ISPRS Journal of Photogrammetry and Remote Sensing 218 (2024) 70–86

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Generalized spatio-temporal-spectral integrated fusion for soil moisture downscaling

Menghui Jiang [a], Huanfeng Shen [a,*], Jie Li [b], Liangpei Zhang [c]

[a] School of Resource and Environmental Sciences, Wuhan University, PR China
[b] School of Geodesy and Geomatics, Wuhan University, PR China
[c] State Key Laboratory of Information Engineering, Survey Mapping and Remote Sensing, Wuhan University, PR China

ARTICLE INFO

Keywords:
Soil moisture (SM)
Generalized spatio-temporal-spectral integrated fusion
Statistical downscaling
Spatio-temporal fusion downscaling
STC-CycleGAN

ABSTRACT

Soil moisture (SM) is one of the key land surface parameters, but the coarse spatial resolution of the passive microwave SM products constrains the precise monitoring of surface changes. The existing SM downscaling methods typically either utilize spatio-temporal information or leverage auxiliary parameters, without fully mining the complementary information between them. In this paper, a generalized spatio-temporal-spectral integrated fusion-based downscaling method is proposed to fully utilize the complementary features between multi-source auxiliary parameters and multi-temporal SM data. Specifically, we define the spectral characteristic of geographic objects as an assemblage of diverse attribute characteristics at specific spatio-temporal locations and scales. Based on this, the SM-related auxiliary parameter data can be treated as the generalized spectral characteristics of SM, and a generalized spatio-temporal-spectral integrated fusion framework is proposed to integrate the spatio-temporal features of the SM products and the generalized spectral features from the auxiliary parameters to generate fine spatial resolution SM data with high quality. In addition, considering the high heterogeneity of multi-source data, the proposed framework is based on a spatio-temporal constrained cycle generative adversarial network (STC-CycleGAN). The proposed STC-CycleGAN network comprises a forward integrated fusion stage and a backward spatio-temporal constraint stage, between which spatio-temporal cycle-consistent constraints are formed. Numerous experiments were conducted on Soil Moisture Active Passive (SMAP) SM products. The qualitative, quantitative, and in-situ site verification results demonstrate the capability of the proposed method to mine the complementary information of multi-source data and achieve high-accuracy downscaling of global daily SM data from 36 km to 9 km.

1. Introduction

Soil moisture (SM) is a crucial variable in the Earth's surface layer that plays a significant role in diverse processes such as the energy budget and water cycle of land ecosystems (Petropoulos et al., 2015), drought monitoring (Hirschi et al., 2011), agricultural irrigation (Martínez-Fernández et al., 2016), and water resource management (Renzullo et al., 2014). Rapid and accurate monitoring of SM is of great significance in the management of climate, hydrology, ecology, and other systems (Jiang et al., 2019; Xiao et al., 2021; Huang et al., 2023).

The traditional in-situ measurements provide precise estimation of point-wise SM but have limited spatial coverage. Furthermore, it is costly to deploy and maintain monitoring networks, making it difficult to comprehensively reflect the spatial distribution of large-scale SM (Dorigo et al., 2015; Srivastava, 2017; Yang et al., 2024). Modern satellite remote sensing technology, characterized by its continuity, periodicity, cost-effectiveness, and global coverage, is extensively employed in SM retrieval (Petropoulos et al., 2015). Microwave remote sensing, due to its longer wavelength, is less susceptible to bad weather conditions and is a promising approach for accurate SM retrieval at a global scale (Engman and Chauhan, 1995; Shangguan et al., 2023). As a result, numerous active and passive microwave systems have been successively deployed to capture surface SM, such as the MetOp A/B Advanced Scatterometer (ASCAT) (Bartalis et al., 2007), the Soil Moisture and Ocean Salinity (SMOS) mission (Kerr et al., 2010), and the Soil Moisture Active Passive mission (SMAP) (Entekhabi et al., 2010a). However, active microwave sensors are sensitive to vegetation structure and surface roughness, leading to less accurate SM products (Sabaghy et al.,

* Corresponding author.
E-mail address: shenhf@whu.edu.cn (H. Shen).

https://doi.org/10.1016/j.isprsjprs.2024.10.012
Received 29 April 2024; Received in revised form 9 October 2024; Accepted 11 October 2024
Available online 19 October 2024

Table 1
Datasets used in this study.

Dataset	Variables	Spatialresolution	Temporalresolution	Usage
SMAP	AP_9	9 km	Daily	Label
	P_{36}, TB,	36 km	Daily	Input
	Longitude, Latitude	9 km	–	Input
MODIS	NDVI	1 km	16 days	Input
	LST	1 km	8 days	Input
	Reflectance	500 m	8 days	Input
NOAA	DEM	1 km	–	Input
In-situ SM	ISMN, ARS Micronet	point	1 h, 5 min	Validation

2018; Xu et al., 2022). In contrast, passive microwave sensors are generally more reliable, due to their reduced sensitivity to local surface features; however, the spatial resolution of the acquired SM data often extends to several tens of kilometers, thus limiting their practical application (Fang et al., 2019; Shangguan et al., 2023).

To estimate SM data at finer spatial resolutions with reliable accuracy, numerous downscaling methods have been developed for the coarse-resolution passive microwave SM products. All these downscaling methods require the integration of new information to achieve resolution enhancement (Jing et al., 2024). Therefore, they can be viewed from the perspective of data fusion and can be broadly classified into three main categories: 1) remote sensing-dynamic model fusion-based methods (also known as data assimilation-based methods); 2) multi-parameter fusion-based methods; and 3) spatio-temporal fusion-based methods. Remote sensing-dynamic model fusion-based downscaling refers to inputting the low spatial resolution (LR) remote sensing observations into a dynamic model to adjust the model trajectories, thereby generating the high spatial resolution (HR) target parameters. Representative methods include the four-dimensional variational data assimilation method (Reichle et al., 2001) and the ensemble Kalman filter assimilation methods (Sahoo et al., 2013; Draper and Reichle, 2015; Mahfouf et al., 2009). However, due to the complex physical processes involved, these kinds of approaches are computationally intensive and difficult to apply at fine scales (Peng et al., 2017).

