soil Moisture Retrievals From the Chinese FY-3D Microwave Radiation Imager, IEEE Trans. Geosci. Remote Sens., 59, 4018–4032, https://doi.org/10.1109/TGRS.2020.3019408, 2021.

Karthikeyan, L. and Mishra, A. K.: Multi-layer high-resolution soil moisture estimation using machine learning over the

885   United States, Remote Sens. Environ., 266, 112706, https://doi.org/10.1016/J.RSE.2021.112706, 2021.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., and Delwart, S.: The SMOS Soil Moisture Retrieval Algorithm, IEEE Trans. Geosci. Remote Sens., 50, 1384–1403, https://doi.org/10.1109/TGRS.2012.2184548, 2012.

Köhli, M., Schrön, M., Zreda, M., Schmidt, U., Dietrich, P., and Zacharias, S.: Footprint characteristics revised for field-

890   scale soil moisture monitoring with cosmic-ray neutrons, Water Resour. Res., 51, 5772–5790, https://doi.org/https://doi.org/10.1002/2015WR017169, 2015.

Lal, P., Singh, G., Das, N. N., Colliander, A., and Entekhabi, D.: Assessment of ERA5-Land Volumetric Soil Water Layer Product Using In Situ and SMAP Soil Moisture Observations, IEEE Geosci. Remote Sens. Lett., 19, 1–5, https://doi.org/10.1109/LGRS.2022.3223985, 2022.

895   LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444, https://doi.org/10.1038/nature14539, 2015.

Li, Q., Zhu, Y., Shangguan, W., Wang, X., Li, L., and Yu, F.: An attention-aware LSTM model for soil moisture and soil temperature prediction, Geoderma, 409, 115651, https://doi.org/https://doi.org/10.1016/j.geoderma.2021.115651, 2022a.

Li, X., Wigneron, J.-P., Fan, L., Frappart, F., Yueh, S. H., Colliander, A., Ebtehaj, A., Gao, L., Fernandez-Moran, R., Liu, X., Wang, M., Ma, H., Moisy, C., and Ciais, P.: A new SMAP soil moisture and vegetation optical depth product (SMAP-IB):

900   Algorithm, assessment and inter-comparison, Remote Sens. Environ., 271, 112921, https://doi.org/https://doi.org/10.1016/j.rse.2022.112921, 2022b.

Li, X., Wigneron, J.-P., Frappart, F., Lannoy, G. De, Fan, L., Zhao, T., Gao, L., Tao, S., Ma, H., Peng, Z., Liu, X., Wang, H., Wang, M., Moisy, C., and Ciais, P.: The first global soil moisture and vegetation optical depth product retrieved from fused SMOS and SMAP L-band observations, Remote Sens. Environ., 282, 113272,

905   https://doi.org/https://doi.org/10.1016/j.rse.2022.113272, 2022c.

Liang, S. and Wang, J. (Eds.): Advanced Remote Sensing, Second Edi., Academic Press, 685–711 pp., https://doi.org/https://doi.org/10.1016/B978-0-12-815826-5.00018-0, 2020.

Liang, S., Cheng, J., Jia, K., Jiang, B., Liu, Q., Xiao, Z., Yao, Y., Yuan, W., Zhang, X., Zhao, X., and Zhou, J.: The Global Land Surface Satellite (GLASS) Product Suite, Bull. Am. Meteorol. Soc., 102, E323–E337, https://doi.org/10.1175/BAMS-

910   D-18-0341.1, 2021.

Ling, X., Huang, Y., Guo, W., Wang, Y., Chen, C., Qiu, B., Ge, J., Qin, K., Xue, Y., and Peng, J.: Comprehensive evaluation of satellite-based and reanalysis soil moisture products using in situ observations over China, Hydrol. Earth Syst. Sci., 25, 4209–4229, https://doi.org/10.5194/hess-25-4209-2021, 2021.

915 Liu, N., Liu, Q., Wang, L., Liang, S., Wen, J., Qu, Y., and Liu, S.: A statistics-based temporal filter algorithm to map spatiotemporally continuous shortwave albedo from MODIS data, Hydrol. Earth Syst. Sci., 17, 2121–2129, https://doi.org/10.5194/hess-17-2121-2013, 2013.

