

Progressive Scale Convolutional Network for Spatio-Temporal Downscaling of Soil Moisture: A Case Study Over the Tibetan Plateau

Ziyu Zhou^{a,1}, Keyan Hu^{b,1}, Ling Zhang^c, Zhaohui Xue^{a,d,*}, Yutian Fang^e and Yusha Zheng^e

^aCollege of Geography and Remote Sensing, Hohai University, Nanjing, 211100, China

^bSchool of Geosciences and Info-Physics, Central South University, Changsha, 410100, Hunan, China

^cSchool of Naval Architecture & Intelligent Manufacturing, Jiangsu Maritime Institute, Nanjing, 211100, China

^dKey Laboratory of Soil and Water Processes in Watershed, Hohai University, Nanjing, 211100, China

^eSchool of Earth Sciences and Engineering, Hohai University, Nanjing, 211100, China

ARTICLE INFO

Keywords:

Soil moisture
Downscaling
Deep learning
Tibetan Plateau
Multisource data

ABSTRACT

Soil moisture (SM) plays a critical role in hydrological and meteorological processes. High-resolution SM can be obtained by combining coarse passive microwave data with fine-scale auxiliary variables. However, the inversion of SM at the temporal scale is hindered by the incompleteness of surface auxiliary factors. To address this issue, first, we introduce validated high temporal resolution ERA5-Land variables into the downscaling process of the low-resolution SMAP SM product. Subsequently, we design a progressive scale convolutional network (PSCNet), at the core of which are two innovative components: a multi-frequency temporal fusion module (MFTF) for capturing temporal dynamics, and a bespoke squeeze-and-excitation (SE) block designed to preserve fine-grained spatial details. Using this approach, we obtained seamless SM products for the Tibetan Plateau (TP) from 2016 to 2018 at 10-km spatial and 3-hour temporal resolution. The experimental results on the TP demonstrated the following: 1) In the satellite product validation, the PSCNet exhibited comparable accuracy and lower error, with a mean R value of 0.881, outperforming other methods. 2) In the in-situ site validation, PSCNet consistently ranked among the top three models for the R metric across all sites, while also showing superior performance in overall error reduction. 3) In the temporal generalization validation, the feasibility of using high-temporal resolution ERA5-Land variables for downscaling was confirmed, as all methods maintained an average relative error within 6% for the R metric and 2% for the ubRMSE metric. 4) In the temporal dynamics and visualization validation, PSCNet demonstrated excellent temporal sensitivity and vivid spatial details. Overall, PSCNet provides a promising solution for spatio-temporal downscaling by effectively modeling the intricate spatio-temporal relationships in SM data.

1. Introduction

Soil moisture (SM) represents the volumetric water content in unsaturated soil (Seneviratne et al., 2010). It constitutes a fundamental element in hydrology and meteorology (Chahine, 1992; McColl et al., 2017), regulating evapotranspiration and impacting seepage, runoff, and groundwater flow (Akbar et al., 2020; Sadeghi et al., 2020). It is also a core variable in the terrestrial-atmospheric energy cycle dynamics (Dobriyal et al., 2012; Petropoulos et al., 2015), modulating the partitioning of available energy into sensible and latent heat fluxes at the Earth's surface (Sabaghy et al., 2018; Xia et al., 2014), thereby influencing the equilibrium of water, energy, and biogeochemical systems (Jung et al., 2017; Ochsner et al., 2013; Vereecken et al., 2022).

Reliable SM data can be obtained through three primary approaches: in-situ measurements, reanalysis datasets, and satellite remote sensing observations (Dorigo et al., 2011; Njoku and Entekhabi, 1996; Zhang et al., 2022). In-situ measurements of SM provide high local accuracy. However, the spatial variability of climatic conditions and significant heterogeneity of surface characteristics not only weaken the spatial representativeness of point-scale measurements across adjacent areas, but also pose significant maintenance challenges for long-term stable measurements (Collow et al., 2012; Loew, 2008; Njoku et al., 2003; Peng et al., 2017). Reanalysis datasets offer excellent spatio-temporal

*Corresponding author

✉ xzh2012@163.com (Z. Xue)

¹Ziyu Zhou and Keyan Hu are co-first authors.

continuity and spatial coverage, but have received less validation against in-situ measurements compared to remote sensing data. Given these limitations, microwave remote sensing techniques have become a reliable and promising approach for global SM retrieval due to their lower uncertainty and reduced bias compared to reanalysis products (Gao et al., 2022; Petropoulos et al., 2015; Zeng et al., 2015).

Remote sensing (RS) retrieves SM by analyzing electromagnetic radiation reflected or emitted from the land surface, utilizing physical or empirical relationships between sensor measurements and surface parameters (Gao et al., 2022; Zeng et al., 2015). Among various RS techniques, microwave sensors can penetrate through clouds and operate under all weather conditions, enabling continuous monitoring of SM within the top 5 cm of soil. Compared to optical sensors, passive microwave systems are less affected by surface roughness variations and offer more reliable temporal consistency for long-term observations. Several dedicated SM satellite missions have been deployed, including the advanced microwave scanning radiometer for earth observing system (AMSR-E), its successor AMSR2, the soil moisture and ocean salinity (SMOS) and soil moisture active passive (SMAP). These missions operate at different microwave frequencies and spatial resolutions, employing varying observation geometries to retrieve SM. However, microwave-derived SM products suffer from coarse spatial resolution (typically tens of kilometers) and multi-day revisit cycles (Abowarda et al., 2021), which constrain their applicability in domains requiring fine-scale data such as carbon cycle research, drought and flood monitoring, water resource management, precision agriculture, extreme climate monitoring, and ecosystem dynamics studies (Carbone et al., 1985; Fang et al., 2018; Li et al., 2021; Nagy et al., 2024; Peng et al., 2021; Piles et al., 2016).

To fill this gap, researchers have developed various approaches to downscale SM products. Current downscaling methods are broadly classified into three types: model-based approaches, active and passive microwave data fusion approaches, and optical/thermal and microwave data fusion approaches. Model-based approaches utilize statistical modeling and land surface process simulation to downscale SM data (Drusch et al., 2005; Ines and Droogers, 2002; Kaheil et al., 2008; Kim and Barros, 2002; Kroes et al., 2000; Mascaro et al., 2011; Peng et al., 2013; Shin and Mohanty, 2013; Verhoest et al., 2015). Active and passive microwave data fusion methods combine passive sensors' global observations with active sensors' fine spatial resolution. However, these methods are constrained by infrequent radar revisits and geometric registration challenges (Aiazzi et al., 2002; Montzka et al., 2016; Njoku et al., 2002; Wagner et al., 2008; Wigneron et al., 2003; Xing et al., 2025; Zhan et al., 2006). The third category, optical/thermal and microwave data fusion methods, enhance the spatial resolution of coarse microwave SM products by leveraging high-resolution optical and thermal observations. Empirical polynomial fitting downscaling methods based on the universal triangle theory have been developed (Carlson et al., 1994; Chauhan et al., 2003). Subsequent research has focused on developing soil and vegetation indices, exploring combinations of auxiliary factors, and optimizing algorithms (Merlin et al., 2010; Nasta et al., 2018; Piles et al., 2011; Zhao and Li, 2013). Additionally, researchers have developed methods based on evapotranspiration process decomposition (Malbêteau et al., 2016; Merlin et al., 2012; Molero et al., 2016), techniques for establishing soil indices and SM correlation patterns (Jiang and Islam, 2003; Kim and Hogue, 2011; Peng et al., 2015), and downscaling approaches utilizing artificial intelligence (AI) technology (Srivastava et al., 2013).

In recent years, AI-based methods have been widely applied to the downscaling and reconstruction of SM due to their capacity to capture complex spatio-temporal nonlinear dependencies between auxiliary variables and SM. Srivastava et al. (2013) pioneered the application of machine learning techniques to SM downscaling, utilizing data from the moderate resolution imaging spectroradiometer (MODIS).

- Within the domain of machine learning, random forest (RF) and support vector machines (SVMs) have been widely applied in SM downscaling and meteorological variable reconstruction (Abbaszadeh et al., 2019; Hutengs and Vohland, 2016; Liu et al., 2017; Im et al., 2016). Other machine learning techniques such as gradient boosting decision trees (GBDT) and eXtreme gradient boosting (XGBoost) have also shown competitive performance in pixel-level SM downscaling (Abowarda et al., 2021; Im et al., 2016; Lei et al., 2022; Liu et al., 2020, 2017; Wei et al., 2019; Zhao et al., 2018; He et al., 2025). These methods effectively capture complex nonlinear relationships between input features and target variables; however, they operate on individual pixels without considering spatial dependencies or temporal continuity. Since SM exhibits strong spatial autocorrelation and temporal persistence, these pixel-wise approaches frequently yield spatially discontinuous and temporally inconsistent results. Methods that incorporate spatio-temporal context are therefore essential for generating coherent predictions (Liu et al., 2018; Luo et al., 2016).
- Within the domain of deep learning, methods such as artificial neural networks (ANN), long short-term memory networks (LSTM), and convolutional neural networks (CNN) have demonstrated advantages in various aspects

such as modeling complex nonlinear relationships, capturing temporal dependencies, and extracting spatial features in SM downscaling (Fang et al., 2017; Orth et al., 2021; Rodriguez-Fernandez et al., 2015; Sobayo et al., 2018). These deep learning approaches demonstrate significant potential for advancing downscaling methodologies. For example, Kolassa et al. (2018) constructed an ANN to map global surface SM from SMAP observations, demonstrating strong data assimilation capabilities, and Gao et al. (2022) collected data from the international soil moisture network (ISMN) to train a deep neural network (DNN), achieving performance that surpasses SMAP products.

The Tibetan Plateau (TP) (Qiu, 2008) is a highly sensitive region to climate change (Yang et al., 2013), playing a pivotal role in shaping the climate evolution of the Asian monsoon and atmospheric circulation processes in the Northern Hemisphere through complex interactions among land, atmosphere, and ocean (Bothe et al., 2011; Jin et al., 2005; Xie et al., 2023). However, the widespread distribution of permanent and seasonal permafrost across the TP (Abowarda et al., 2021), coupled with harsh natural conditions, results in sparse ground observation stations that are difficult to maintain, leading to severe deficiencies in in-situ SM measurements for this region. Consequently, there is an urgent need to develop high-precision SM downscaling technologies. Therefore, Jiang et al. (2017) employed a back propagation neural network (BPNN) algorithm with five microwave SM products, demonstrating that BPNN outperforms traditional regression techniques and yields superior correlation coefficients (R) with in-situ SM data over the TP. Moreover, Wei et al. (2019) and Zhao et al. (2022) utilized the GBDT method in conjunction with 26 indices related to varying levels of vegetation water content or SM, and compared its performance with deep belief networks (DBN), ResNet, BPNN, and RF, successfully downscaling SMAP data on the TP from a 36 km to 1 km resolution. Alternatively, Shangguan et al. (2023) introduced the MATCH method, which enhances the downscaling performance of ANN, RF, ResNet, and LSTM through a combined machine learning approach for TP applications. More recently, Shangguan et al. (2024) constructed five machine learning downscaling models annually and minimized the error variance of SM data to merge the datasets, resulting in daily 1 km resolution SM data on the TP.

Constrained by the inadequate temporal resolution of satellite-derived surface auxiliary variables, most existing studies have focused on the spatial downscaling of SM. Reanalysis datasets generate hourly-resolution surface and near-surface parameters through data assimilation and numerical modeling. By utilizing validated surface auxiliary variables from reanalysis data to compensate for the temporal resolution limitations of satellite-derived surface auxiliary variables, this strategy provides the data foundation for multi-scale datasets for spatio-temporal downscaling. However, the integration of multi-source data with heterogeneous spatio-temporal characteristics necessitates specialized network architectures. To address this challenge, we first propose the progressive scale convolutional network (PSCNet) that systematically handles the trade-off between spatial scales and feature dimensions. We subsequently introduce a temporal convolutional module to enhance the network's temporal representation capability. Finally, through validation with original microwave satellite data, in-situ measurements, temporal generalization testing, and temporal dynamics visualization, our method demonstrates competitive results compared to existing approaches. The contributions of our research can be summarized as follows:

- We pioneered the integration of validated high spatio-temporal resolution reanalysis variables for spatio-temporal downscaling on the TP, achieving results at 10-km and 3-hour resolution. Furthermore, we demonstrated the feasibility and effectiveness of utilizing multi-temporal auxiliary variables within this framework.
- We conducted a systematic comparative analysis of spatially explicit models versus point-based models for SM reconstruction through multi-perspective validation including in-situ site validation, satellite data validation, inter-model validation, and visual assessment.
- We developed a separable spatio-temporal convolutional framework and introduced PSCNet, specifically designed for downscaling tasks with end-to-end spatio-temporal modeling capability that supports arbitrary input sizes, ensuring spatial coherence and eliminating tile boundary artifacts.

2. Study Area and Data

2.1. Study Area

The TP is the highest and largest plateau in the world, covering an area of about 3.08×10^6 square kilometers, with an average elevation of about 4320 m. It is situated between latitudes 26° and 40° north and longitudes 67° and 104°

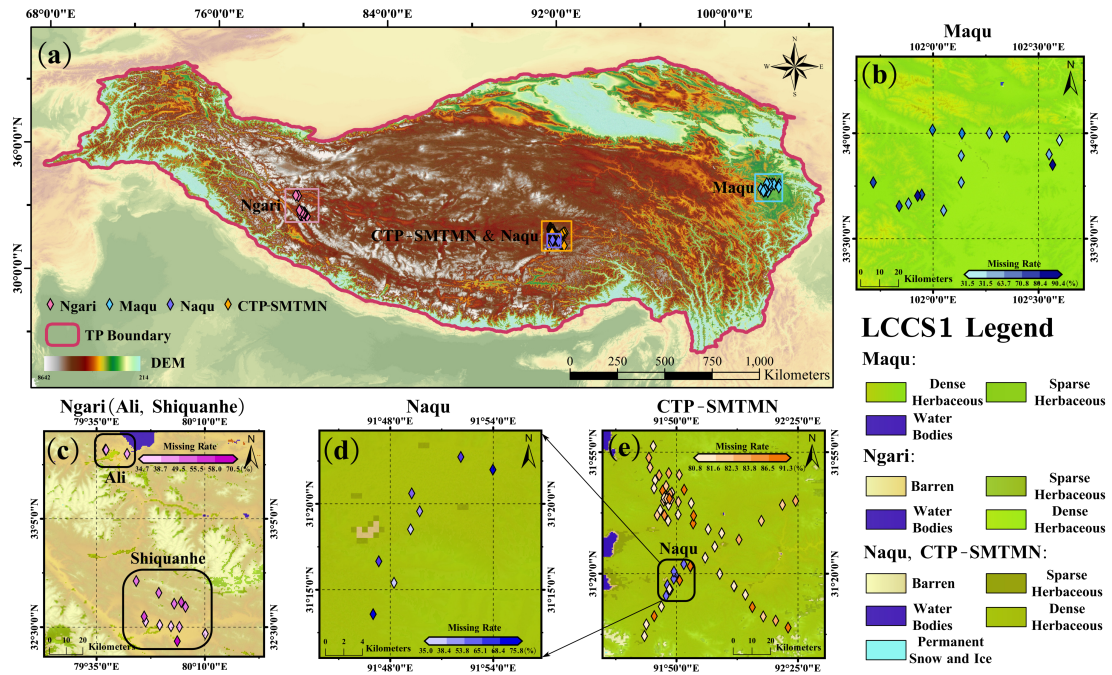


Figure 1: Elevation distribution and observation networks over the TP. (a) Overview showing elevation and four networks. (b-e) Site distribution with land cover types for Maqu, Ngari, Naqu, and CTP-SMTMN (abbreviated as CTP) networks, respectively. Symbol colors denote data missing rates. Background combines land cover types (30% transparency) with DEM topography. Legend describes FAO land cover classification system 1 (LCCS1) types within each network.

east. TP lies at the convergence of Asian mountain ranges, bounded by the Kunlun-Qilian Mountains in the north, the Himalayas in the south, the Karakoram Range and the Pamir Plateau in the west, and the Hengduan Mountains in the east, serving as the birthplace of many major Asian rivers (Zhang et al., 2021). It is also a region with complex climatic conditions and an uneven distribution of vegetation. It's permafrost and seasonally frozen ground cover more than 40% and 56%, respectively (Zou et al., 2017), while SM is highly heterogeneous due to its complex topography and diverse land cover.