Due to the simplicity and efficiency, the current dominant approach is statistical downscaling, also known as multi-parameter fusion-based downscaling (Jing et al., 2022; Jing et al., 2024), which introduces HR auxiliary parameters to fuse with the LR SM data to obtain the HR SM data. This type of method typically constructs a statistical relationship model between SM and auxiliary parameters at a low resolution and applies it to a high resolution under the scale-invariance assumption of the relationship. From the perspective of auxiliary data, on the one hand, researchers have leveraged active microwave auxiliary data and proposed a series of change detection methods to retrieve SM from radar backscatter data (Narayan et al., 2006; Piles et al., 2009; Das et al., 2010). On the other hand, researchers have employed optical/thermal infrared auxiliary data and proposed downscaling methods to estimate SM from surface parameters such as land surface temperature (LST) and vegetation indices (Chauhan et al., 2003; Sánchez-Ruiz, et al., 2014; Peng et al., 2016; Kim et al., 2017). In addition, geoinformation data, such as terrain and elevation data, have frequently been used as an additional ancillary data source (Ranney et al., 2015; Hoehn et al., 2017). However, most of the above-mentioned traditional statistical methods suffer from the limitation of a linear assumption. Recently, due to the powerful nonlinear feature extraction capability of machine learning algorithms, researchers have utilized machine learning approaches such as random forest (RF), artificial neural networks (ANN), and residual networks (ResNets) to retrieve SM data from multi-source auxiliary data, and have achieved notable results (Wei et al., 2019; Liu et al., 2020; Jiang et al., 2022a; Zhang et al., 2022; Zhao et al., 2022). Overall, the performance of multi-parameter fusion-based downscaling is heavily reliant on the quality of the auxiliary parameters, which can suffer from spatial and temporal instability (Sabaghy et al., 2018).

The other universal category is spatio-temporal fusion-based downscaling, which exploits the complementarity in the spatial and temporal resolutions among multiple sensors (Gao et al., 2006; Tang et al., 2021). These methods establish spatio-temporal transformation relationships based on HR and LR surface parameter pairs from reference dates and the LR surface parameter from the target date to achieve surface parameter downscaling of the target date. The spatio-temporal fusion-based methods have been widely employed in the downscaling of remote sensing parameters such as surface reflectance, LST, and vegetation indices (Tewes et al., 2015; Shen et al., 2016; Cheng et al., 2017). Regarding SM, Jiang et al. (2019) first proposed a spatio-temporal fusion model (STFM) to extend the 9-km SMAP SM product. They created an HR-LR baseline SM pair from the historical 9-km and 36-km SMAP SM products and applied non-local filtering to fuse the baseline pair with daily 36-km SM data, ultimately obtaining a 9-km product from 2015 to 2017. Building upon this foundation, Yang et al. (2022) proposed an STFM based on virtual image pairs (VIPSTF) for simulating the 9-km SMAP SM product after a malfunction in the active radar sensor. They synthesized 10 sets of virtual HR-LR baseline SM pairs using the historical 9-km and 36-km SM products and coupled a spatial weighting scheme to generate 8-day synthesized 9-km SMAP SM data from 2015 to 2020. The spatio-temporal fusion-based approaches, which require no additional heterogeneous auxiliary data, exhibit high feasibility and potential for the downscaling of surface parameters. However, they have difficulty creating high-quality HR-LR baseline pairs and addressing the dramatic changes in surface parameters where there is a long interval between the target date and the baseline date (i.e., reference date).

As mentioned above, multi-parameter fusion-based downscaling and spatio-temporal fusion-based downscaling have their respective pros and cons, exhibiting pronounced complementarity but lacking a unified framework. To address this issue, this paper presents a broad definition of the spectral characteristics of geographic objects and treats auxiliary parameters as the generalized spectral features of SM. Building upon this, multi-parameter fusion-based downscaling that fuses the HR auxiliary parameters and the LR SM can be considered as generalized spatial-spectral fusion-based downscaling, and a generalized spatio-temporal-spectral fusion-based downscaling model is proposed to unify the spatio-temporal fusion-based downscaling and multi-parameter fusion-based downscaling. The main innovations of this paper can be summarized as follows:

(1) This paper presents a broad definition of the spectral characteristics of geographic objects as a collection of diverse attributes at specific spatio-temporal locations and scales. Furthermore, a generalized spatio-temporal-spectral integrated fusion framework that merges the generalized spectral features from auxiliary parameters and the spatio-temporal features from the target surface parameter is proposed to obtain downscaled surface parameter data with high resolution and high quality.
(2) A spatio-temporal constrained cycle generative adversarial network (STC-CycleGAN) is proposed to fuse the complementary spatial, temporal, and generalized spectral features from multi-source data. The proposed network consists of a GAN-based forward fusion stage to generate the downscaling results, a GAN-based backward constraint stage to implement spatio-temporal cycle-consistent constraints on the downscaling results, and loss functions with cycle-consistency and gradient penalty terms to ensure stable training.
(3) The SM parameter was taken as a research example in this study, and qualitative, quantitative, and in-situ site validations were conducted to verify the effectiveness of the proposed model and framework.

The rest of this paper is structured as follows. In Section II, we

Table 2
Details of the in-situ site networks.

Network	Country	Sensor	Depth (m)	Number
USCRN	USA	Stevens Hydraprobe II Sdi-12	0.05–0.05	63
SCAN	USA	Hydraprobe Analog, Hydraprobe Digital Sdi-12	0.05–0.05	88
TXSON	USA	CS655	0.05–0.05	39
HOBE	Denmark	Decagon 5TE	0.00–0.05	12
TWENTE	Netherlands	5TM, EC-TM	0.05–0.05	6
RSMN	Romania	5TM	0.00–0.05	9
SMOSMANIA	France	ThetaProbe ML2X	0.05–0.05	8
ARS Micronet	USA	Stevens Hydra Probe	0.05–0.05	35

introduce the datasets, while Section III describes the adopted method. Section IV presents the qualitative and quantitative global experimental results and the in-situ site validation analysis. Following this, Section V discusses aspects such as auxiliary parameters and fusion strategies. Finally, we present our conclusions in Section VI.

2. Datasets

This study focused on the SMAP products and utilized auxiliary parameter data mainly from the Moderate Resolution Imaging Spectroradiometer (MODIS) for model training, validation, and testing. In-situ measurements from the International Soil Moisture Network (ISMN) and the USDA Agricultural Research Service's Micronet (ARS Micronet) were employed for verification. Table 1 lists the datasets utilized in this study.

2.1. SMAP soil moisture datasets

The SMAP satellite, launched by NASA on January 31, 2015, carries two sensors: an L-band radar system and an L-band radiometer. The L-band active radar system provides 3-km resolution SM (A_3) data with lower accuracy, while the L-band passive radiometer offers 36-km resolution SM (P_{36}) data with higher accuracy. Based on this, the SMAP team combined the data from the active and passive microwave sensors to produce a 9-km SM (AP_9) product with a balance between spatial resolution and accuracy. Unfortunately, due to damage to the active radar system on July 7, 2015, SMAP can only provide usable A_3 and AP_9 products from April 13, 2015, to July 7, 2015. In this context, in this study, we maximized the use of historically available AP_9 data as input and label data to train the downscaling network model, aiming to achieve its daily continuity beyond the historical data period.

The SMAP satellite collects data at 6:00 AM and 6:00 PM local time during its descending and ascending passes, respectively. Data from the 6:00 AM descending pass were utilized in this study, due to their higher reliability. In addition, the SMAP satellite can achieve global coverage approximately every two to three days, with an accurate revisit period of eight days. However, this results in large strip gaps in the daily SM products, as shown in Fig. 1(a)–(b). In this study, to construct a stable HR-LR SM pair at the reference date for spatio-temporal fusion, we referred to Jiang et al. (2019) and composed a baseline HR-LR SM pair with the widest coverage using the historically available data mainly from June 30, 2015, to July 7, 2015, denoted as (AP_9^{base}, P_{36}^{base}).