Ma, H. and Liang, S.: Development of the GLASS 250-m leaf area index product (version 6) from MODIS data using the bidirectional LSTM deep learning model, Remote Sens. Environ., 273, 112985, https://doi.org/10.1016/J.RSE.2022.112985, 2022.

920 Merlin, O., Rudiger, C., Al Bitar, A., Richaume, P., Walker, J. P., and Kerr, Y. H.: Disaggregation of SMOS Soil Moisture in Southeastern Australia, IEEE Trans. Geosci. Remote Sens., 50, 1556–1571, https://doi.org/10.1109/TGRS.2011.2175000, 2012.

Miralles, D. G., Bonte, O., Koppa, A., Baez-Villanueva, O. M., Tronquo, E., Zhong, F., Beck, H. E., Hulsman, P., Dorigo, W., Verhoest, N. E. C., and Haghdoost, S.: GLEAM4: global land evaporation and soil moisture dataset at 0.1° resolution 925 from 1980 to near present, Sci. Data, 12, 416, https://doi.org/10.1038/s41597-025-04610-y, 2025.

Montzka, C., Bogena, H. R., Zreda, M., Monerris, A., Morrison, R., Muddu, S., and Vereecken, H.: Validation of Spaceborne and Modelled Surface Soil Moisture Products with Cosmic-Ray Neutron Probes, Remote Sens., 9, https://doi.org/10.3390/rs9020103, 2017.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., 930 Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 2021, 1–50, https://doi.org/10.5194/essd-2021-82, 2021.

Orth, R. and Seneviratne, S. I.: Analysis of soil moisture memory from observations in Europe, J. Geophys. Res. Atmos., 117, D15115, https://doi.org/https://doi.org/10.1029/2011JD017366, 2012.

935 Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M. H., Crow, W. T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M. W. J., de Rosnay, P., Dorigo, W., Gruber, A., Hagemann, S., Hirschi, M., Kerr, Y. H., Lovergine, F., Mahecha, M. D., Marzahn, P., Mattia, F., Musial, J. P., Preuschmann, S., Reichle, R. H., Satalino, G., Silgram, M., van Bodegom, P. M., Verhoest, N. E. C., Wagner, W., Walker, J. P., Wegmüller, U., and Loew, A.: A roadmap for high-resolution satellite soil moisture applications – confronting product characteristics with user requirements, Remote Sens. 940 Environ., 252, 112162, https://doi.org/10.1016/J.RSE.2020.112162, 2021a.

Peng, J., Tanguy, M., Robinson, E. L., Pinnington, E., Evans, J., Ellis, R., Cooper, E., Hannaford, J., Blyth, E., and Dadson, S.: Estimation and evaluation of high-resolution soil moisture from merged model and Earth observation data in the Great Britain, Remote Sens. Environ., 264, 112610, https://doi.org/https://doi.org/10.1016/j.rse.2021.112610, 2021b.

Piles, M., Camps, A., Vall-Llossera, M., Corbella, I., Panciera, R., Rudiger, C., Kerr, Y. H., and Walker, J.: Downscaling

945   SMOS-derived soil moisture using MODIS visible/infrared data, IEEE Trans. Geosci. Remote Sens., 49, 3156–3166, https://doi.org/10.1109/TGRS.2011.2120615, 2011.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

950   Qu, Y., Liu, Q., Liang, S., Wang, L., Liu, N., and Liu, S.: Direct-Estimation Algorithm for Mapping Daily Land-Surface Broadband Albedo From MODIS Data, IEEE Trans. Geosci. Remote Sens., 52, 907–919, https://doi.org/10.1109/TGRS.2013.2245670, 2014.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The global land data assimilation system, Bull. Am.

955   Meteorol. Soc., 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Sabaghy, S., Walker, J. P., Renzullo, L. J., and Jackson, T. J.: Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities, Remote Sens. Environ., 209, 551–580, https://doi.org/https://doi.org/10.1016/j.rse.2018.02.065, 2018.

Schmugge, T., Gloersen, P., Wilheit, T., and Geiger, F.: Remote sensing of soil moisture with microwave radiometers, J.