2.2. Data

Surface radiation, meteorological factors, topography, vegetation, and soil characteristics are considered important SM inversion (Crow et al., 2012; Peng et al., 2016; Zhao and Li, 2013). Therefore, we synthesized multiple satellite products, including SMAP L3 SM product (SPL3SMP, Version 8), MODIS land surface parameters, and topographic data from Hydrological data and maps based on shuttle elevation derivatives at multiple scales (HydroSHEDS), combined with climate data from european centre for medium-range weather forecasts (ECMWF) distributing ERA5-Land hourly reanalysis products and gauge-adjusted precipitation data provided by the global satellite mapping of precipitation (GSMaP). Table 1 details the specific datasets used to construct the multi-temporal scale training dataset.

2.2.1. SMAP Soil Moisture Data

The SMAP satellite, launched by NASA on January 31, 2015, was designed to retrieve global surface SM and freeze/thaw state. It carries both an active L-band microwave radar and a passive radiometer. After the radar failed, the radiometer continues operations in a sun-synchronous orbit with equatorial crossings at 6:00 AM (descending) and 6:00 PM (ascending) local solar time, achieving global coverage every 2-3 days. The vast permanent and seasonal permafrost region on the TP presents challenges for L-band radiometer SM retrieval, resulting in extensive gaps in SPL3SMP across this region (Fig. 1). This study utilizes the SMAP L3 radiometer-derived product SPL3SMP (Version 8), which provides twice-daily global soil moisture data at a 36 km resolution on an EASE-Grid. During processing, we filtered data to remove observations flagged for frozen conditions (surface temperature less than 273.15 K), open water

Table 1

List of surface variables, data sources, and spatial and temporal resolutions used to construct the multi-source dataset.

Type	Variable name	Source	Temporal resolution	Spatial resolution
Interpolation variable	SM	SPL3SMP	Descending/Ascending Pass	36km
Auxiliary Dynamic Variables	HOY	—	Derived	—
	LST-Day	MYD11A1	Daily	1km
	LST-Night	MYD11A1	Daily	1km
	Albedo-BSA	MCD43C3	Daily	5km
	Albedo-WSA	MCD43C3	Daily	5km
	NDVI	MYD13A1	16 Day	500m
	EVI	MYD13A1	16 Day	500m
	Precipitation	GSMaP-Gauge	Hourly	10km
	Flux	ERA5-Land	Hourly	10km
	Snowmelt	ERA5-Land	Hourly	10km
	Runoff	ERA5-Land	Hourly	10km
	Surface radiation	ERA5-Land	Hourly	10km
	Evaporation	ERA5-Land	Hourly	10km
Auxiliary Static Variables	Longitude	—	Static	—
	Latitude	—	Static	—
	Clay	SoilGrids250m 2.0	Static	250m
	Sand	SoilGrids250m 2.0	Static	250m
	Silt	SoilGrids250m 2.0	Static	250m
	SOC	SoilGrids250m 2.0	Static	250m
	DEM	HydroSHEDS	Static	30m
	Slope	HydroSHEDS	Static	30m
	Aspect	HydroSHEDS	Static	30m
Land Cover	LC-Prop1:LCCS1	MCD12Q1	Yearly	500m
	K-GCC	—	Static	1km

HOY: Hour of Year, LST: Land Surface Temperature, NDVI: Normalized Difference Vegetation Index, EVI: Enhanced Vegetation Index, K-GCC: Köppen-Geiger Climate Classification, DEM: Digital Elevation Model, SOC: Soil Organic Carbon Content.

coverage >10% (Lei et al., 2022), and values failing the valid minimum threshold (< 0.02). Data were downloaded following reprojection to a geographic coordinate system using NASA Earthdata Search's reprojection tool.

2.2.2. MODIS Auxiliary Data

MODIS, aboard the Terra and Aqua satellites, is widely used for global environmental monitoring (Ustin and Middleton, 2021). This multispectral sensor covers the visible, near-infrared, and thermal infrared spectral bands. Its data for Collection 6 have completed the third phase of quality validation (Skakun et al., 2018). The surface variables employed during the experiment comprised LST, albedos, NDVI, and EVI, supplemented by the auxiliary LCCS1 variable for land cover classification analysis.

LST is derived from the global 1 km resolution MYD11A1 product from the Aqua satellite, utilizing both daytime and nighttime surface temperature bands. Albedos are calculated using the daily MCD43C3 product, which integrates 16 days of Terra and Aqua MODIS data and employs all bands from both white-sky and black-sky observations. NDVI and EVI, on the other hand, are derived from MYD13A1 data synthesized using the 16-day maximum value compositing (MVC) method at a 500 m spatial resolution. MCD12Q1 provides annual land cover types at a 500-meter resolution from 2001 to 2022. The data acquisition period for all of the aforementioned datasets spans from January 1, 2016, to January 1, 2019. All of these datasets are freely accessible on the Google Earth Engine.

2.2.3. ERA5-Land Auxiliary Data

The ERA5-Land dataset, released by ECMWF, represents advancements in surface physics modeling and multi-source data assimilation techniques, providing hourly data from 1950 to the present at a spatial resolution of $0.1^\circ \times 0.1^\circ$ with global land surface coverage. ERA5-Land is produced by running the ECMWF land surface model

Table 2

Summary of in-situ observation networks used for validation.

Network	Number of stations	Climate classification*	Data period (valid: Apr-Oct)	Missing data rate (3-hour intervals)
Naqu	9	Dwd: Cold, dry winter, very cold winter	2016.01.01-2018.10.31	59.6%
CTP	52	ET: Polar, tundra	2016.01.01-2016.09.20	83.6%
Ali	3	BWk: Arid, desert, cold	2016.01.01-2018.10.31	46.5%
Shiquanhe	12	BWk: Arid, desert, cold	2016.01.01-2018.10.31	47.5%
Maqu	15	ET and Dwd	2016.01.01-2018.10.30	66.9%

* Climate classification based on Köppen-Geiger Global 1 km climate classification maps (Beck et al., 2023).

(H-TESSEL) forced by downscaled meteorological fields from the ERA5 climate reanalysis, with elevation corrections for thermodynamic variables. The accuracy of ERA5-Land surface variables has been extensively validated across different regions (Crowhurst et al., 2021; Gualtieri, 2021; Hersbach, 2016; Hua et al., 2019; Zhang et al., 2022). For SM downscaling in this study, we selected ERA5-Land surface variables specifically validated over the TP region and frequently used in SM retrieval. These variables include surface temperature (1-7 cm), surface solar radiation, surface net radiant heat flux, surface latent heat flux, surface sensible heat flux, and total evaporation. (Dong et al., 2020; Li et al., 2023; Martens et al., 2020; Yilmaz, 2023; Zhang et al., 2020; Zheng et al., 2022).

2.2.4. Other Auxiliary Data

For comprehensive SM modeling, we incorporated additional precipitation and soil property datasets. Precipitation is a primary controlling factor for the abrupt changes in SM during short periods. High-temporal-resolution precipitation data enable accurate characterization of rapid SM variations (Dai et al., 2022; Wang et al., 2018; Yin et al., 2023). The GSMP-Gauge Precipitation dataset (Jiang et al., 2023) provides global data at 0.1° resolution at half-hourly intervals. Compared to other products in the GPM series, GSMP-Gauge exhibits the highest accuracy in mainland China (Kubota et al., 2020; Zhou et al., 2020). Soil texture (clay, sand, silt, SOC) provides a fundamental description of the soil's hydraulic conductivity and permeability (Nemes et al., 2005). Combined with topographic factors (DEM, slope, and aspect), these variables can serve as static features to construct a soil hydrodynamic model. Accordingly, we selected the global SoilGrids250m soil property dataset and the HydroSHEDS DEM product. SoilGrids250m offers global predictions of soil properties and classifications at a 250m resolution, while HydroSHEDS 3 arc-second enhances the global shuttle radar topography mission (SRTM) product with hydrological feature optimization and custom corrections such as void-filling, filtering, stream burning, and upscaling techniques (Lehner et al., 2008), which effectively address the notable radar shadow effects in mountainous and aquatic zones. We employ the HydroSHEDS 3 arc-second Void-filling DEM product and derive slope and aspect values through terrain analysis.

2.2.5. In-situ Measurements

The ISMN is a centralized global SM data system maintained by the Vienna University of Technology (Dorigo et al., 2017). This system provides hourly site-based observation data in UTC and has been widely validated (Ochsner et al., 2013). Our precision assessment relies on site data from five in-situ measurement networks on the TP: Naqu, Maqu, two Ngari sub-networks (Ali and Shiquanhe, treated separately due to their significant spatial separation, shown in Fig. 1), and CTP, as detailed in Table 2.

Systematic differences in measurement results from satellites and sparse observation networks are influenced by factors associated with different measurement modes, such as spatial scales, observational mechanisms, probing depths, and sensitivity to instantaneous rainfall (Colliander et al., 2017; Parinussa et al., 2011). Furthermore, these networks are located on seasonal permafrost (Zou et al., 2017), presenting significant challenges to data integrity and reliability. Therefore, to obtain robust site observation results, the following data processing rules have been established. 1) Select sites with the shallowest surface depth of less than 5cm. 2) Use ISMN data marked as "Good" (Dorigo et al., 2013). 3)

Only select data from the melt season between April 1 and November 1 from 2016 to 2019 for verification, excluding sites where the data missing rate exceeds 95%. 4) When more than two sites are located within the same grid point, select the most representative site to reduce the disparity of SM assessment weights (Dorigo et al., 2015; Zhang et al., 2019).

2.3. Data Processing

In this study, we integrate multi-source datasets with heterogeneous spatio-temporal resolutions to construct a temporally unified dataset that spans from 03:00 on January 1, 2016, to 21:00 on December 31, 2018. To meet the requirements of spatial downscaling, we further organize the data into two distinct datasets: a 36km training dataset (w/ SM) and a 10km inference dataset (w/o SM), both at 3-hour temporal intervals

For the general data processing workflow, initial processing of different data types involves cropping, stitching, quality control, and reprojection to unify coordinate systems to the WGS84 system. Subsequently, spatial resolutions are standardized using bilinear interpolation to resample the data onto regular latitude-longitude grids at resolutions of 36km and 10km to ensure alignment with the resolutions of SMAP SM data and ERA5-Land data, respectively. Finally, all auxiliary datasets were temporally resampled to a 3-hour resolution using interpolation methods. The SMAP SM data, serving as the training target, was kept at its native observation times of approximately 06:00 and 18:00 local solar time daily. To establish a uniform temporal framework for the analysis, these local solar time observations are subsequently referenced by the nominal 06:00 and 18:00 UTC timestamps. Additionally, we incorporate longitude and latitude as spatial context variables and HOY as a temporal context variable to enhance model performance.

For model training, we first apply Z-Score normalization to the 36km training dataset, then standardize the 10km inference dataset using the same statistical parameters. Training samples are extracted as 32x32 spatial patches with a temporal length of 5 and a stride of 10 pixels. The training data is partitioned into training (70%), validation (15%), and test (15%) sets. Data augmentation techniques, including flipping, transposing, and mirroring, are applied to the training samples.

For ground truth validation, ISMN site data is preprocessed and temporally aggregated to a 3-hour resolution.

3. Methodology

The downscaling framework and experimental procedure in this study are illustrated in Fig. 2, comprising four main modules: (a) Data preparation, which integrates multi-source datasets with heterogeneous spatio-temporal resolutions. (b) Data processing involves reprojecting, resampling, quality control, and data clipping steps. (c) The proposed PSCNet consists of the multi-frequency temporal fusion (MFTF) module and the squeeze-and-excitation (SE) block. (d) A multi-step validation of the downscaled products, where the models are divided into two categories: spatially explicit models (e.g., PSCNet, UNet) and point-based models (e.g., LSTM, RF). The comprehensive quality assessment is carried out through various perspectives, including in-situ site validation, satellite data validation, and inter-model validation, among others.