2.2. MODIS auxiliary datasets

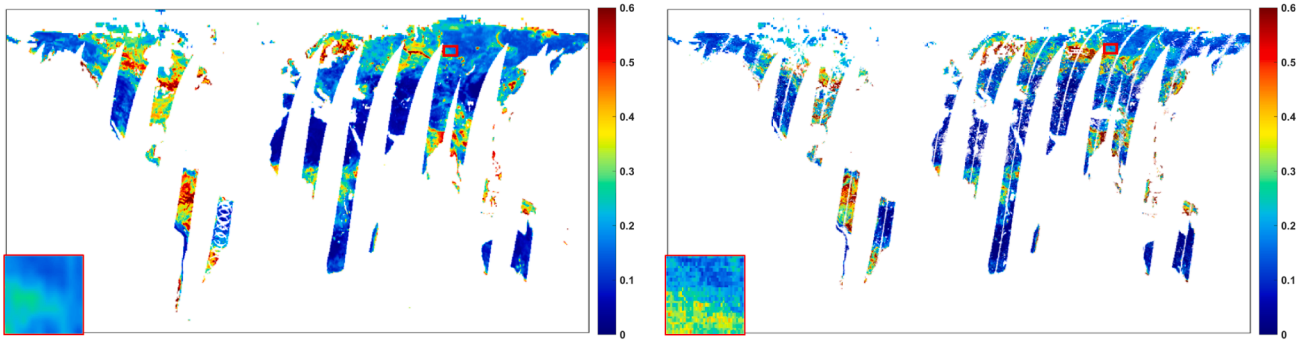
MODIS data, as a crucial component of the NASA Earth Observing System, have been extensively utilized in atmospheric, terrestrial, and oceanic researches (Justice et al., 2002). In this study, various SM-related MODIS surface parameter products, including normalized difference vegetation index (NDVI), LST, and surface reflectance, were utilized as auxiliary parameters (Chen et al., 2017; Hu et al., 2020; Xu et al., 2022). To mitigate the missing data caused by adverse weather conditions and ensure the data quality, we employed the 8-day or 16-day composite MODIS products. Specifically, we utilized the 16-day composite NDVI product from MOD13A2 at a 1-km spatial resolution, the 8-day composite LST product from MOD11A2 at a 1-km spatial resolution, and the 8-day composite surface reflectance product from MOD09A1 at a 500-m spatial resolution. All these products were acquired from the Terra satellite, with an overpass time of around 10:30 AM, which is close to the SMAP descending overpass time (6:00 AM) and is less affected by solar radiation.

2.3. Other auxiliary datasets

In addition, some other auxiliary data were also considered in this study, including a brightness temperature product and geoinformation data, such as terrain, longitude, and latitude. Specifically, the vertical polarization brightness temperature data (TB) obtained from the level-3 36-km passive microwave product of SMAP and the latitude and longitude obtained from the level-3 9-km microwave product of SMAP were utilized in this study. The global digital elevation model (DEM) product of a 30-arc-second resolution (approximately 1 km) released by NOAA (https://www.ngdc.noaa.gov/) was utilized as the terrain factor.

2.4. In-situ soil moisture datasets

The ISMN database is a global in-situ measurement SM database created by the Vienna University of Technology, which is maintained through international cooperation (Dorigo et al., 2013). The database provides SM data measured by multiple sensors at different depths



(a) P_{36} (2015/07/04) (b) AP_9 (2015/07/04)

Fig. 1. SMAP SM products.

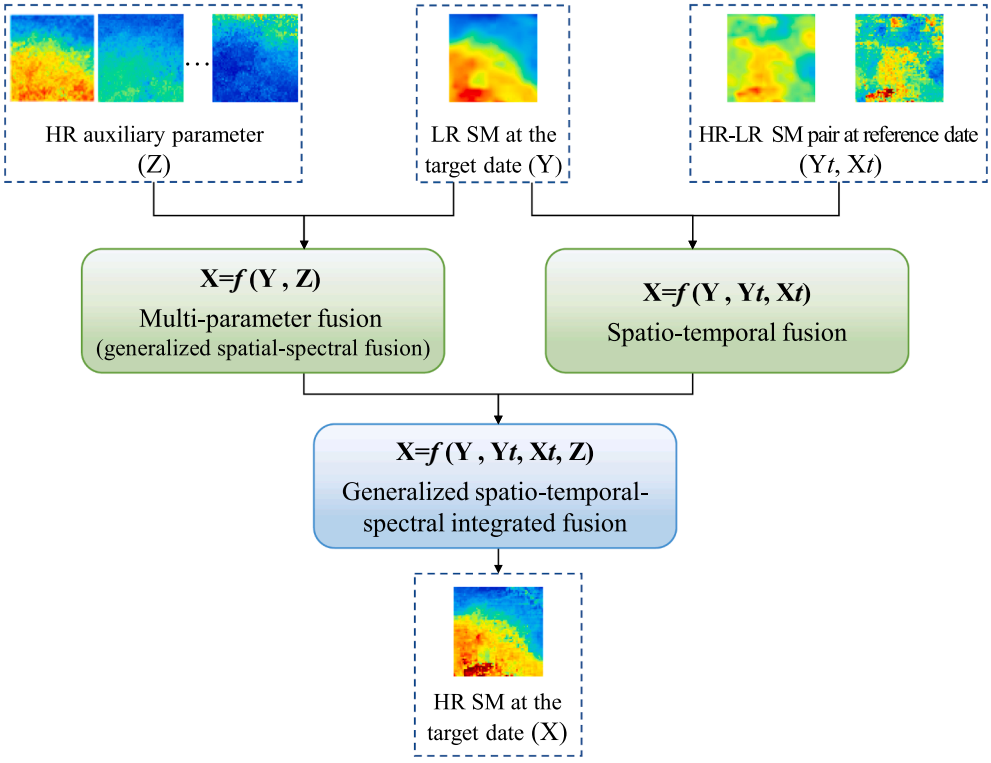


Fig. 2. The formulation of generalized spatio-temporal-spectral integrated fusion for SM downscaling.