960   Geophys. Res., 79, 317–323, https://doi.org/10.1029/JB079i002p00317, 1974.

Schoener, G. and Stone, M. C.: Impact of antecedent soil moisture on runoff from a semiarid catchment, J. Hydrol., 569, 627–636, https://doi.org/https://doi.org/10.1016/j.jhydrol.2018.12.025, 2019.

Sungmin, O. and Orth, R.: Global soil moisture data derived through machine learning trained with in-situ measurements, Sci. Data, 8, 170, https://doi.org/10.1038/s41597-021-00964-1, 2021.

965   Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I.: Attention is All you Need, in: Advances in Neural Information Processing Systems, 2017.

Wang, F., Tian, D., Lowe, L., Kalin, L., and Lehrter, J.: Deep Learning for Daily Precipitation and Temperature Downscaling, Water Resour. Res., 57, e2020WR029308, https://doi.org/https://doi.org/10.1029/2020WR029308, 2021.

Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., and Sun, L.: Transformers in Time Series: A Survey, arXiv

970   [preprint], arXiv:2202.07125v5, 2023.

Wigneron, J.-P., Li, X., Frappart, F., Fan, L., Al-Yaari, A., De Lannoy, G., Liu, X., Wang, M., Le Masson, E., and Moisy, C.: SMOS-IC data record of soil moisture and L-VOD: Historical development, applications and perspectives, Remote Sens. Environ., 254, 112238, https://doi.org/https://doi.org/10.1016/j.rse.2020.112238, 2021.

Xing, Z., Fan, L., Zhao, L., De Lannoy, G., Frappart, F., Peng, J., Li, X., Zeng, J., Al-Yaari, A., Yang, K., Zhao, T., Shi, J.,

975   Wang, M., Liu, X., Hu, G., Xiao, Y., Du, E., Li, R., Qiao, Y., Shi, J., Wen, J., Ma, M., and Wigneron, J. P.: A first assessment of satellite and reanalysis estimates of surface and root-zone soil moisture over the permafrost region of Qinghai-Tibet Plateau, Remote Sens. Environ., 265, 112666, https://doi.org/10.1016/J.RSE.2021.112666, 2021.

Xing, Z., Li, X., Fan, L., Colliander, A., Frappart, F., de Rosnay, P., Fernandez-Moran, R., Liu, X., Wang, H., Zhao, L., and Wigneron, J.-P.: Assessment of 9 km SMAP soil moisture: Evidence of narrowing the gap between satellite retrievals and

980   model-based reanalysis, Remote Sens. Environ., 296, 113721, https://doi.org/https://doi.org/10.1016/j.rse.2023.113721, 2023.

Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., Huang, J., Li, H., and Lin, T.: DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping, Remote Sens. Environ., 247, 111946, https://doi.org/https://doi.org/10.1016/j.rse.2020.111946, 2020.

Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K. C., Ying, Y., and Lin, T.: Towards interpreting multi-temporal deep

985   learning   models   in   crop   mapping,   Remote   Sens.   Environ.,   264,   112599, https://doi.org/https://doi.org/10.1016/j.rse.2021.112599, 2021.

Xu, M., Yao, N., Yang, H., Xu, J., Hu, A., Gustavo Goncalves de Goncalves, L., and Liu, G.: Downscaling SMAP soil moisture using a wide & deep learning method over the Continental United States, J. Hydrol., 609, 127784, https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.127784, 2022.

990   Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P.   D.:   A   high-accuracy   map   of   global   terrain   elevations,   Geophys.   Res.   Lett.,   44,   5844–5853, https://doi.org/10.1002/2017GL072874, 2017.

Yin, H., Guo, Z., Zhang, X., Chen, J., and Zhang, Y.: RR-Former: Rainfall-runoff modeling based on Transformer, J. Hydrol., 609, 127781, https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.127781, 2022.

995   Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., and Zhang, L.: Deep learning in environmental remote sensing: Achievements and challenges, Remote Sens. Environ., 241, 111716, https://doi.org/https://doi.org/10.1016/j.rse.2020.111716, 2020.