3.1. Problem Formulation

The SM downscaling process can be modeled as the learning and transfer of spatio-temporal mapping relationships using multi-source auxiliary variables. Given the original multi-source auxiliary dataset $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n\}$, where each $\mathcal{D}_k$ has unique spatio-temporal resolution, spatio-temporal alignment generates a unified low-resolution dataset $\mathcal{D}_L = \{X_L, Y_L\}$ (containing auxiliary variables X_L and SM ground truth Y_L) and a high-resolution dataset $\mathcal{D}_H = \{X_H\}$ (containing only auxiliary variables X_H). The extracted spatio-temporal data block $x^{(i)} \in \mathbb{R}^{T \times H \times W \times C}$ characterizes region i 's multi-variable state over a temporal sequence of length T , where (H, W) is the spatial scale and C is the variable dimension. In the low-resolution domain, we construct a network model f_θ to learn relationships between SM and auxiliary variables, where the parameter set $\theta = \{\theta_{\text{temp}}, \theta_{\text{spat}}\}$ models the spatio-temporal dimensions: The temporal encoder $f_{\theta_{\text{temp}}}$, implemented as our MFTF module, aggregates the time series into contextual features, then the spatial decoder $f_{\theta_{\text{spat}}}$ reconstructs these features into SM predictions:

$$\hat{Y}_L^{(i)} = f_{\theta_{\text{spat}}} \left(f_{\theta_{\text{temp}}} \left(\{x_{L,1}^{(i)}, \dots, x_{L,T}^{(i)}\} \right) \right) \in \mathbb{R}^{H_L \times W_L} \quad (1)$$

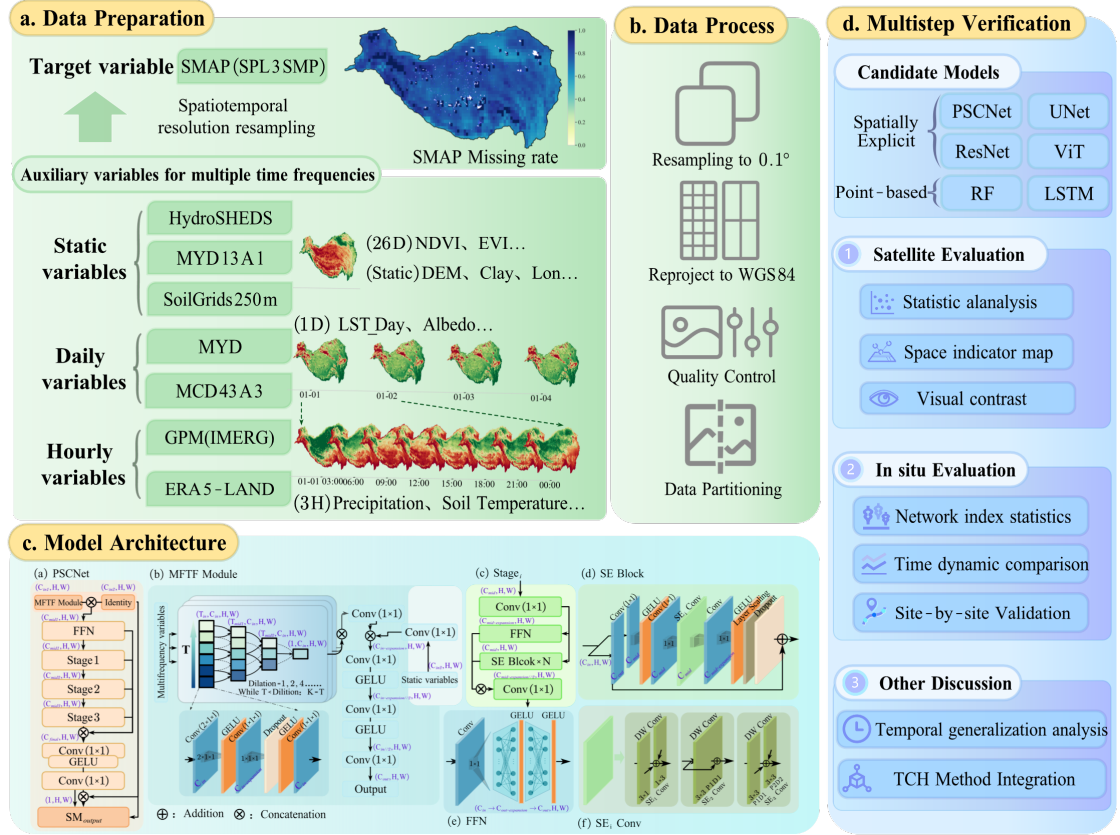


Figure 2: Overall framework and methodological workflow of SM spatial downscaling.

The model is optimized by minimizing the loss function:

$$\theta^* = \arg \min_{\theta} \sum_i \mathcal{L}(\hat{Y}_L^{(i)}, Y_L^{(i)}) \quad (2)$$

Finally, the trained model enables cross-scale transfer by applying the learned spatio-temporal mapping f_{θ^*} to the high-resolution auxiliary variables X_H to generate high-resolution SM:

$$\hat{Y}_H^{(i)} = f_{\theta^*}(X_H^{(i)}) \in \mathbb{R}^{H_H \times W_H} \quad (3)$$

3.2. Model Architecture

Achieving effective spatio-temporal mapping in practice requires careful consideration of network architecture design. The key challenge in downscaling network design is to preserve cross-scale mapping relationships while flexibly capturing inter-variable dependencies across dimensions. Moreover, robust spatio-temporal modeling is essential for ensuring the spatio-temporal consistency of the results, which in turn hinges on efficient information distillation and redundancy removal. To address these challenges, we examine the network architecture from several perspectives:

- **Maintaining Spatial Integrity without Downsampling.** The U-shaped architecture is widely used in modern CNN designs, where progressive downsampling enhances the model's ability to capture spatial hierarchies. However, even with skip connections, downsampling operations inevitably lead to the loss of fine-grained spatial details essential for high-resolution reconstruction. Additionally, patch-wise reconstruction can introduce boundary artifacts and spatial discontinuities. To address these issues, we adopt a convolutional network architecture that omits downsampling operations, thereby preserving the completeness of spatial information.

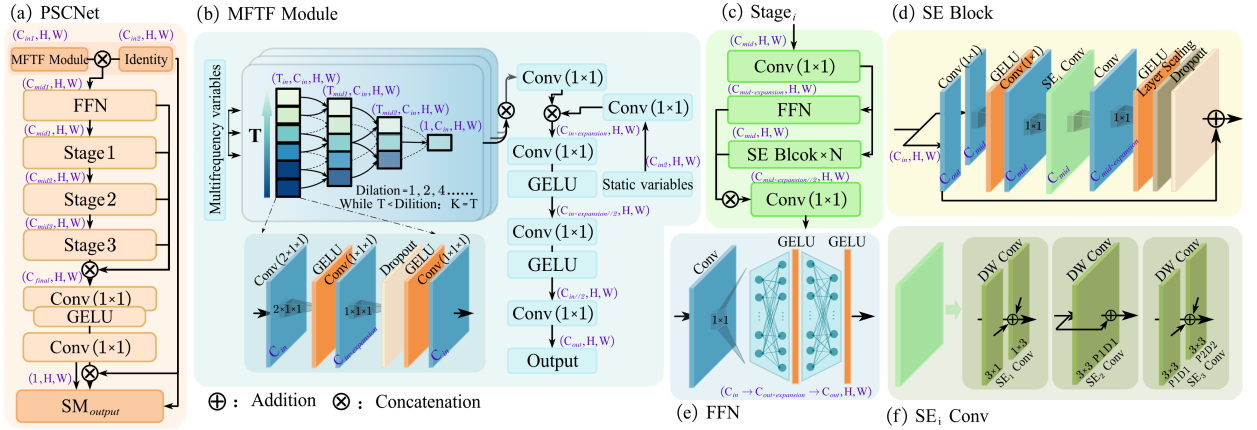


Figure 3: The architecture of the proposed PSCNet. The main figure (a) illustrates the overall structure, while (b-f) provide detailed views of its key components: the MFTF Module, the Stage block, the SE block, the feed-forward network (FFN), and the SE Convolutional layer.

- **Stabilizing Generalization without Batch Normalization.** In the downscaling process, data sampling demonstrates strong sequential dependency. Batch normalization (BN) may introduce variance shifts that result in numerical instability during inference (Li et al., 2019). To mitigate this issue, we remove BN layers from the network to enhance stability.
- **Enhancing Information Retention with Gaussian Error Linear Unit (GELU) and Inverted Residuals.** Activation functions play a crucial role in enabling ANNs to capture nonlinear relationships (Glorot and Bengio, 2010). ReLU activations can cause information loss through hard thresholding, particularly problematic when preserving subtle spatial patterns in downscaling tasks (Howard et al., 2017). To address this limitation, we replace all ReLU layers with GELU activations and incorporate inverted residual structures to maintain richer feature representations (Liu et al., 2022).
- **Reconstructing Inference Patterns with Hybrid Multi-Level Network Architectures.** Localized feature extraction demonstrates strong performance in downscaling tasks (Abbaszadeh et al., 2019; Kim et al., 2018; Zhuo and Han, 2017); however, spatio-temporal modeling with larger receptive fields yields superior spatial coherence. To combine the strengths of both approaches, we design a multi-stage hybrid architecture embodied in our Stage module (Fig. 3 (c)), which explicitly couples a FFN for localized inference with a convolutional branch built upon SE Blocks to extract hierarchical spatial features (Ding et al., 2024). Dilated and factorized convolutions further enlarge the receptive field efficiently while preserving model compactness.
- **Modeling spatio-temporal Dependencies.** Effective spatio-temporal downscaling hinges on the model's ability to capture complex, non-linear dependencies. Our MFTF module (Fig. 3 (b)) achieves this by first establishing an absolute spatio-temporal context using positional encodings (Hour of Year, longitude, and latitude). This context-aware data is then processed by a temporal convolutional network (TCN) (Bai et al., 2018), which serves as the core engine for dependency modeling. Its capacity is enhanced by internal feature enrichment layers, while no-padding convolutions are used to progressively distill the learned temporal relationships across various scales into a final, feature-rich spatial representation. If the length of the current time block, T_l , is smaller than the dilation length, D_l , a convolution kernel of length T_l is used to reduce the temporal dimension to 1.

Collectively, these design principles culminate in the PSCNet architecture, illustrated in Fig. 3.

3.3. Loss Function

Patch-based training enhances spatial coherence within patches but frequently introduces artifacts along patch boundaries. Therefore, we implement a multi-task loss function combining RMSE and structural similarity index measure (SSIM) to produce downscaled SM data better aligned with human visual perception.

RMSE quantifies pixel-level errors. To prioritize edge regions and mitigate mosaic artifacts during reconstruction, an edge-weighting kernel W_e is incorporated into the L_{RMSE} calculation. The kernel $W_e(i, j)$, and L_{RMSE} are defined as follows:

$$W_e(i, j) = 1 + (\text{ratio} - 1) \cdot \frac{2 \cdot \sqrt{\left(i - \frac{H-1}{2}\right)^2 + \left(j - \frac{W-1}{2}\right)^2}}{\sqrt{H^2 + W^2}} \quad (4)$$

$$L_{\text{RMSE}} = \sqrt{\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W W_e(i, j) \cdot (\hat{y}(i, j) - y(i, j))^2} \quad (5)$$

Here, $\hat{y}$ and y denote the predicted and true SM values, while H and W specify the height and width of the patch. The parameter ratio is the weighting coefficient.

SSIM assesses perceptual quality based on human vision and exhibits strong generalizability (Wang et al., 2004). Incorporating SSIM into a hybrid loss function effectively suppresses reconstruction artifacts and improves visual quality (Huang et al., 2020). The L_{SSIM} and composite loss L_{FULL} are defined as:

$$L_{\text{SSIM}} = 1 - \frac{(2\mu_{\hat{y}}\mu_y + C_1)(2\sigma_{\hat{y}y} + C_2)}{(\mu_{\hat{y}}^2 + \mu_y^2 + C_1)(\sigma_{\hat{y}}^2 + \sigma_y^2 + C_2)} \quad (6)$$

$$L_{\text{FULL}} = \alpha L_{\text{RMSE}} + (1 - \alpha) L_{\text{SSIM}} \quad (7)$$

where $\mu_{\hat{y}}$ and μ_y are the means of the predicted and true SM values, respectively, while $\sigma_{\hat{y}}^2$ and σ_y^2 are their variances, $\sigma_{\hat{y}y}$ is the covariance, and C_1 and C_2 are small stabilizing constants. The hyperparameter α controls the blending ratio. Statistics $(\mu, \sigma^2, \sigma_{xy})$ are computed over the model's output patches.

4. Experimental Results

The following sections outline the evaluation scenarios: 1) a downscaling quality assessment of SM inversion products based on SMAP microwave satellite data, 2) an assessment of the feasibility of reconstructing high-temporal resolution data using in-situ measurements, and 3) a visual assessment of downscaled products.

This study employs four commonly used evaluation metrics for SM downscaling (Liu et al., 2020; Qu et al., 2021), namely the R, bias, RMSE, and ubRMSE. The specific formulas for these metrics are provided in A.1.

4.1. Model Implementation and Configuration

The models used in this study are primarily divided into two categories: 1) spatially explicit models that leverage contextual spatial information, including models designed to capture spatial dependencies, such as ResNet, UNet, and Vision Transformer, as well as our proposed spatio-temporal model, PSCNet; and 2) point-based models, which treat each location independently, including temporal sequence regression models such as LSTM and pixel-wise methods like RF.

For the spatially explicit models, during training, we apply progressive masking to SM data, transitioning from dual-task learning (variable inference and missing data reconstruction) to pure variable inference to enhance spatial reasoning capabilities (Papamakarios et al., 2017; Perez and Wang, 2017). Furthermore, during inference, the PSCNet architecture is capable of processing inputs of arbitrary size, allowing us to employ a full-image inference mode to reinforce spatial coherence, while other models perform inference on 32×32 patches, consistent with their training configuration. For point-based models, LSTM uses a time window of 56 steps (corresponding to one week), with a hidden layer dimension of 512 and an initial learning rate of 0.0005. For RF, the hyperparameters are set to 300 decision trees, with a maximum depth of 10 for each tree and a minimum sample size of 10 and 6 for node splitting and leaf nodes, respectively. Finally, to ensure fair performance comparison in downscaling tasks, we align all models to comparable FLOPS or parameter counts.

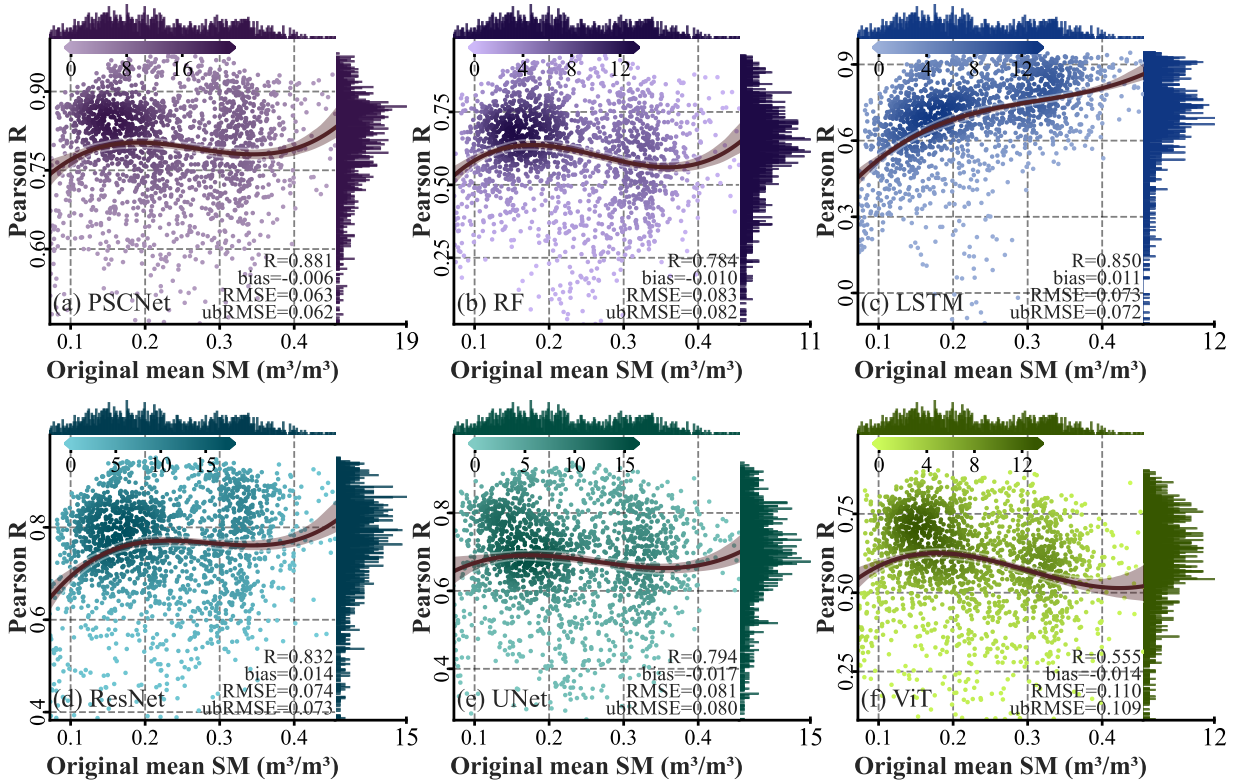


Figure 4: Scatter density plots showing R between downscaled products from ten downscaling methods and original SMAP microwave satellite observations. Overall statistical metrics are displayed in the lower right corner of each subplot. (a) PSCNet; (b) RF; (c) LSTM; (d) ResNet; (e) UNet; (f) ViT.