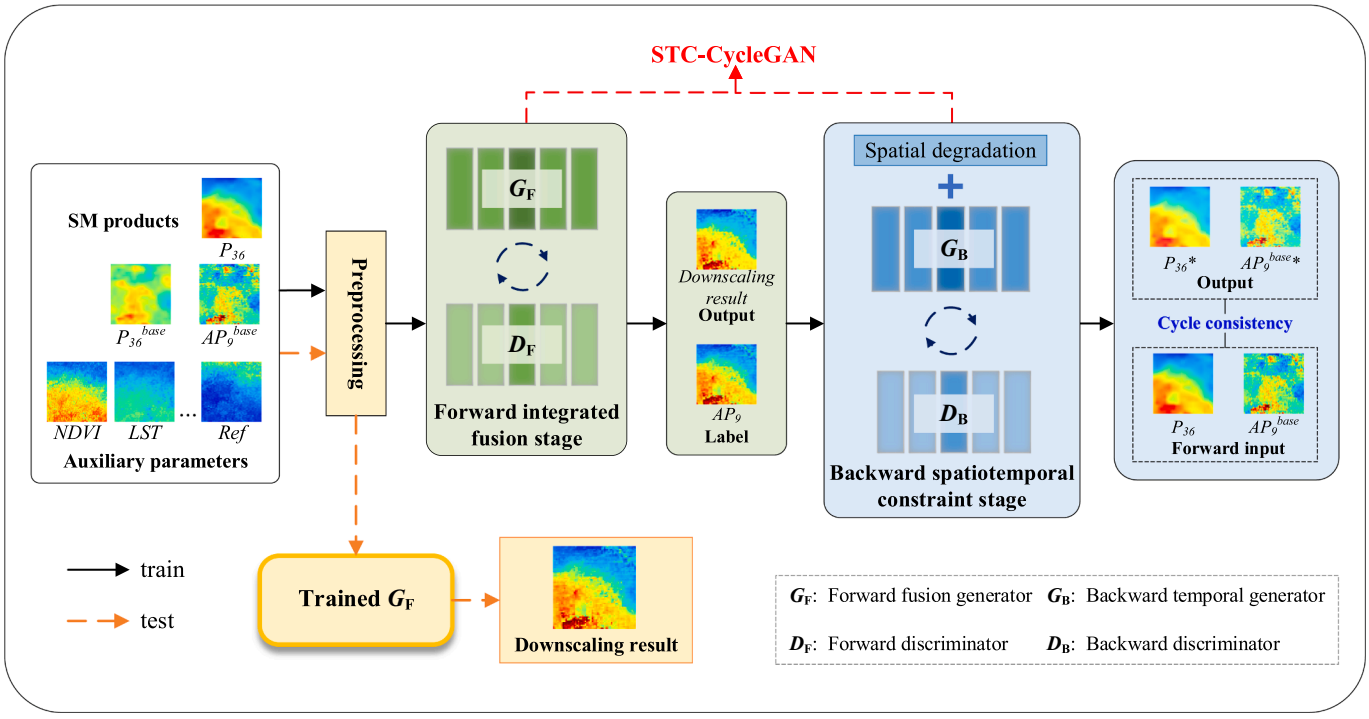


Fig. 3. The flowchart of the proposed STC-CycleGAN-based generalized spatio-temporal-spectral integrated fusion framework.

(2–100 cm) in various regions worldwide, and represents a reliable verification data source for remote sensing satellite SM retrieval. In order to match the measurement depth of the SMAP SM products, we selected hundreds of high-quality sites from seven ISMN networks for verification, each with a measurement depth of 0–5 cm. Moreover, 35 SM sites from ARS Micronet, which is jointly operated by the USDA Agricultural Research Service, Oklahoma State University, and the Oklahoma Climatological Survey, were also utilized in this study, due to their consistent quality.

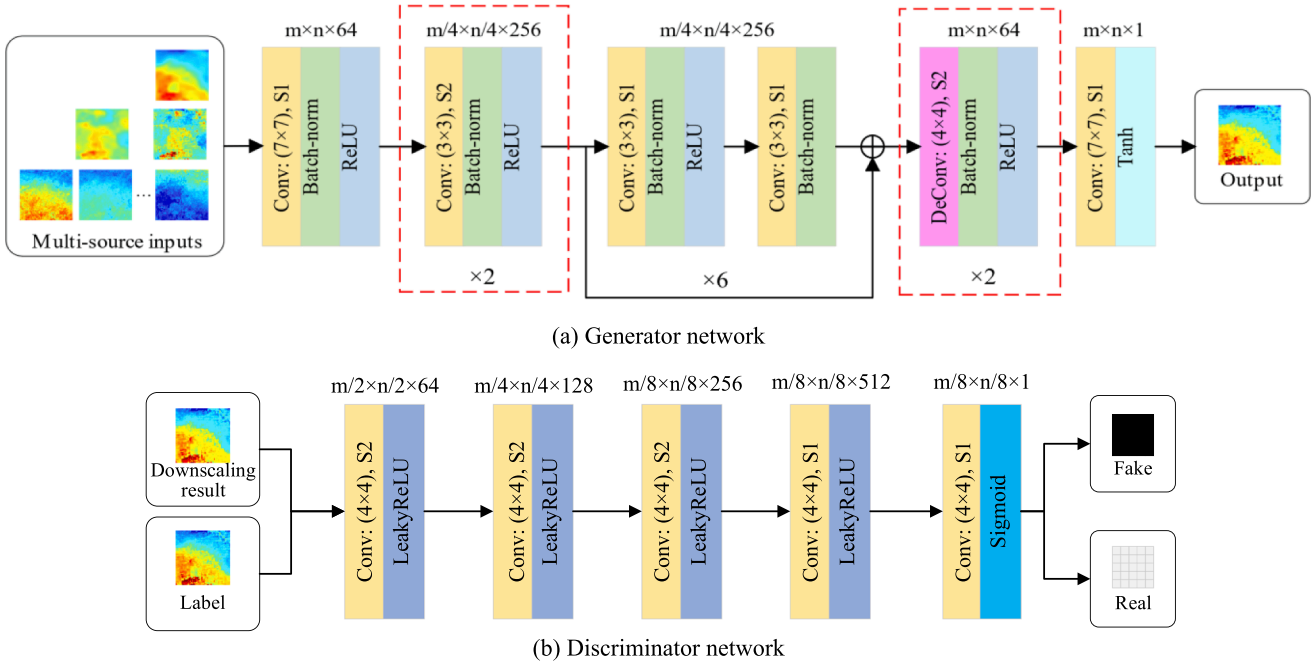


Fig. 4. The network structure.

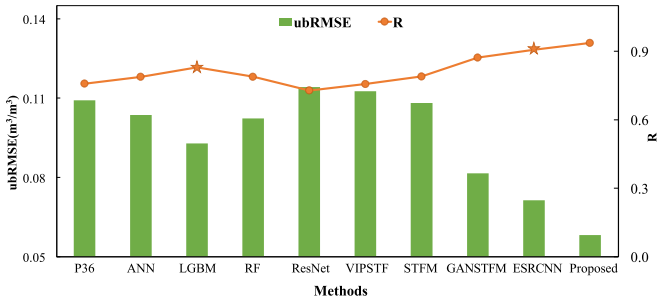


Fig. 5. Combined plots of R values and ubRMSE values for the various comparison methods on the T1 test set.