Zeng, A., Chen, M., Zhang, L., and Xu, Q.: Are Transformers Effective for Time Series Forecasting?, in: Proceedings of the AAAI Conference on Artificial Intelligence, 11121–11128, https://doi.org/10.1609/aaai.v37i9.26317, 2023.

1000   Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J.: Dive into Deep Learning, arXiv [preprint], arXiv:2106.11342, 2021.

Zhang, K., Chen, H., Ma, N., Shang, S., Wang, Y., Xu, Q., and Zhu, G.: A global dataset of terrestrial evapotranspiration and soil moisture dynamics from 1982 to 2020, Sci. Data, 11, 445, https://doi.org/10.1038/s41597-024-03271-7, 2024a.

Zhang, Q., Yuan, Q., Jin, T., Song, M., and Sun, F.: SGD-SM 2.0: an improved seamless global daily soil moisture long-term dataset from 2002 to 2022, Earth Syst. Sci. Data, 14, 4473–4488, https://doi.org/10.5194/essd-14-4473-2022, 2022.

1005   Zhang, Y., Liang, S., Ma, H., He, T., Wang, Q., Li, B., Xu, J., Zhang, G., Liu, X., and Xiong, C.: Generation of global 1 km daily soil moisture product from 2000 to 2020 using ensemble learning, Earth Syst. Sci. Data, 15, 2055–2079, https://doi.org/10.5194/essd-15-2055-2023, 2023.

Zhang, Y., Liang, S., Ma, H., He, T., Tian, F., Zhang, G., and Xu, J.: A seamless global 5 km surface soil moisture product from 1982 to 2021, Zenodo [data set], https://doi.org/10.5281/zenodo.14198201, 2024b.

1010 Zhao, H., Li, J., Yuan, Q., Lin, L., Yue, L., and Xu, H.: Downscaling of soil moisture products using deep learning: Comparison and analysis on Tibetan Plateau, J. Hydrol., 607, 127570, https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.127570, 2022.

Zheng, C., Jia, L., and Zhao, T.: A 21-year dataset (2000–2020) of gap-free global daily surface soil moisture at 1-km grid resolution, Sci. Data, 10, 139, https://doi.org/10.1038/s41597-023-01991-w, 2023.

1015 Zheng, J., Zhao, T., Lü, H., Shi, J., Cosh, M. H., Ji, D., Jiang, L., Cui, Q., Lu, H., Yang, K., Wigneron, J.-P., Li, X., Zhu, Y., Hu, L., Peng, Z., Zeng, Y., Wang, X., and Kang, C. S.: Assessment of 24 soil moisture datasets using a new in situ network in the Shandian River Basin of China, Remote Sens. Environ., 271, 112891, https://doi.org/https://doi.org/10.1016/j.rse.2022.112891, 2022.

Zheng, Y., Coxon, G., Woods, R., Power, D., Rico-Ramirez, M. A., McJannet, D., Rosolem, R., Li, J., and Feng, P.:
1020 Evaluation of reanalysis soil moisture products using cosmic ray neutron sensor observations across the globe, Hydrol. Earth Syst. Sci., 28, 1999–2022, https://doi.org/10.5194/hess-28-1999-2024, 2024.

Zhou, Y., Zhang, Y., Wang, R., Chen, H., Zhao, Q., Liu, B., Shao, Q., Cao, L., and Sun, S.: Deep learning for daily spatiotemporally continuity of satellite surface soil Moisture over eastern China in summer, J. Hydrol., 619, 129308, https://doi.org/https://doi.org/10.1016/j.jhydrol.2023.129308, 2023.

1025 Zhuo, W., Huang, J., Li, L., Zhang, X., Ma, H., Gao, X., Huang, H., Xu, B., and Xiao, X.: Assimilating Soil Moisture Retrieved from Sentinel-1 and Sentinel-2 Data into WOFOST Model to Improve Winter Wheat Yield Estimation, Remote Sens., 11, 1618, https://doi.org/10.3390/rs11131618, 2019.

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., and Rosolem, R.: COSMOS: the COsmic-ray Soil Moisture Observing System, Hydrol. Earth Syst. Sci., 16, 4079–4099, https://doi.org/10.5194/hess-16-4079-2012, 2012.
1030