4.2. Model Evaluation Using SMAP Microwave Satellite Data

To evaluate the downscaled products, we first spatially aggregate them from the 10 km resolution back to the original 36 km SMAP grid. Although spatially aggregating downscaled products back to the original 36 km resolution introduces uncertainty, especially in regions with significant SM variability, this approach remains effective for assessing how well the downscaled products preserve original spatial patterns. During the training phase, the original 36 km SMAP microwave satellite data serve as the low-resolution training target, while in the inference phase, 10 km auxiliary variables are fed into the model to generate 10 km SM products. These products are then aggregated to match the spatial resolution of the original 36 km SMAP data. After addressing missing values, R values are computed at each spatial location. The data are subsequently flattened to calculate statistical metrics between the two datasets, and a scatter density plot is created between the upsampled products and the original SMAP data.

Fig. 4 shows correlation density scatter plots between downscaled products and original SMAP observations. The mean SM values cluster around $0.2 \text{ (m}^3/\text{m}^3\text{)}$ and $0.3 \text{ (m}^3/\text{m}^3\text{)}$, with largely consistent scatter patterns observed across most models in the downscaling task. Among these, the proposed PSCNet model achieves the best performance, with a Pearson R of 0.881, an RMSE of 0.063, and an ubRMSE of 0.062. Among the baseline models, the point-based RF delivers a performance with a Pearson R of 0.784, which is comparable to some models that leverage spatial context, although its overall accuracy is moderate. The LSTM model, in contrast, exhibits a unique performance pattern distinct from all other methods. It shows enhanced accuracy in regions with high SM averages but performs poorly where SM is low, a behavior likely related to its extended training sequence of 56 time steps creating a strong temporal dependency. For the other deep learning baselines designed to capture spatial dependencies, ResNet and UNet achieve the most competitive results with Pearson R values of 0.832 and 0.794 respectively, placing their performance closest to our proposed PSCNet model. Finally, ViT exhibits markedly inferior performance across all key evaluation metrics, exemplified by a low Pearson R of 0.555 and the highest error values among all models.

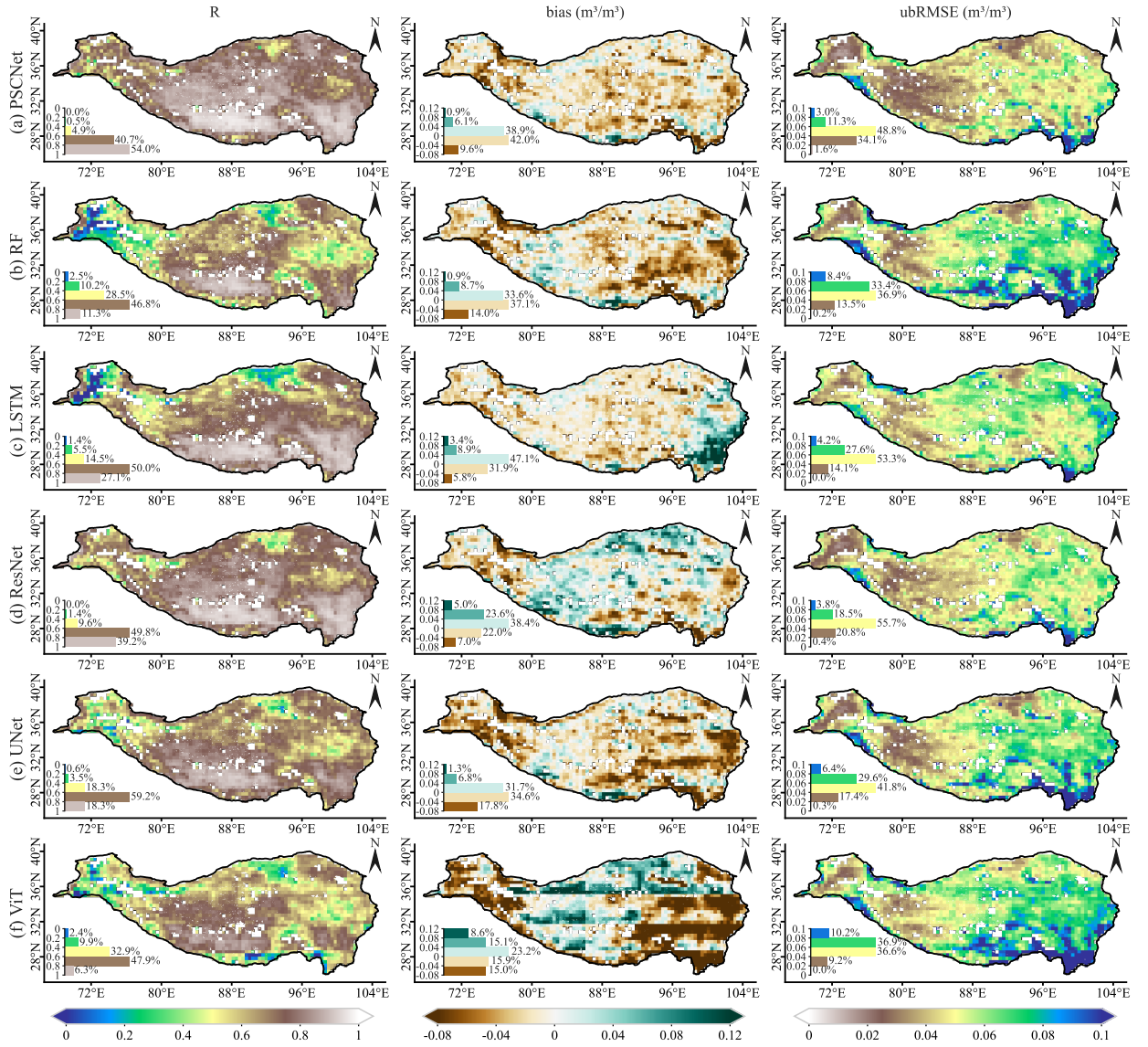


Figure 5: Spatial distribution maps of validation metrics (R, bias, and ubRMSE from left to right) for downscaled products from different models, with histograms showing pixel distribution within each metric range.

Fig. 5 displays the spatial distribution maps of validation metrics for the different downscaled products. The spatial patterns reveal that model performance correlates strongly with regional moisture conditions. This correlation is driven by the region's distinct geography; the Himalayas block moisture-laden South Asian monsoon flows, while the westerlies desiccate leeward regions after crossing the Iranian Plateau. These climatic drivers result in a pronounced southeast-to-northwest SM gradient across the TP, with humid conditions in the southeast giving way to extreme aridity in basins from the Karakoram to the Qaidam. Correspondingly, the spatial patterns of the validation metrics for all models mirror this moisture gradient, consistently showing higher accuracy in the moisture-abundant southeast and degraded performance in the arid northwest.

Specifically, regarding the R metric, all models exhibit weaker performance in the Pamir Plateau and the Qaidam Basin, with the point-based models being the most severely affected, showing significant correlation distortion near the Pamir Plateau. In contrast, PSCNet maintains the highest correlation in these challenging areas. Quantitatively, 94.6% of the spatial domain for PSCNet has an R value greater than 0.6, with over 54% of the region surpassing an R

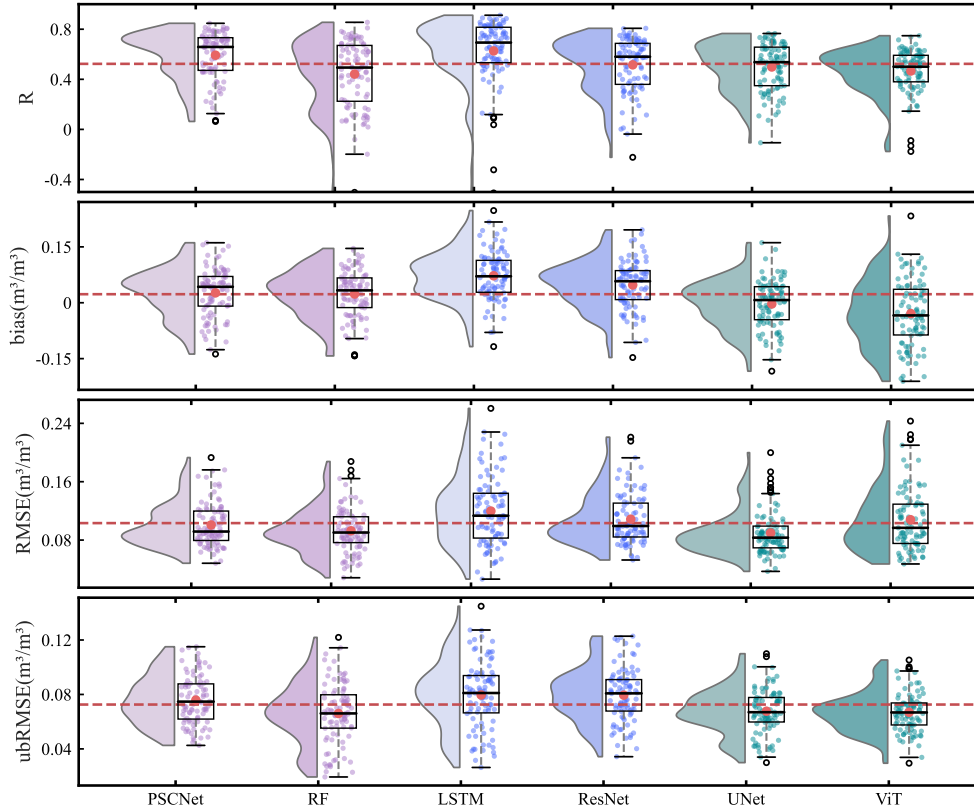


Figure 6: Cloud and rain plots of statistical metrics comparing the different downscaling models using in-situ site data (split violin plots with density distributions on the left and scatter points with box plots on the right, red dashed lines represent overall mean values, and red dots represent each method's mean performance in the dataset).

of 0.9. Regarding bias, a general overestimation is observed near the Tanggula and Kunlun Mountains, while a general underestimation is prevalent in the northern TP. Among the models, PSCNet and LSTM exhibit the smallest overall bias and show excellent performance. In contrast, RF and UNet generally show an overestimation, while ResNet shows a tendency for underestimation. Furthermore, the results from ViT suffer from severe spatial artifacts, representing the poorest performance. In terms of ubRMSE, all models show higher errors near the Hengduan Mountains. PSCNet achieves the best overall ubRMSE, with only 14.3% of its domain showing a value greater than 0.06. This is a considerable improvement over the next-best model, ResNet (22.3%), and is 17.9 percentage points lower than the average of all models (32.2%).

Overall, among all downscaled products, PSCNet demonstrates the best capability in preserving the features of the original microwave satellite data, achieving the top performance across all metrics in this spatial comparison. While higher SM in humid regions often implies greater variability, which poses a significant challenge, PSCNet effectively captures these complex dynamics, maintaining robust performance where other models falter.

4.3. Model Evaluation Using In-situ Soil Moisture Data

4.3.1. Comparative Evaluation of Models at Overall and Network Scales

To further evaluate the performance and generalization capability of the downscaling models, we directly compare their 3-hour downscaled results against in-situ SM observations from five networks on the TP (Maqu, Naqu, Ali, Shiquanhe, and CTP) during the thawing season. The comprehensive results are presented visually through raincloud plots in Fig. 6 and detailed numerically in Table 3.

The raincloud plots in Fig. 6 provide a comprehensive overview of the statistical distribution for each model's performance against in-situ data. Among the models, LSTM exhibits the highest median value and an excellent overall distribution for the R metric against in-situ observations; however, it shows clear disadvantages in other metrics. In

Table 3

Statistical results of the downscaled products from different models against in-situ network data (R, bias, RMSE, and ubRMSE). The top three performers for each metric are highlighted in shades of green from dark to light.

Products	Metrics	PSCNet	LSTM	ResNet	UNet	ViT	RF
Maqu	R	0.380	0.419	0.354	0.347	0.391	0.221
	bias	-0.003	0.079	0.025	-0.011	-0.066	0.002
	RMSE	0.105	0.138	0.109	0.103	0.118	0.105
	ubRMSE	0.075	0.091	0.077	0.070	0.076	0.073
Naqu	R	0.690	0.790	0.652	0.634	0.586	0.772
	bias	0.071	0.090	0.091	0.023	0.025	0.077
	RMSE	0.118	0.121	0.131	0.078	0.073	0.101
	ubRMSE	0.091	0.078	0.093	0.071	0.066	0.060
Ali	R	0.725	0.620	0.549	0.474	0.535	0.378
	bias	0.018	0.012	0.031	-0.007	0.018	0.015
	RMSE	0.048	0.051	0.059	0.052	0.052	0.056
	ubRMSE	0.045	0.049	0.050	0.051	0.049	0.054
Shiquanhe	R	0.500	0.534	0.462	0.469	0.417	0.453
	bias	0.048	0.032	0.072	0.034	0.098	0.024
	RMSE	0.084	0.059	0.098	0.060	0.117	0.048
	ubRMSE	0.060	0.042	0.060	0.043	0.059	0.031
CTP	R	0.726	0.782	0.672	0.632	0.575	0.699
	bias	0.011	0.061	0.037	-0.023	-0.056	0.028
	RMSE	0.102	0.118	0.110	0.101	0.118	0.098
	ubRMSE	0.081	0.084	0.085	0.072	0.069	0.072

contrast, the PSCNet model integrates the advantages of temporal and spatial models, not only maintaining a temporal correlation comparable to that of LSTM but also exhibiting strong performance in error metrics, with RMSE and ubRMSE distributions significantly lower than LSTM's. Furthermore, the UNet, ViT, and RF models demonstrate a strong capability for error minimization, highlighting their effectiveness in leveraging spatial context for error reduction.

The statistical results of the downscaled products from different models against the in-situ sites are detailed in Table 3. The overall results are largely consistent with the findings from the raincloud plots. The LSTM model excels in temporal correlation, but its strong performance on other metrics is largely confined to the Ali and Shiquanhe networks. Conversely, models such as RF, UNet, and ViT often demonstrate robust performance in the bias, RMSE, and ubRMSE metrics. While our proposed PSCNet does not always emerge as the top performer in every network, it consistently maintains competitive and stable results across all sites.