3. Method

3.1. Generalized spatio-temporal-spectral integrated fusion formulation

To estimate the HR SM data (X) from the LR SM data (Y), two categories of commonly used downscaling methods have been developed. One is multi-parameter fusion-based downscaling, which involves fusing the HR auxiliary parameters (Z) and the LR SM to obtain the HR SM, represented as $X = f(Y,Z)$, with $f(\cdot)$ denoting a mapping function. The other category is spatio-temporal fusion-based downscaling, which aims to fuse the LR SM of the target date and the HR-LR SM pair (X_t, Y_t) at the reference date to obtain the HR SM of the target date, expressed as: $X = f(Y,Y_t,X_t)$. Multi-parameter fusion-based downscaling utilizes auxiliary parameters to reflect the real-time surface states and estimate the HR SM data, but is constrained by the quality of these auxiliary parameters, and thus suffers from spatial and temporal instability. Spatio-temporal fusion-based downscaling is based on the spatio-temporal information of SM, and requires no auxiliary parameters. However, this approach struggles with drastic SM changes between the target and the reference dates. There is strong complementarity between these two categories, but a lack of unification.

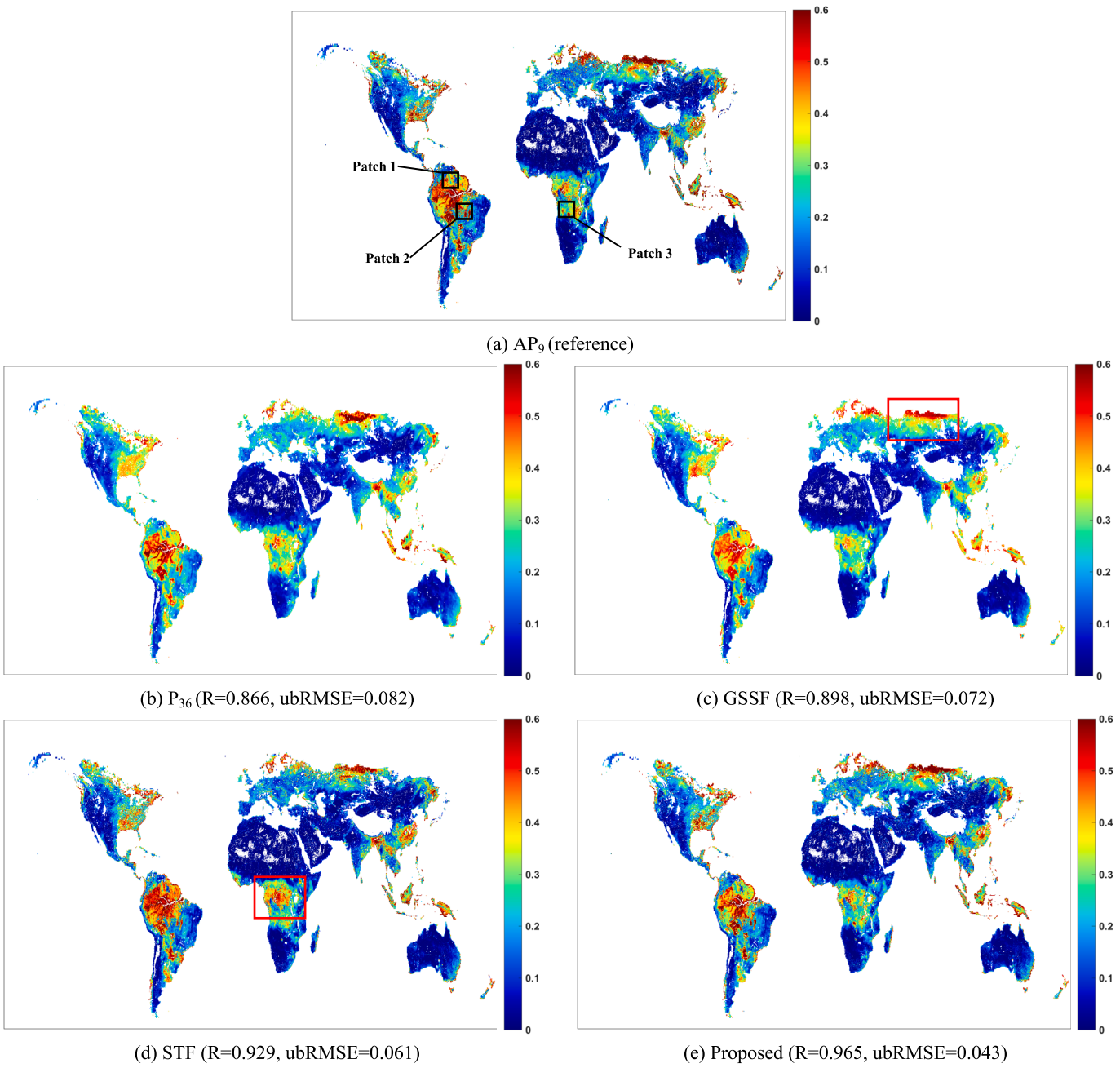
In this study, we aimed to conduct generalized spatio-temporal-spectral fusion to integrate the complementary features between multi-source auxiliary parameters and multi-temporal SM products to generate HR SM data with high stability and high quality. To this end, we broadly defined the spectral characteristics of geographic objects as a collection of diverse attributes arranged according to certain rules at specific spatio-temporal locations and scales. Thus, the auxiliary parameter data (Z) can be regarded as the generalized spectral charac-

Table 4
Average quantitative evaluation on the T1 test set.

Method	R	bias	RMSE	ubRMSE
Ideal data	1	0	0	0
P_{36}	0.7576	−0.0021	0.1098	0.1092
Generalized spatial-spectral fusion (GSSF)	0.8297	0.0089	0.0933	0.0928
Spatio-temporal fusion (STF)	0.9065	−0.0020	0.0714	0.0713
Generalized spatio-temporal-spectral fusion (Proposed)	**0.9368**	**−0.0014**	**0.0583**	**0.0582**

Table 3
Training, validation, and test sets and in-situ datasets.

Dataset	size	Time range	Missing data rate		
			AP_9	P_{36}	MODIS products
Training	40 × 40 × 27920	2015/05/01–2015/06/29	≤40 %	≤40 %	≤2%
Validation	40 × 40 × 1095	2015/04/15–2015/04/30	≤40 %	≤40 %	≤2%
Test	1496 × 3856 × 16	2015/04/15–2015/04/30 (T1)	–	–	–
	1496 × 3856 × 115	2015/07/08–2015/10/31 (T2)			
In-situ	260 points	2015/04/15–2015/07/07 (T1)2015/07/08–2015/10/31 (T2)	–	–	–
Product comparison	1496 × 3856 × 667	2016, 2022	–	–	–



(a) AP_9 (reference)

(b) P_{36} (R=0.866, ubRMSE=0.082)

(c) GSSF (R=0.898, ubRMSE=0.072)

(d) STF (R=0.929, ubRMSE=0.061)

(e) Proposed (R=0.965, ubRMSE=0.043)

Fig. 6. The composed downscaling results on the T1 test set.

teristics of SM, and the multi-parameter fusion-based downscaling can be regarded as a generalized spatial-spectral fusion-based approach, as shown in Fig. 2. Generalized spatio-temporal-spectral integrated fusion is proposed to merge the generalized spectral characteristics from auxiliary parameters (Z) and the spatio-temporal features from multi-temporal SM data (Y, Y_t, X_t) for precise SM downscaling, which can be formulated as: $X = f(Y, Y_t, X_t, Z)$.