The preceding analysis has detailed the statistical performance of the various downscaling models against in-situ station data. However, this performance is not entirely consistent with the models' fitting capabilities against the original SMAP microwave satellite data. This discrepancy stems from the inherent systematic measurement bias between satellite-based SMAP data and ground-based in-situ observations. To further investigate these temporal behaviors, we utilize temporal dynamics visualization to examine the patterns of the downscaled data across different networks. Fig. 7 shows the temporal dynamics plots constructed by sampling and averaging stations within each network, displaying three data types: in-situ observations, SMAP satellite data, and downscaled products. SMAP satellite data have a temporal frequency of 12 hours, while in-situ observations and downscaled products have a temporal frequency of 3 hours. The original SMAP microwave products exhibit varying degrees of overestimation compared to in-situ measurements across different networks, with this issue being particularly pronounced in the Naqu network. Moreover, SMAP products show substantial instability in their temporal dynamics, particularly during the rainy season with high precipitation.

A comparison of the downscaled products with the in-situ SM data reveals several key points: 1) The original SMAP microwave data exhibit a significant bias against the in-situ network measurements. This bias is particularly pronounced in networks with higher precipitation (e.g., Maqu, Naqu, and CTP) compared to the more arid networks (Shiquanhe and Ali). 2) Consequently, the SM estimates from the downscaled products also show a general overestimation relative to the in-situ data, as their training target is the biased SMAP data itself, rather than the in-situ observations.

Compared with the original SMAP satellite product sequence, the downscaled results from the various models show distinct characteristics in their temporal fidelity. LSTM and PSCNet effectively track the temporal dynamics of

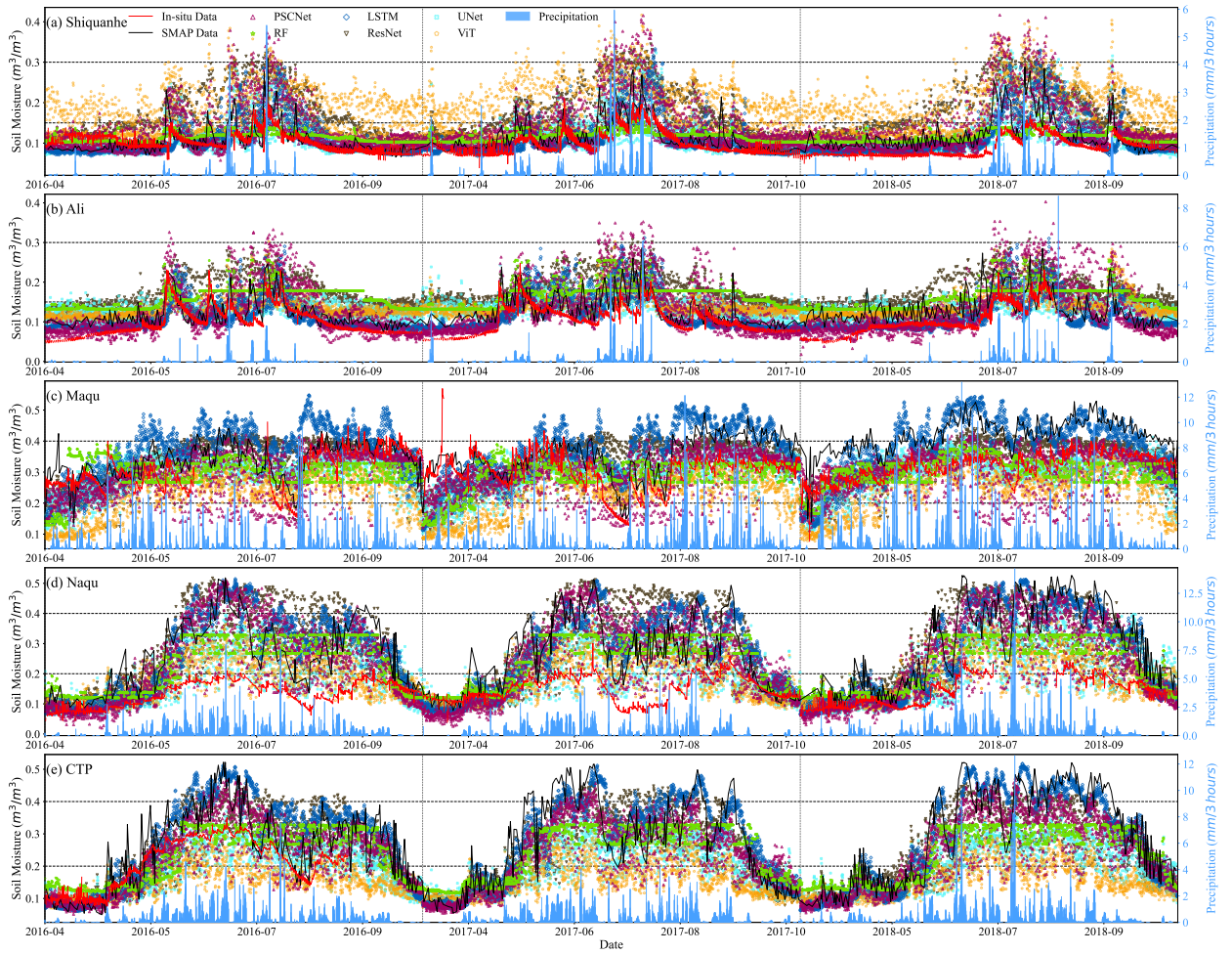


Figure 7: Time series of SM from in-situ networks, SMAP satellite products, and downscaled products, and precipitation data, for (a) Shiquanhe; (b) Ali; (c) Maqu; (d) Naqu; and (e) CTP across five networks spanning the freeze-thaw season of 2016.4–2018.12.

the SMAP sequence, with their performance being particularly strong in arid networks. In high-precipitation networks, however, their behaviors differ during the rainy season: PSCNet is more responsive to the magnitude of precipitation-induced changes, which can lead to greater prediction variability, whereas LSTM provides smoother temporal tracking, albeit with a slight overestimation. The other baseline models struggle to a greater extent in capturing these dynamics. The RF model provides overly smoothed estimates with insufficient sensitivity to sharp fluctuations. Similarly, ResNet and UNet fail to capture the full range of temporal variations, and this issue is most severe with ViT, whose predictions are erratic and substantially deviate from the observed sequence.

Overall, the systematic bias present between the original SMAP data and in-situ observations is propagated into all downscaled products through the training process, becoming most evident during the rainy season. Among the baseline models, LSTM demonstrates excellent temporal capture capabilities, while others such as RF, UNet, and ResNet are particularly effective at reducing errors and bias, our proposed PSCNet uniquely synthesizes these capabilities, delivering both robust temporal reconstruction and consistently low errors.

4.3.2. Evaluation of the Proposed PSCNet Model at Individual Sites

The network-level metrics offer a valuable macro-level overview. However, they inherently average out and obscure the granular, site-specific behaviors of the model. Therefore, a site-level validation is conducted here to provide a more nuanced and robust assessment of PSCNet's performance under diverse local conditions.

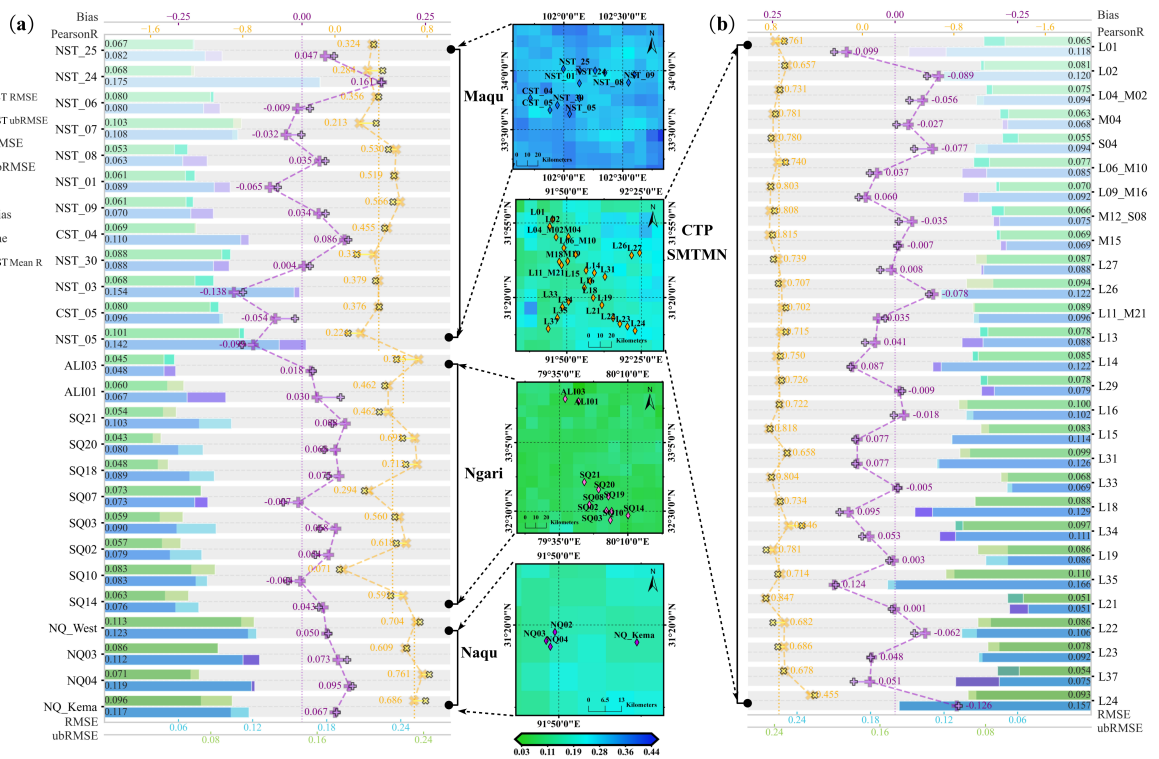


Figure 8: Validation of representative sites within the network using the PSCNet model. Panels (a) and (b) display site-specific error bar plots. (a) presents sites from the Maqu, Ngari, and Naqu networks, while (b) is dedicated to the CTP network. The figure's central section illustrates the annual averages of the SM products and their spatial distribution within each network. For validation metrics, RMSE and ubRMSE mean values are presented as stacked bar charts, visualizing the relative differences between PSCNet and the Mean model using varying transparency. Conversely, for R and Bias mean values, the line plots display scatter points for PSCNet, and for each site, an additional point representing the Mean model's performance is plotted. Furthermore, reference line (hereinafter referred to as the Zero Line) indicating the network-level mean R for all models is included in the line plots. All Mean model values (for both bar and line plots) and the Zero Line are computed from PSCNet, RF, LSTM, and Temporal. In these visualizations, only PSCNet's metrics are explicitly labeled.

As shown in Fig. 8, the performance metrics exhibit distinct characteristics across the different networks. In networks with lower annual mean SM values, such as the Ali network, all models generally demonstrate better and more consistent metrics (bias, RMSE, ubRMSE). Conversely, PSCNet demonstrates superior performance in networks with higher annual mean SM values, where other models typically fail. For instance, PSCNet maintains consistent advantages in the Maqu network, which exhibits the highest mean SM values. Finally, PSCNet's consistent improvement in the R metric across a majority of sites, which appears independent of specific network characteristics, can be attributed to its learning paradigm of effectively integrating spatial context with temporal dependencies.

4.4. Spatial Results of Soil Moisture Downscaling

Based on the comprehensive visualization of the entire TP, we further present enlarged maps of the Qaidam Basin. The Qaidam Basin is a vast, enclosed intermontane rift basin characterized by a dry plateau continental climate, making it one of the plateau's most sensitive and responsive regions to climate change. As shown in Fig. 9, in the original imagery, the periphery of the Qaidam Basin displays higher SM, while the interior features a pronounced, narrow, low-SM valley.

Visual assessment of the downscaled products will be conducted based on two criteria: (1) consistency with SMAP microwave satellite products; and (2) image smoothness and texture features. It should be noted that image smoothness and texture richness represent subjective visual comparisons and do not directly indicate reconstruction accuracy.

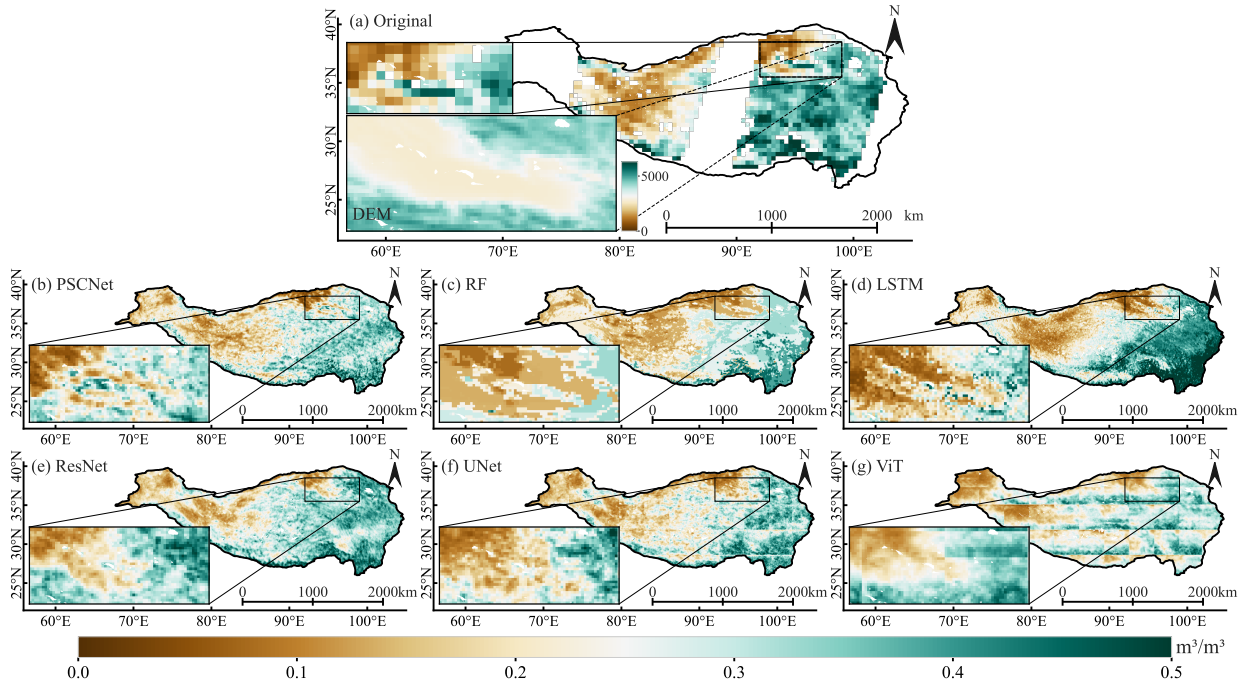


Figure 9: Visualization of SM downscaling results (open water areas masked, 2018-09-04 06:00 UTC). The Qaidam Basin was selected as the detailed visualization region. (a) Original SMAP satellite SM with additional DEM map. (b)-(g) Visualization of downscaled SM results using different downscaling methods.

In terms of consistency with SMAP microwave satellite products, as illustrated in Fig. 9, the PSCNet, RF, and LSTM models demonstrate the highest visual fidelity in matching topographical contours. However, the RF and LSTM models tends to underestimate SM values, which obscures the distinct high-SM valley in the central part of the basin. The ResNet model produces moderate results; it fails to reconstruct finer details, making the basin's distinct valley nearly indiscernible. The worst performances come from the UNet and ViT models, which exhibit noticeable tiling effects across the TP landscape. Specifically within the Qaidam Basin, the UNet model aligns more closely with topographical contours, while ViT displays the poorest performance.