The significant differences in scale and modality among the multi-source auxiliary parameters and SM products pose a challenge to the collaborative expression of their complementary information. Owing to the strong feature extraction and representation capabilities, machine learning excels in characterizing complex nonlinear relationships within data, while the GAN and its variant models have demonstrated outstanding performances in fusing multi-modal data (Aggarwal et al., 2021; Wu and Biljecki, 2023). Therefore, in this paper, based on a machine learning framework, we propose a spatio-temporal constrained cycle generative adversarial network (STC-CycleGAN) to mine the complementary spatio-temporal-spectral features from multi-source auxiliary parameters and SM products to achieve high-precision global daily SM downscaling.

3.2. Spatio-temporal constrained cycle generative adversarial network (STC-CycleGAN)

In this section, X represents the ideal HR SM data of the target date, namely AP_9; Y represents the LR SM data of the target date to be downscaled, namely P_{36}; X_t and Y_t represent the composed baseline SM pair of the reference date, namely AP_9^{base} and P_{36}^{base}, respectively; and Z represents the various auxiliary parameters, including NDVI, LST, reflectance (Ref), longitude (Lon), latitude (Lat), DEM, and TB.

The flowchart of the proposed framework is depicted in Fig. 3. The inputs undergo necessary preprocessing steps, including resampling and

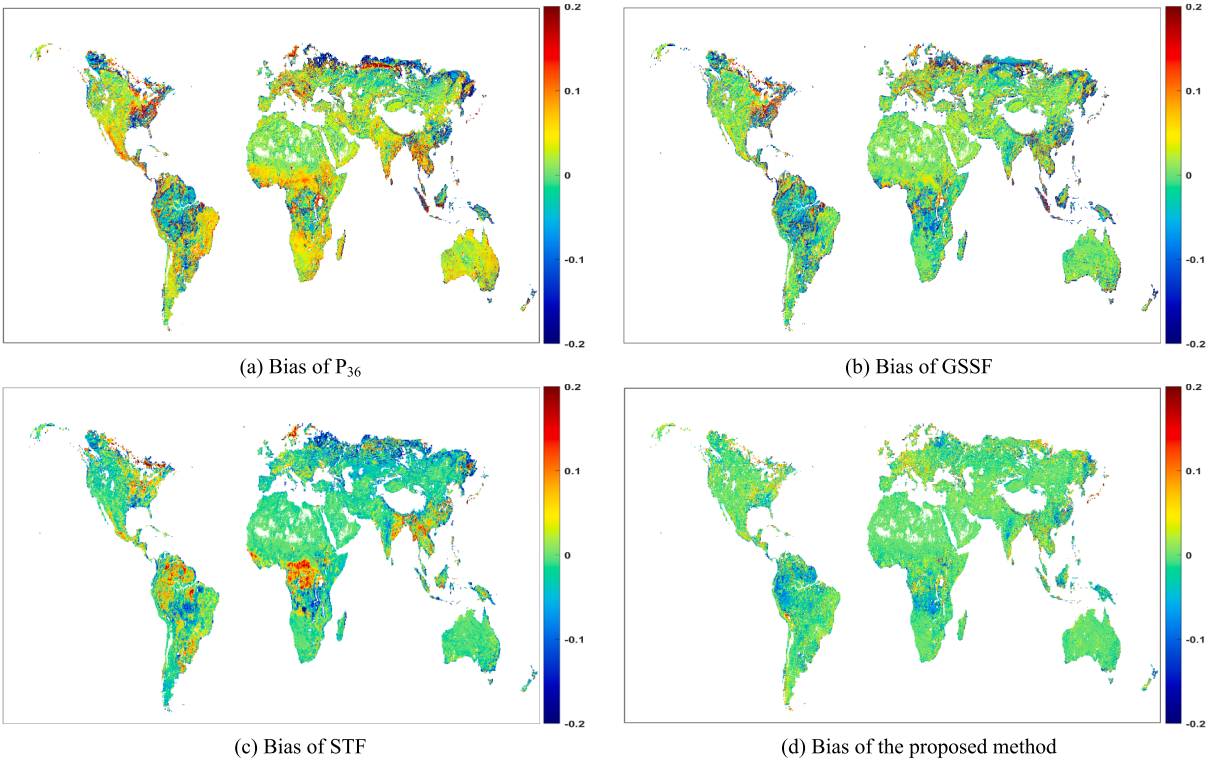


Fig. 7. Bias between the downscaling results and the reference in Fig. 6. Bias > 0 indicates SM overestimation (red), bias < 0 indicates SM underestimation (blue), bias ≈ 0 indicates accurate SM estimation (green).

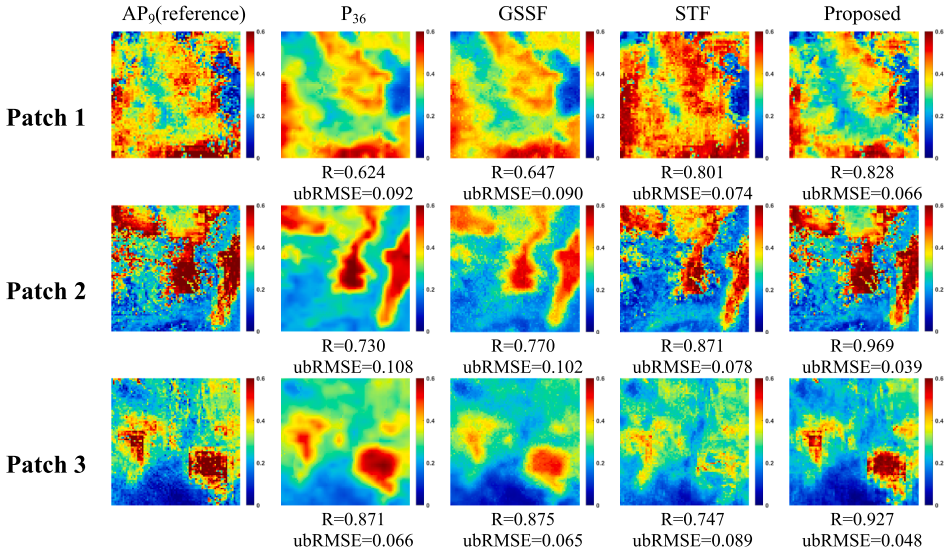


Fig. 8. Detailed spatial information of the downscaled SM (Fig. 6(a), black box areas).

reprojection, before being fed into the proposed STC-CycleGAN network. The STC-CycleGAN network is made up of a forward integrated fusion stage for generating downscaling results and a backward spatio-temporal constraint stage to implement spatio-temporal cycle constraints.

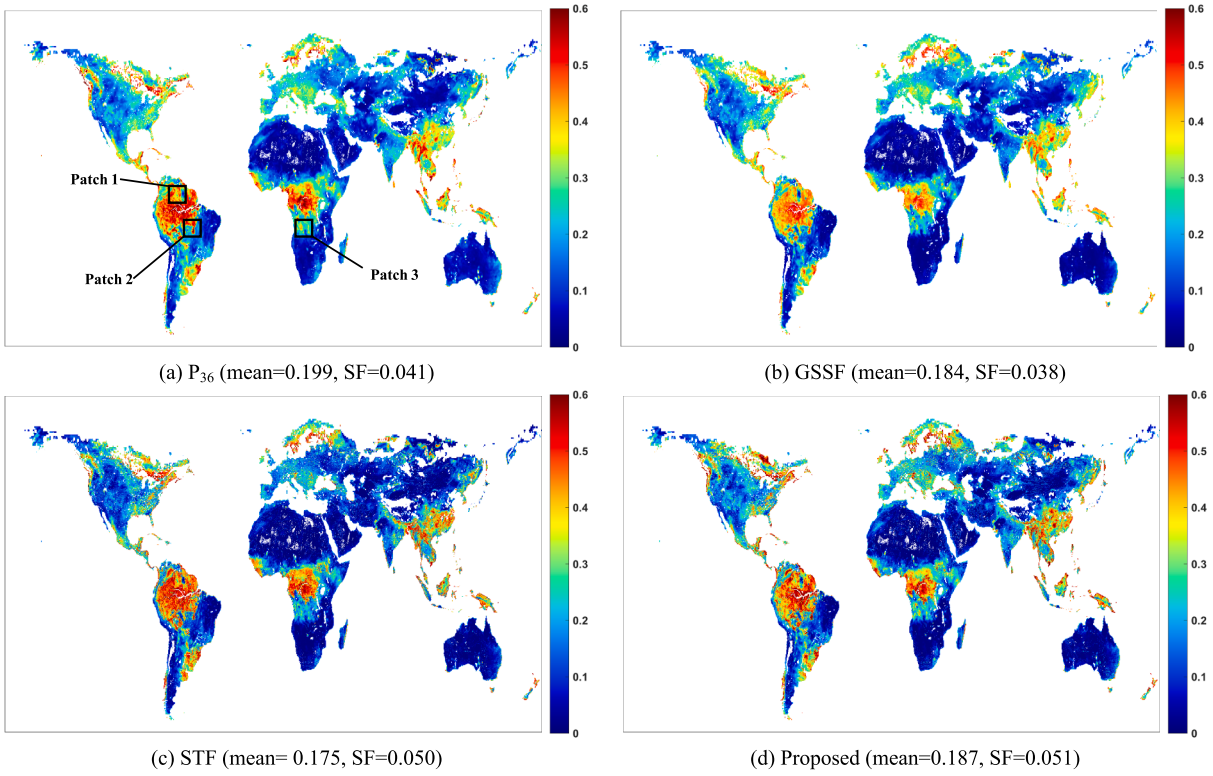
Table 5
Average quantitative evaluation on the T2 test set.

Method	R	bias	RMSE	ubRMSE	mean	SF
Ideal data	1	0	0	0	0.2125 (P_{36})	$+\infty$
GSSF	**0.9622**	0.0106	**0.0366**	**0.0349**	0.1988	0.0321
STF	0.9419	0.0211	0.0494	0.0447	0.1950	0.0396
Proposed	0.9424	**0.0049**	0.0460	0.0457	**0.2061**	**0.0421**

3.2.1. *The forward integrated fusion stage*

As shown in Fig. 3, the forward integrated fusion stage comprises a forward fusion generator network and a forward discriminator network, which optimize their parameters through a minimax game. Specifically, the forward fusion generator aims to produce a downscaled result by integrating the multi-source inputs, as described in Eq. (1):

$$X_d = \boldsymbol{G_F}((Y, Y_t, X_t, Z); \Theta_{GF}) \quad (1)$$



(a) P_{36} (mean=0.199, SF=0.041) (b) GSSF (mean=0.184, SF=0.038)

(c) STF (mean= 0.175, SF=0.050) (d) Proposed (mean=0.187, SF=0.051)

Fig. 9. The composed downscaling results for the last 16 days of the T2 period.

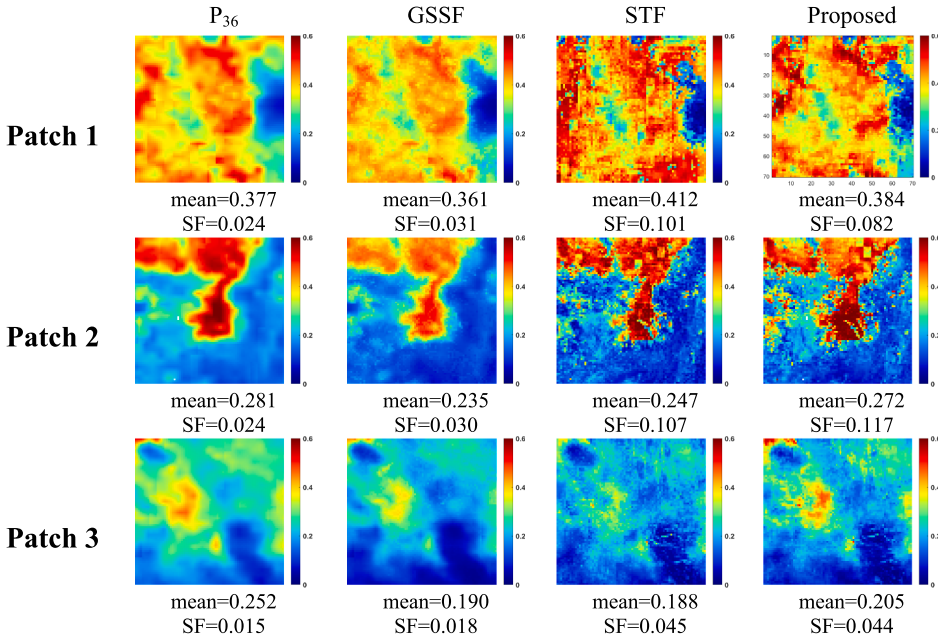


Fig. 10. Detailed spatial information of the downscaled SM (Fig. 9(a), black box areas).

where $\boldsymbol{G_F}(\cdot)$ represents the forward fusion generator network function, and Θ_{GF} denotes its trainable parameters. (Y, Y_t, X_t, Z) is the multi-source input, and X_d is the output downscaled result.

The forward discriminator network aims to differentiate between the downscaled result and the label data (i.e., X_d and X, respectively) to drive the forward integrated fusion generator to produce downscaling results that are closer to the label data. The network optimizes its parameters using the following loss function:

Table 6
The average accuracy of the downscaled SM against in-situ measurements.