In terms of image smoothness and texture features, the quality of texture reconstruction should be evaluated based on the degree of restoration to the SMAP microwave satellite data. Models with point-based training strategies, such as RF and LSTM, show good pixel-wise accuracy but produce fragmented or discontinuous spatial patterns that compromise overall texture coherence. Models that leverage spatial context can theoretically generate more coherent results, but the final output is significantly influenced by both the model's architecture and its inference mode. For instance, both PSCNet and ResNet employ a full-image inference strategy, which effectively mitigates the edge artifacts common in patch-based methods and yields more natural image textures. However, while this strategy enables ResNet to produce texturally coherent outputs, the early integration of spatial information inherent in its architectural design concurrently leads to an over-smoothing effect that sacrifices fine-detail reconstruction. PSCNet, conversely, leverages an architecture purposefully designed to overcome this limitation, successfully preserving fine-grained details and achieving a superior balance between texture fidelity and spatial clarity.

5. Discussion

In the discussion section, we examine model performance from two key perspectives: 1. temporal generalization analysis across different time points to evaluate model robustness in temporal extrapolation scenarios, and 2. inter-model uncertainty analysis using the three-cornered hat (TCH) method to quantify the consistency among different models.

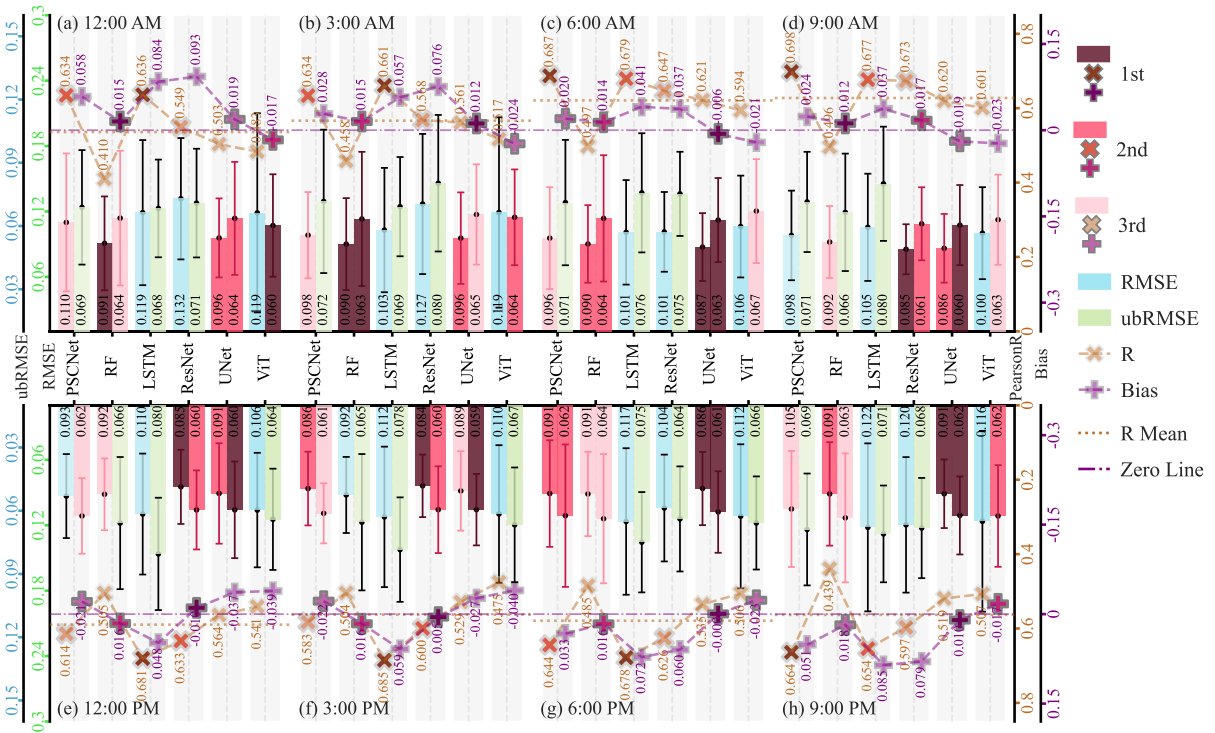


Figure 10: Plot of statistical metrics for comparisons between in-situ measurements and downscaled products across eight timestamps (0, 3, 6, 9, 12, 15, 18, 21) UTC with error bars representing the 25th and 75th percentiles of the data distribution.

Table 4
Relative Generalization Errors ($\overline{RE}$) for all methods across temporal intervals without training data (excluding 6:00 and 18:00 UTC), calculated according to Equation A.2. Positive values indicate better temporal generalization capability compared to training timestamps, while negative values indicate performance degradation.

Metric	PSCNet	LSTM	ResNet	UNet	ViT	RF
$\overline{RE}_{ubRMSE}$ (%)	-1.052	1.018	4.273	-0.150	4.518	-2.706
$\overline{RE}_R$ (%)	-4.167	-1.937	-5.176	-5.007	-5.328	-5.153

5.1. Temporal Generalization Analysis

A primary challenge in this study stems from the temporal discrepancy between the training data and the downscaling target. The training dataset, derived from SMAP satellite products, is limited to observations at approximately 6:00 and 18:00 UTC. The objective, however, is to reconstruct data across the entire diurnal cycle at a 3-hour resolution. This temporal gap raises critical concerns about the models' generalization capability when extrapolating to untrained timestamps. Specifically, we hypothesize that the reconstruction performance may degrade as the temporal distance from the two anchor training times increases, leading to larger biases. Therefore, this section aims to quantitatively assess this temporal generalization error to evaluate the robustness and practical applicability of each model.

In this section, we assess the feasibility of the temporal downscaling task by evaluating the performance differences among various models against in-situ station data. Specifically, we aggregate in-situ station data from multiple observation networks, partitioning daily measurements into eight discrete timestamps. We first examine the performance of different models across multiple statistical metrics at each temporal interval through comprehensive visualization analysis. Subsequently, we define temporal generalization error formulas to quantitatively assess the overall generalization capability of each model across non-training timestamps.

The visualization analysis reveals distinct temporal generalization patterns across different model architectures. In Fig. 10, the sequential model LSTM achieves the highest R values in more than half of the timestamps. However,

Table 5

Detailed Relative Generalization Errors (RE, in %) for all models. This table presents the RE at individual non-training timestamps and the overall mean error ($\overline{RE}$), calculated according to Equation A.2. Bold values indicate the worst-case degradation (minimum RE) for each model and metric.

Model	RE_R (%)							RE_{ubRMSE} (%)						
	00:00	03:00	09:00	12:00	15:00	21:00	$\overline{RE}$	00:00	03:00	09:00	12:00	15:00	21:00	$\overline{RE}$
PSCNet	-4.69	-4.79	4.91	-7.71	-12.45	-0.27	-4.17	-3.17	-7.42	-7.03	6.34	8.22	3.25	-1.05
RF	-16.59	-6.73	0.94	2.74	2.68	-10.62	-4.60	-0.04	0.58	-4.35	-3.39	-2.74	0.50	-1.57
LSTM	-6.31	-2.63	-0.28	0.27	0.96	-3.64	-1.94	9.31	8.20	-6.22	-6.83	-4.12	5.76	1.02
ResNet	-13.71	-10.73	5.82	-0.48	-5.75	-6.20	-5.18	-1.87	-15.42	12.25	14.28	14.46	1.94	4.27
UNet	-12.99	-2.96	7.17	-2.51	-8.55	-10.19	-5.01	-3.08	-6.08	2.09	3.44	3.59	-0.86	-0.15
ViT	-12.11	-6.03	9.19	-1.63	-13.56	-7.82	-5.33	9.36	3.49	5.19	3.56	-0.41	5.93	4.52

it exhibits weaker performance in terms of bias, RMSE, and ubRMSE. In contrast, the spatial models UNet and ResNet, along with the point-based model RF, demonstrate strong performance in terms of bias, RMSE, and ubRMSE, consistently outperforming the sequential model on these metrics during most time periods. Building on these complementary strengths, PSCNet, with its spatio-temporal architecture, successfully integrates the advantages of both sequential and spatial models. Specifically, it maintains a significant advantage in R, achieving the highest values at three timestamps (6:00, 9:00, and 21:00) and ranking within the top three at the remaining timestamps. Furthermore, it outperforms the sequential model in terms of bias, RMSE, and ubRMSE across most networks.

The quantitative results in Table 5 provide a detailed assessment of the temporal generalization capability of each model. In terms of ($\overline{RE}_R$), models that incorporate sequential modeling, namely LSTM and PSCNet, demonstrate less performance degradation than their counterparts, with the average degradation across all models being approximately 5%. In contrast, for $\overline{RE}_{ubRMSE}$, the spatially explicit models perform exceptionally well, with ViT and ResNet yielding the best results (4.52% and 4.27%, respectively). The average degradation for the other models is approximately 1%, suggesting that the ubRMSE metric is less sensitive to performance decay from temporal extrapolation. A closer look at timestamp-specific errors reveals that for the RE_R metric, the most significant degradation for a majority of models clusters at 00:00. This is likely attributable to two primary factors: its maximal temporal distance from the training anchors (06:00 and 18:00 UTC) and the discrepancy between instantaneous nocturnal conditions and daily-updated auxiliary variables. In contrast, the temporal distribution of degradation for RE_{ubRMSE} does not exhibit such a distinct pattern. Overall, these results validate the feasibility of using the integration of validated high-spatio-temporal resolution reanalysis variables for spatio-temporal downscaling on the TP, as the average temporal generalization error remains within 6% for the R metric and 2% for the ubRMSE metric.

5.2. Uncertainty Analysis

Due to the scarcity of in-situ SM measurement data on the TP, comprehensively assessing the reconstruction performance of various models presents a significant challenge. One approach is to use the TCH method to examine inter-model consistency when only limited in-situ observations are available (Tavella and Premoli, 1994). Notably, the TCH method quantifies inter-model uncertainty rather than absolute accuracy, with lower uncertainty indicating proximity to the ensemble mean but not necessarily superior performance. For this uncertainty analysis, we selected PSCNet-series models and point-based models for comparison. Fig. 11 demonstrates the uncertainties among the five models calculated using the TCH method.

The TCH analysis reveals that the spatial distribution of uncertainty across various models exhibits a consistent primary pattern: uncertainty gradually increases from the Pamir Plateau, through the Karakoram Range and Northern TP, to the Hengduan Mountains, Tanggula Mountains, and the eastern TP. This spatial pattern directly corresponds to the gradient of SM values, indicating that greater inter-model disagreement occurs in regions with higher SM. Beyond this overarching trend, the varied spatial distributions also highlight the region-specific strengths and weaknesses of different modeling approaches. For instance, models such as PSCNet, LSTM, and ResNet, which demonstrated superior performance in high-SM regions during the spatial validation (Fig. 5), paradoxically exhibit far greater uncertainty in these areas compared to other models. This is particularly the case for the LSTM model, which is prone to higher uncertainty in regions with dramatic elevation changes, such as the Gangdise, Hengduan, and Bayan Har Mountains. Furthermore, these inter-model discrepancies are sensitive to the temporal scale, becoming more pronounced at the 3-hourly resolution (Fig. 11 (a)) than at the daily scale (Fig. 11 (b)), which is particularly evident in the UNet and

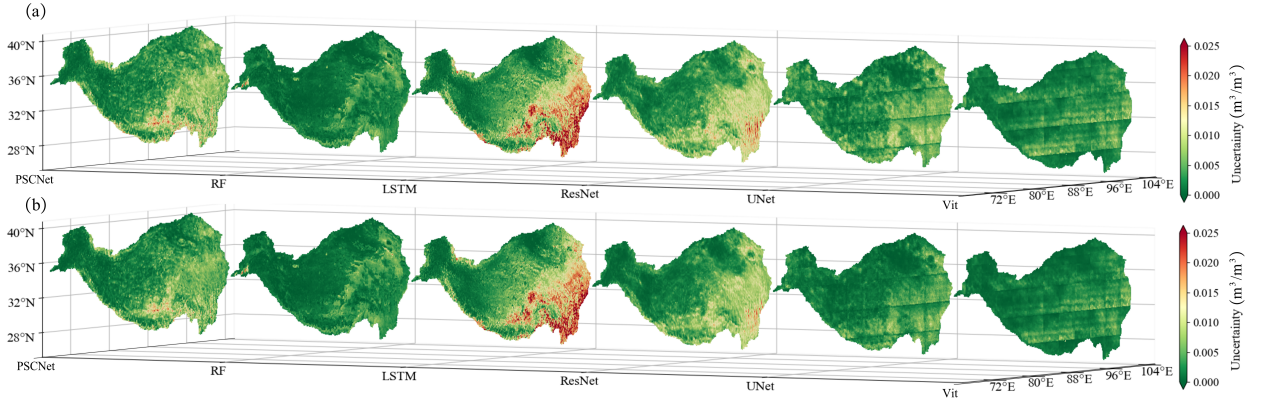


Figure 11: Uncertainties calculated by the TCH methods, (a) (c), (b) (d) Plots of the spatial distribution of the TCH uncertainties of each method after a sampling frequency of 3H and daily averaging, respectively.

ViT products. In contrast, the RF model demonstrates greater stability, maintaining favorable uncertainty performance across both resolutions.

6. Conclusions and future work

Focusing on high-resolution spatio-temporal downscaling, this study investigates two modeling paradigms: spatially explicit models and point-based models, to reconstruct SM data over the TP at 10-km spatial and 3-hour temporal resolution from 2016 to 2018. Our results indicate that the proposed PSCNet spatio-temporal model outperforms various spatial models in satellite product validation, in-situ site validation, and other validation experiments, demonstrating superior accuracy and stability. Additionally, the model shows robust performance in predicting previously unseen time points. In temporal dynamics and visualization validation, PSCNet demonstrates excellent temporal sensitivity and vivid spatial details. These results empirically validate the enhanced generalization capability of our model, emphasizing its potential for application in increasingly complex environments.

The observed performance improvements can be attributed to the model's robust spatio-temporal modeling of SM. By leveraging multi-source auxiliary variables and multi-frequency temporal data, the model effectively captures the nonlinear mapping relationships of SM. This representation enables the model to handle complex scenarios driven by rainfall variability and adapt to dynamic meteorological conditions. These findings underscore the potential of spatio-temporal modeling in advancing the understanding of SM dynamics.

Although this study successfully achieves temporal upsampling to a 3-hour resolution, there is still considerable room for improvement in spatial reconstruction at the 10-km scale. Future work will focus on developing enhanced spatio-temporal reconstruction frameworks. Current limitations include the lack of systematic variable screening that incorporates surface auxiliary features, as well as limited model evaluation across different land cover types, human activities, and climatic conditions. These issues warrant further investigation in subsequent research.

A. Appendix

A.1. Statistical indicators

The specific expressions of the four evaluation indicators used in the study are as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}} \quad (\text{A.1})$$

$$\text{ubRMSE} = \sqrt{\frac{\sum_{i=1}^N [(x_i - \bar{x}) - (y_i - \bar{y})]^2}{N}} \quad (\text{A.2})$$

$$R = \frac{\sum_{i=1}^N ((x_i - \bar{x}) \cdot (y_i - \bar{y}))}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (\text{A.3})$$

$$\text{bias} = \frac{1}{N} \sum_{i=1}^N (x_i - y_i) \quad (\text{A.4})$$

x_i and y_i respectively represent the actual observed SM data and the predicted SM data corresponding to the observation time i , while $\bar{x}$ and $\bar{y}$ respectively represent the overall mean of the actual and predicted SM data during the observation period.

A.2. Relative generalization error index

In this paper, the relative generalization error of various indicators is calculated as follows

$$\bar{X}_{\text{metric}} = \frac{X_{\text{metric}}^6 + X_{\text{metric}}^{18}}{2} \quad (\text{A.5})$$

In the formula, X_{metric}^i represents the value of the metric at time i , and $\bar{X}_{\text{metric}}$ is the mean value of the metric at the time points present in the training dataset, serving as a baseline for validation. This includes metrics such as R and ubRMSE . Following this, the Relative Error (RE) at an individual non-training timestamp j is calculated. The formulas are defined such that a positive value indicates improved performance relative to the baseline:

$$\text{RE}_{\text{ubRMSE}}(j) = \frac{\bar{X}_{\text{ubRMSE}} - X_{\text{ubRMSE}}^j}{\bar{X}_{\text{ubRMSE}}} \quad (\text{A.6})$$

$$\text{RE}_R(j) = \frac{X_R^j - \bar{X}_R}{\bar{X}_R} \quad (\text{A.7})$$

Finally, the mean Relative Generalization Error ($\overline{\text{RE}}$) is computed by averaging the individual RE values across the $N = 6$ non-training timestamps $T = \{0, 3, 9, 12, 15, 21\}$:

$$\overline{\text{RE}}_{\text{metric}} = \frac{1}{N} \sum_{j \in T} \text{RE}_{\text{metric}}(j) \quad (\text{A.8})$$

CRedit authorship contribution statement

Ziyu Zhou: Methodology, Software, Data curation, Visualization, Writing – review & editing. **Keyan Hu:** Methodology, Formal analysis, Visualization, Writing – original draft. **Ling Zhang:** Formal analysis, Writing – review & editing. **Zhaohui Xue:** Conceptualization, Supervision, Methodology, Funding acquisition, Writing – review & editing. **Yutian Fang:** Data curation, Formal analysis, Visualization. **Yusha Zheng:** Formal analysis, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known financial or personal relationships that could be perceived as competing interests, which might have influenced the research presented in this paper.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (42271324, 42201406), the Natural Science Foundation of Jiangsu Province, China (BK20221506), the Natural Science Foundation for Colleges and Universities of Jiangsu Province under Grant 22KJB570006, and the Qing Lan Project of Jiangsu Province of China.

References

- Abbaszadeh, P., Moradkhani, H., Zhan, X., 2019. Downscaling smap radiometer soil moisture over the conus using an ensemble learning method. *Water Resources Research* 55, 324–344.
- Abowarda, A.S., Bai, L., Zhang, C., Long, D., Li, X., Huang, Q., Sun, Z., 2021. Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale. *Remote Sensing of Environment* 255, 112301.
- Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., 2002. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Transactions on geoscience and remote sensing* 40, 2300–2312.
- Akbar, R., Short Gianotti, D.J., Salvucci, G.D., Entekhabi, D., 2020. Partitioning of historical precipitation into evaporation and runoff based on hydrologic dynamics identified with recent smap satellite measurements. *Water Resources Research* 56, e2020WR027307.
- Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.
- Beck, H.E., McVicar, T.R., Vergopolan, N., Berg, A., Lutsko, N.J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A.I., Miralles, D.G., 2023. High-resolution (1 km) köppen-geiger maps for 1901–2099 based on constrained cmip6 projections. *Scientific data* 10, 724.
- Bothe, O., Fraedrich, K., Zhu, X., 2011. Large-scale circulations and tibetan plateau summer drought and wetness in a high-resolution climate model. *International Journal of Climatology* 31, 832–846.
- Carbone, R., Carpenter, M., Burghart, C., 1985. Doppler radar sampling limitations in convective storms. *Journal of Atmospheric and Oceanic Technology* 2, 357–361.
- Carlson, T.N., Gillies, R.R., Perry, E.M., 1994. A method to make use of thermal infrared temperature and ndvi measurements to infer surface soil water content and fractional vegetation cover. *Remote sensing reviews* 9, 161–173.
- Chahine, M.T., 1992. The hydrological cycle and its influence on climate. *Nature* 359, 373–380.
- Chauhan, N., Miller, S., Ardanuy, P., 2003. Spaceborne soil moisture estimation at high resolution: A microwave-optical/ir synergistic approach. *International Journal of Remote Sensing* 24, 4599–4622.
- Colliander, A., Cosh, M.H., Misra, S., Jackson, T.J., Crow, W.T., Chan, S., Bindlish, R., Chae, C., Collins, C.H., Yueh, S.H., 2017. Validation and scaling of soil moisture in a semi-arid environment: Smap validation experiment 2015 (smavex15). *Remote Sensing of Environment* 196, 101–112.
- Collow, T.W., Robock, A., Basara, J.B., Illston, B.G., 2012. Evaluation of smos retrievals of soil moisture over the central united states with currently available in situ observations. *Journal of Geophysical Research: Atmospheres* 117.
- Crow, W.T., Berg, A.A., Cosh, M.H., Loew, A., Mohanty, B.P., Panciera, R., De Rosnay, P., Ryu, D., Walker, J.P., 2012. Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics* 50.
- Crowhurst, D., Dadson, S., Peng, J., Washington, R., 2021. Contrasting controls on congo basin evaporation at the two rainfall peaks. *Climate Dynamics* 56, 1609–1624.
- Dai, L., Fu, R., Guo, X., Du, Y., Zhang, F., Cao, G., 2022. Soil moisture variations in response to precipitation across different vegetation types on the northeastern qinghai-tibet plateau. *Frontiers in Plant Science* 13, 854152.
- Ding, X., Zhang, Y., Ge, Y., Zhao, S., Song, L., Yue, X., Shan, Y., 2024. Unireplknet: A universal perception large-kernel convnet for audio video point cloud time-series and image recognition, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5513–5524.
- Dobriyal, P., Qureshi, A., Badola, R., Hussain, S.A., 2012. A review of the methods available for estimating soil moisture and its implications for water resource management. *Journal of Hydrology* 458, 110–117.
- Dong, J., Crow, W.T., Tobin, K.J., Cosh, M.H., Bosch, D.D., Starks, P.J., Seyfried, M., Collins, C.H., 2020. Comparison of microwave remote sensing and land surface modeling for surface soil moisture climatology estimation. *Remote Sensing of Environment* 242, 111756.
- Dorigo, W., Gruber, A., De Jeu, R., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R., et al., 2015. Evaluation of the esa cci soil moisture product using ground-based observations. *Remote Sensing of Environment* 162, 380–395.
- Dorigo, W., Van Oevelen, P., Wagner, W., Drusch, M., Mecklenburg, S., Robock, A., Jackson, T., 2011. A new international network for in situ soil moisture data. *Eos, Transactions American Geophysical Union* 92, 141–142.
- Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., et al., 2017. Esa cci soil moisture for improved earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment* 203, 185–215.
- Dorigo, W., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiova, A., Sanchis-Dufau, A.D., Zamojski, D., Cordes, C., Wagner, W., Drusch, M., 2013. Global automated quality control of in situ soil moisture data from the international soil moisture network. *Vadose Zone Journal* 12, vzj2012–0097.
- Drusch, M., Wood, E.F., Gao, H., 2005. Observation operators for the direct assimilation of trmm microwave imager retrieved soil moisture. *Geophysical Research Letters* 32.
- Fang, B., Lakshmi, V., Bindlish, R., Jackson, T.J., 2018. Amsr2 soil moisture downscaling using temperature and vegetation data. *Remote Sensing* 10, 1575.
- Fang, K., Shen, C., Kifer, D., Yang, X., 2017. Prolongation of smap to spatiotemporally seamless coverage of continental us using a deep learning neural network. *Geophysical Research Letters* 44, 11–030.

- Gao, L., Gao, Q., Zhang, H., Li, X., Chaubell, M.J., Ebtehaj, A., Shen, L., Wigneron, J.P., 2022. A deep neural network based smap soil moisture product. *Remote Sensing of Environment* 277, 113059.
- Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks, in: *Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings*. pp. 249–256.
- Gualtieri, G., 2021. Reliability of era5 reanalysis data for wind resource assessment: A comparison against tall towers. *Energies* 14, 4169.
- He, K., Zhao, W., Brocca, L., Quintana-Seguí, P., Chen, X., 2025. Smpd-merg: a hybrid downscaling model for high-resolution daily precipitation estimation via merging surface soil moisture and multi-source precipitation data. *IEEE Transactions on Geoscience and Remote Sensing*.
- Hersbach, H., 2016. The era5 atmospheric reanalysis., in: *AGU fall meeting abstracts*, pp. NG33D–01.
- Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Hua, W., Zhou, L., Nicholson, S.E., Chen, H., Qin, M., 2019. Assessing reanalysis data for understanding rainfall climatology and variability over central equatorial africa. *Climate Dynamics* 53, 651–669.
- Huang, Y., Song, R., Xu, K., Ye, X., Li, C., Chen, X., 2020. Deep learning-based inverse scattering with structural similarity loss functions. *IEEE Sensors Journal* 21, 4900–4907.
- Hutengs, C., Vohland, M., 2016. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment* 178, 127–141.
- Im, J., Park, S., Rhee, J., Baik, J., Choi, M., 2016. Downscaling of amsr-e soil moisture with modis products using machine learning approaches. *Environmental Earth Sciences* 75, 1–19.
- Ines, A.V., Droogers, P., 2002. Inverse modelling in estimating soil hydraulic functions: a genetic algorithm approach. *Hydrology and Earth System Sciences* 6, 49–66.
- Jiang, H., Shen, H., Li, H., Lei, F., Gan, W., Zhang, L., 2017. Evaluation of multiple downscaled microwave soil moisture products over the central tibetan plateau. *Remote Sensing* 9, 402.
- Jiang, L., Islam, S., 2003. An intercomparison of regional latent heat flux estimation using remote sensing data. *International Journal of Remote Sensing* 24, 2221–2236.
- Jiang, S.h., Wei, L.y., Ren, L.l., Zhang, L.q., Wang, M.h., Cui, H., 2023. Evaluation of imerg, tmpa, era5, and cpc precipitation products over mainland china: Spatiotemporal patterns and extremes. *Water Science and Engineering* 16, 45–56.
- Jin, L., Ganopolski, A., Chen, F., Claussen, M., Wang, H., 2005. Impacts of snow and glaciers over tibetan plateau on holocene climate change: Sensitivity experiments with a coupled model of intermediate complexity. *Geophysical Research Letters* 32.
- Jung, M., Reichstein, M., Schwalm, C.R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., et al., 2017. Compensatory water effects link yearly global land co2 sink changes to temperature. *Nature* 541, 516–520.
- Kaheil, Y.H., Gill, M.K., McKee, M., Bastidas, L.A., Rosero, E., 2008. Downscaling and assimilation of surface soil moisture using ground truth measurements. *IEEE Transactions on Geoscience and Remote Sensing* 46, 1375–1384.
- Kim, D., Moon, H., Kim, H., Im, J., Choi, M., 2018. Intercomparison of downscaling techniques for satellite soil moisture products. *Advances in Meteorology* 2018, 4832423.
- Kim, G., Barros, A.P., 2002. Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data. *Remote Sensing of Environment* 83, 400–413.
- Kim, J., Hogue, T.S., 2011. Improving spatial soil moisture representation through integration of amsr-e and modis products. *IEEE Transactions on Geoscience and Remote Sensing* 50, 446–460.
- Kolassa, J., Reichle, R., Liu, Q., Alemohammad, S., Gentile, P., Aida, K., Asanuma, J., Bircher, S., Caldwell, T., Colliander, A., et al., 2018. Estimating surface soil moisture from smap observations using a neural network technique. *Remote sensing of environment* 204, 43–59.
- Kroes, J., Wesseling, J., Van Dam, J., 2000. Integrated modelling of the soil–water–atmosphere–plant system using the model swap 2·0 an overview of theory and an application. *Hydrological processes* 14, 1993–2002.
- Kubota, T., Aonashi, K., Ushio, T., Shige, S., Takayabu, Y.N., Kachi, M., Arai, Y., Tashima, T., Masaki, T., Kawamoto, N., et al., 2020. Global satellite mapping of precipitation (gsmap) products in the gpm era. *Satellite precipitation measurement: Volume 1*, 355–373.
- Lehner, B., Verdin, K., Jarvis, A., 2008. New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union* 89, 93–94.
- Lei, F., Senyurek, V., Kurum, M., Gurbuz, A.C., Boyd, D., Moorhead, R., Crow, W.T., Eroglu, O., 2022. Quasi-global machine learning-based soil moisture estimates at high spatio-temporal scales using cygnss and smap observations. *Remote Sensing of Environment* 276, 113041.
- Li, W., Kiaghadi, A., Dawson, C., 2021. High temporal resolution rainfall–runoff modeling using long-short-term-memory (lstm) networks. *Neural Computing and Applications* 33, 1261–1278.
- Li, X., Chen, S., Hu, X., Yang, J., 2019. Understanding the disharmony between dropout and batch normalization by variance shift, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2682–2690.
- Li, Z., Yang, X., Tang, H., 2023. Evaluation of the hourly era5 radiation product and its relationship with aerosols over china. *Atmospheric Research* 294, 106941.
- Liu, Y., Jing, W., Wang, Q., Xia, X., 2020. Generating high-resolution daily soil moisture by using spatial downscaling techniques: A comparison of six machine learning algorithms. *Advances in Water Resources* 141, 103601.
- Liu, Y., Yang, Y., Jing, W., Yue, X., 2017. Comparison of different machine learning approaches for monthly satellite-based soil moisture downscaling over northeast china. *Remote Sensing* 10, 31.
- Liu, Y., Yu, J., Han, Y., 2018. Understanding the effective receptive field in semantic image segmentation. *Multimedia Tools and Applications* 77, 22159–22171.
- Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A convnet for the 2020s, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986.

- Loew, A., 2008. Impact of surface heterogeneity on surface soil moisture retrievals from passive microwave data at the regional scale: The upper danube case. *Remote Sensing of Environment* 112, 231–248.
- Luo, W., Li, Y., Urtasun, R., Zemel, R., 2016. Understanding the effective receptive field in deep convolutional neural networks. *Advances in neural information processing systems* 29.
- Malbêteau, Y., Merlin, O., Molero, B., Rüdiger, C., Bacon, S., 2016. Dispatch as a tool to evaluate coarse-scale remotely sensed soil moisture using localized in situ measurements: Application to smos and amsr-e data in southeastern australia. *International Journal of Applied Earth Observation and Geoinformation* 45, 221–234.
- Martens, B., Schumacher, D.L., Wouters, H., Muñoz-Sabater, J., Verhoest, N.E., Miralles, D.G., 2020. Evaluating the land-surface energy partitioning in era5. *Geoscientific Model Development* 13, 4159–4181.
- Mascaro, G., Vivoni, E.R., Deidda, R., 2011. Soil moisture downscaling across climate regions and its emergent properties. *Journal of Geophysical Research: Atmospheres* 116.
- McColl, K.A., Alemohammad, S.H., Akbar, R., Konings, A.G., Yueh, S., Entekhabi, D., 2017. The global distribution and dynamics of surface soil moisture. *Nature Geoscience* 10, 100–104.
- Merlin, O., Al Bitar, A., Walker, J.P., Kerr, Y., 2010. An improved algorithm for disaggregating microwave-derived soil moisture based on red, near-infrared and thermal-infrared data. *Remote Sensing of Environment* 114, 2305–2316.
- Merlin, O., Rudiger, C., Al Bitar, A., Richaume, P., Walker, J.P., Kerr, Y.H., 2012. Disaggregation of smos soil moisture in southeastern australia. *IEEE Transactions on Geoscience and Remote Sensing* 50, 1556–1571.
- Molero, B., Merlin, O., Malbêteau, Y., Al Bitar, A., Cabot, F., Stefan, V., Kerr, Y., Bacon, S., Cosh, M., Bindlish, R., et al., 2016. Smos disaggregated soil moisture product at 1 km resolution: Processor overview and first validation results. *Remote Sensing of Environment* 180, 361–376.
- Montzka, C., Jagdhuber, T., Horn, R., Bogena, H.R., Hajsek, I., Reigber, A., Vereecken, H., 2016. Investigation of smap fusion algorithms with airborne active and passive l-band microwave remote sensing. *IEEE Transactions on Geoscience and Remote Sensing* 54, 3878–3889.
- Nagy, A., Kiss, N.É., Buday-Bódi, E., Magyar, T., Cavazza, F., Gentile, S.L., Abdullah, H., Tamás, J., Fehér, Z.Z., 2024. Precision estimation of crop coefficient for maize cultivation using high-resolution satellite imagery to enhance evapotranspiration assessment in agriculture. *Plants* 13, 1212.
- Nasta, P., Penna, D., Brocca, L., Zuecco, G., Romano, N., 2018. Downscaling near-surface soil moisture from field to plot scale: A comparative analysis under different environmental conditions. *Journal of Hydrology* 557, 97–108.
- Nemes, A., Rawls, W.J., Pachepsky, Y.A., 2005. Influence of organic matter on the estimation of saturated hydraulic conductivity. *Soil Science Society of America Journal* 69, 1330–1337.
- Njoku, E.G., Entekhabi, D., 1996. Passive microwave remote sensing of soil moisture. *Journal of hydrology* 184, 101–129.
- Njoku, E.G., Jackson, T.J., Lakshmi, V., Chan, T.K., Nghiem, S.V., 2003. Soil moisture retrieval from amsr-e. *IEEE transactions on Geoscience and remote sensing* 41, 215–229.
- Njoku, E.G., Wilson, W.J., Yueh, S.H., Dinardo, S.J., Li, F.K., Jackson, T.J., Lakshmi, V., Bolten, J., 2002. Observations of soil moisture using a passive and active low-frequency microwave airborne sensor during sgp99. *IEEE Transactions on Geoscience and Remote Sensing* 40, 2659–2673.
- Ochsner, T.E., Cosh, M.H., Cuenca, R.H., Dorigo, W.A., Draper, C.S., Hagimoto, Y., Kerr, Y.H., Larson, K.M., Njoku, E.G., Small, E.E., et al., 2013. State of the art in large-scale soil moisture monitoring. *Soil Science Society of America Journal* 77, 1888–1919.
- Orth, R., et al., 2021. Global soil moisture data derived through machine learning trained with in-situ measurements. *Scientific Data* 8, 1–14.
- Papamakarios, G., Pavlakou, T., Murray, I., 2017. Masked autoregressive flow for density estimation. *Advances in neural information processing systems* 30.
- Parinussa, R.M., Meesters, A.G., Liu, Y.Y., Dorigo, W., Wagner, W., de Jeu, R.A., 2011. Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model. *IEEE Geoscience and Remote Sensing Letters* 8, 779–783.
- Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M.H., Crow, W.T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M.W., et al., 2021. A roadmap for high-resolution satellite soil moisture applications—confronting product characteristics with user requirements. *Remote Sensing of Environment* 252, 112162.
- Peng, J., Borsche, M., Liu, Y., Loew, A., 2013. How representative are instantaneous evaporative fraction measurements of daytime fluxes? *Hydrology and Earth System Sciences* 17, 3913–3919.
- Peng, J., Loew, A., Chen, X., Ma, Y., Su, Z., 2016. Comparison of satellite-based evapotranspiration estimates over the tibetan plateau. *Hydrology and earth system sciences* 20, 3167–3182.
- Peng, J., Loew, A., Merlin, O., Verhoest, N.E., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics* 55, 341–366.
- Peng, J., Loew, A., Zhang, S., Wang, J., Niesel, J., 2015. Spatial downscaling of satellite soil moisture data using a vegetation temperature condition index. *IEEE Transactions on Geoscience and Remote Sensing* 54, 558–566.
- Perez, L., Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.
- Petropoulos, G.P., Ireland, G., Barrett, B., 2015. Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth, Parts a/b/c* 83, 36–56.
- Piles, M., Camps, A., Vall-Llossera, M., Corbella, I., Panciera, R., Rudiger, C., Kerr, Y.H., Walker, J., 2011. Downscaling smos-derived soil moisture using modis visible/infrared data. *IEEE transactions on Geoscience and Remote Sensing* 49, 3156–3166.
- Piles, M., Petropoulos, G.P., Sánchez, N., González-Zamora, Á., Ireland, G., 2016. Towards improved spatio-temporal resolution soil moisture retrievals from the synergy of smos and msg seviri spaceborne observations. *Remote Sensing of Environment* 180, 403–417.
- Qiu, J., 2008. China: The third pole.
- Qu, Y., Zhu, Z., Montzka, C., Chai, L., Liu, S., Ge, Y., Liu, J., Lu, Z., He, X., Zheng, J., et al., 2021. Inter-comparison of several soil moisture downscaling methods over the qinghai-tibet plateau, china. *Journal of hydrology* 592, 125616.

- Rodriguez-Fernandez, N.J., Aires, F., Richaume, P., Kerr, Y.H., Prigent, C., Kolassa, J., Cabot, F., Jimenez, C., Mahmoodi, A., Drusch, M., 2015. Soil moisture retrieval using neural networks: Application to smos. *IEEE Transactions on Geoscience and Remote Sensing* 53, 5991–6007.
- Sabaghy, S., Walker, J.P., Renzullo, L.J., Jackson, T.J., 2018. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sensing of Environment* 209, 551–580.
- Sadeghi, M., Ebtehaj, A., Crow, W.T., Gao, L., Purdy, A.J., Fisher, J.B., Jones, S.B., Babaeian, E., Tuller, M., 2020. Global estimates of land surface water fluxes from smos and smap satellite soil moisture data. *Journal of Hydrometeorology* 21, 241–253.
- Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., 2010. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews* 99, 125–161.
- Shangguan, Y., Min, X., Shi, Z., 2023. Inter-comparison and integration of different soil moisture downscaling methods over the qinghai-tibet plateau. *Journal of Hydrology* 617, 129014.
- Shangguan, Y., Min, X., Wang, N., Tong, C., Shi, Z., 2024. A long-term, high-accuracy and seamless 1km soil moisture dataset over the qinghai-tibet plateau during 2001–2020 based on a two-step downscaling method. *GIScience & Remote Sensing* 61, 2290337.
- Shin, Y., Mohanty, B.P., 2013. Development of a deterministic downscaling algorithm for remote sensing soil moisture footprint using soil and vegetation classifications. *Water Resources Research* 49, 6208–6228.
- Skakun, S., Justice, C.O., Vermote, E., Roger, J.C., 2018. Transitioning from modis to viirs: an analysis of inter-consistency of ndvi data sets for agricultural monitoring. *International journal of remote sensing* 39, 971–992.
- Sobayo, R., Wu, H.H., Ray, R., Qian, L., 2018. Integration of convolutional neural network and thermal images into soil moisture estimation, in: 2018 1st International Conference on Data Intelligence and Security (ICDIS), IEEE. pp. 207–210.
- Srivastava, P.K., Han, D., Ramirez, M.R., Islam, T., 2013. Machine learning techniques for downscaling smos satellite soil moisture using modis land surface temperature for hydrological application. *Water resources management* 27, 3127–3144.
- Tavella, P., Premoli, A., 1994. Estimating the instabilities of n clocks by measuring differences of their readings. *Metrologia* 30, 479. URL: <https://dx.doi.org/10.1088/0026-1394/30/5/003>, doi:10.1088/0026-1394/30/5/003.
- Ustin, S.L., Middleton, E.M., 2021. Current and near-term advances in earth observation for ecological applications. *Ecological Processes* 10, 1.
- Vereecken, H., Amelung, W., Bauke, S.L., Bogaen, H., Brüggemann, N., Montzka, C., Vanderborght, J., Bechtold, M., Blöschl, G., Carminati, A., et al., 2022. Soil hydrology in the earth system. *Nature Reviews Earth & Environment* 3, 573–587.
- Verhoest, N.E., Van Den Berg, M.J., Martens, B., Lievens, H., Wood, E.F., Pan, M., Kerr, Y.H., Al Bitar, A., Tomer, S.K., Drusch, M., et al., 2015. Copula-based downscaling of coarse-scale soil moisture observations with implicit bias correction. *IEEE Transactions on geoscience and remote sensing* 53, 3507–3521.
- Wagner, W., Pathe, C., Doubkova, M., Sabel, D., Bartsch, A., Hasenauer, S., Blöschl, G., Scipal, K., Martínez-Fernández, J., Löw, A., 2008. Temporal stability of soil moisture and radar backscatter observed by the advanced synthetic aperture radar (asar). *Sensors* 8, 1174–1197.
- Wang, Y., Yang, J., Chen, Y., Wang, A., De Maeyer, P., 2018. The spatiotemporal response of soil moisture to precipitation and temperature changes in an arid region, china. *Remote Sensing* 10, 468.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 600–612.
- Wei, Z., Meng, Y., Zhang, W., Peng, J., Meng, L., 2019. Downscaling smap soil moisture estimation with gradient boosting decision tree regression over the tibetan plateau. *Remote Sensing of Environment* 225, 30–44.
- Wigneron, J.P., Calvet, J.C., Pellarin, T., Van de Griend, A.A., Berger, M., Ferrazzoli, P., 2003. Retrieving near-surface soil moisture from microwave radiometric observations: current status and future plans. *Remote Sensing of Environment* 85, 489–506.
- Xia, Y., Sheffield, J., Ek, M.B., Dong, J., Chaney, N., Wei, H., Meng, J., Wood, E.F., 2014. Evaluation of multi-model simulated soil moisture in nldas-2. *Journal of Hydrology* 512, 107–125.
- Xie, Y., Huang, J., Wu, G., Liu, Y., Dong, W., Lu, M., He, B., Su, Z., Bao, Q., Zhao, Q., et al., 2023. Oceanic repeaters boost the global climatic impact of the tibetan plateau. *Science Bulletin* 68, 2225–2235.
- Xing, Z., Zhao, L., Fan, L., De Lannoy, G., Bai, X., Liu, X., Peng, J., Frappart, F., Yang, K., Li, X., et al., 2025. Retrieval of 1 km surface soil moisture from sentinel-1 over bare soil and grassland on the qinghai-tibetan plateau. *Remote Sensing of Environment* 318, 114563.
- Yang, K., Qin, J., Zhao, L., Chen, Y., Tang, W., Han, M., Lazhu, Chen, Z., Lv, N., Ding, B., et al., 2013. A multiscale soil moisture and freeze–thaw monitoring network on the third pole. *Bulletin of the American Meteorological Society* 94, 1907–1916.
- Yilmaz, M., 2023. Accuracy assessment of temperature trends from era5 and era5-land. *Science of the Total Environment* 856, 159182.
- Yin, D., Song, X., Zhu, X., Guo, H., Zhang, Y., Zhang, Y., 2023. Spatiotemporal analysis of soil moisture variability and its driving factor. *Remote Sensing* 15, 5768.
- Zeng, J., Li, Z., Chen, Q., Bi, H., Qiu, J., Zou, P., 2015. Evaluation of remotely sensed and reanalysis soil moisture products over the tibetan plateau using in-situ observations. *Remote Sensing of environment* 163, 91–110.
- Zhan, X., Houser, P.R., Walker, J.P., Crow, W.T., 2006. A method for retrieving high-resolution surface soil moisture from hydros l-band radiometer and radar observations. *IEEE Transactions on Geoscience and Remote Sensing* 44, 1534–1544.
- Zhang, P., Zheng, D., van der Velde, R., Wen, J., Ma, Y., Zeng, Y., Wang, X., Wang, Z., Chen, J., Su, Z., 2022. A dataset of 10-year regional-scale soil moisture and soil temperature measurements at multiple depths on the tibetan plateau. *Earth System Science Data* 14, 5513–5542.
- Zhang, R., Kim, S., Sharma, A., 2019. A comprehensive validation of the smap enhanced level-3 soil moisture product using ground measurements over varied climates and landscapes. *Remote sensing of environment* 223, 82–94.
- Zhang, X., Lu, N., Jiang, H., Yao, L., 2020. Evaluation of reanalysis surface incident solar radiation data in china. *Scientific Reports* 10, 3494.
- Zhang, Y., Li, B., Liu, L., Zheng, D., 2021. Redetermine the region and boundaries of tibetan plateau. *Geogr. Res* 40, 1543–1553.
- Zhao, H., Li, J., Yuan, Q., Lin, L., Yue, L., Xu, H., 2022. Downscaling of soil moisture products using deep learning: Comparison and analysis on tibetan plateau. *Journal of Hydrology* 607, 127570.
- Zhao, W., Li, A., 2013. A downscaling method for improving the spatial resolution of amsr-e derived soil moisture product based on msg-seviri data. *Remote Sensing* 5, 6790–6811.

- Zhao, W., Sánchez, N., Lu, H., Li, A., 2018. A spatial downscaling approach for the smap passive surface soil moisture product using random forest regression. *Journal of hydrology* 563, 1009–1024.
- Zheng, J., Zhao, T., Lü, H., Shi, J., Cosh, M.H., Ji, D., Jiang, L., Cui, Q., Lu, H., Yang, K., et al., 2022. Assessment of 24 soil moisture datasets using a new in situ network in the shandian river basin of china. *Remote Sensing of Environment* 271, 112891.
- Zhou, Z., Guo, B., Xing, W., Zhou, J., Xu, F., Xu, Y., 2020. Comprehensive evaluation of latest gpm era imerg and gsmap precipitation products over mainland china. *Atmospheric Research* 246, 105132.
- Zhuo, L., Han, D., 2017. Multi-source hydrological soil moisture state estimation using data fusion optimisation. *Hydrology and Earth System Sciences* 21, 3267–3285.
- Zou, D., Zhao, L., Sheng, Y., Chen, J., Hu, G., Wu, T., Wu, J., Xie, C., Wu, X., Pang, Q., et al., 2017. A new map of permafrost distribution on the tibetan plateau. *The Cryosphere* 11, 2527–2542.