Networks	Method	T1				T2			
		R	bias	RMSE	ubRMSE	R	bias	RMSE	ubRMSE
	Ideal data	1	0	0	**0**	1	0	0	0
USCRN	P_{36}	**0.6576**	−0.0062	**0.0802**	**0.0418**	**0.5998**	−0.0245	**0.0734**	**0.0437**
	AP_9	0.5437	0.0310	0.0998	0.0578	–	–	–	–
	GSSF	0.5952	0.0169	0.0823	0.0451	0.5880	−0.0232	0.0735	0.0443
	STF	0.5927	0.0190	0.0879	0.0450	0.5764	**−0.0202**	0.0785	0.0449
	Proposed	0.6396	**0.0061**	0.0860	0.0442	0.5918	−0.0227	0.0798	0.0449
SCAN	P_{36}	**0.5882**	−0.0153	**0.0784**	**0.0458**	**0.5748**	−0.0232	0.0754	**0.0487**
	AP_9	0.4807	0.0415	0.1035	0.0648	–	–	–	–
	GSSF	0.5305	**0.0119**	**0.0784**	0.0489	0.5631	−0.0192	**0.0748**	0.0496
	STF	0.5236	0.0241	0.0864	0.0483	0.5519	−0.0079	0.0828	0.0491
	Proposed	0.5536	0.0172	0.0816	0.0493	0.5659	**−0.0071**	0.0799	0.0495
Others of ISMN	P_{36}	**0.6670**	−0.0047	0.0708	0.0409	**0.7532**	−0.0218	**0.0704**	**0.0414**
	AP_9	0.5528	0.0256	0.0968	0.0638	–	–	–	–
	GSSF	0.5759	0.0166	0.0788	0.0479	0.7382	**−0.0199**	0.0707	0.0428
	STF	0.6114	0.0150	0.0744	0.0451	0.7271	−0.0235	0.0721	0.0439
	Proposed	0.6565	**0.0001**	**0.0693**	**0.0406**	0.7403	−0.0236	0.0726	0.0425
ARS Micronet	P_{36}	0.8169	−0.0145	**0.0690**	0.0432	**0.7619**	0.0070	**0.0648**	**0.0405**
	AP_9	0.7567	0.0233	0.0804	0.0576	–	–	–	–
	GSSF	0.8048	0.0038	0.0718	0.0481	0.7364	0.0245	0.0691	0.0436
	STF	0.8189	**−0.0017**	0.0714	**0.0401**	0.7543	0.0314	0.0690	0.0412
	Proposed	**0.8279**	−0.0103	0.0691	0.0420	0.7574	**0.0034**	0.0666	0.0415
Average	P_{36}	**0.6456**	−0.0105	**0.0766**	**0.0434**	**0.6338**	−0.0204	**0.0728**	**0.0450**
	AP_9	0.5393	0.0336	0.0989	0.0618	–	–	–	–
	GSSF	0.5844	0.0135	0.0790	0.0475	0.6202	−0.0164	0.0731	0.0462
	STF	0.5889	0.0184	0.0832	0.0459	0.6116	**−0.0108**	0.0782	0.0461
	Proposed	0.6248	**0.0080**	0.0794	0.0454	0.6248	−0.0139	0.0772	0.0460

$$L_{D_F} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{D_F}(\mathrm{X}_d, \Theta_{\mathrm{DF}}) - \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{D_F}(\mathrm{X}, \Theta_{\mathrm{DF}}) + \lambda \frac{1}{N}\sum_{i=1}^{N} \left(\|\nabla_{\hat{\mathrm{X}}} \boldsymbol{D_F}(\hat{\mathrm{X}}, \Theta_{\mathrm{DF}})\|_2 - 1\right)^2 \tag{2}$$

where $\boldsymbol{D_F}(\cdot)$ represents the forward discriminator network function, and Θ_{DF} denotes its trainable parameters. In this study, given the instability of GAN training and the high heterogeneity of the input data, we adopted techniques from WGAN and WGAN-GP (Arjovsky et al., 2017; Gulrajani et al., 2017), utilizing the Wasserstein distance and a gradient penalty term in the loss function to enhance the stability and accuracy. Specifically, the first two terms in Eq. (2) represent the critic losses on the generated (fake) data and the label (real) data to measure the discriminator's performance. The third term is the gradient penalty, where $\hat{\mathrm{X}} = \varepsilon \mathrm{X} + (1-\varepsilon)\mathrm{X}_d$ is the interpolation between X_d and X using a random number ε in the range [0,1]. λ is the penalty coefficient, which is empirically set to 10.

3.2.2. The backward spatio-temporal constraint stage

The ideal downscaled SM should maintain an overall value distribution consistent with the LR SM at the target date (i.e., Y) and preserve the detail richness consistent with the HR SM at the reference date (i.e., X_t). To this end, the backward spatio-temporal constraint stage is introduced following the forward integrated fusion stage, which generates partial input observations in reverse from the downscaling result, forming a spatio-temporal cycle-consistent constraint with the forward input. Specifically, the backward stage utilizes a spatial degradation branch to regenerate P_{36} backward from the downscaling result and utilizes a temporal generator network to regenerate AP_9 backward from the downscaling result. These two processes can be written as:

$$\begin{cases} \mathrm{Y}^* = \mathbf{A}\mathrm{X}_d + \mathbf{N} \\ \mathrm{X}_t^* = \boldsymbol{G_B}(\mathrm{X}_d; \Theta_{\mathrm{GB}}) \end{cases} \tag{3}$$

where the upper term represents the spatial degradation branch of the downscaling result, which is typically formulated by the downsampling matrix **A** and the noise **N**, and is achieved through spatial resampling operations in the proposed network. Y^* is the reverse-generated P_{36} (i.e., P_{36}* in Fig. 3). $\boldsymbol{G_B}(\cdot)$ in the lower term represents the backward temporal generator network function used to implement temporal transformation on the downscaling result, with Θ_{GB} denoting the corresponding trainable parameters. X_t^* denotes the reverse-generated AP_9^{base} (i.e., $AP_9^{\mathrm{base}*}$ in Fig. 3). Thus, a cycle forms between the forward inputs $(\mathrm{Y},\ \mathrm{X}_t)$ and the backward outputs $(\mathrm{Y}^*, \mathrm{X}_t^*)$.

The backward discriminator network engages in an adversarial game with the backward temporal generator and the spatial degradation branch. It aims to differentiate between $(\mathrm{Y},\ \mathrm{X}_t)$ and $(\mathrm{Y}^*, \mathrm{X}_t^*)$ to indirectly guide the forward integrated fusion stage to generate downscaling results that maintain the valid spatio-temporal information of the forward inputs $(\mathrm{Y},\ \mathrm{X}_t)$. The loss function of the backward discriminator network can be formulated as:

$$L_{D_B} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{D_B}\left((\mathrm{Y}^*, \mathrm{X}_t^*), \Theta_{\mathrm{DB}}\right) - \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{D_B}((\mathrm{Y},\ \mathrm{X}_t), \Theta_{\mathrm{DB}}) + \lambda \frac{1}{N}\sum_{i=1}^{N} \left(\left\|\nabla_{(\hat{\mathrm{Y}}, \hat{\mathrm{X}}_t)} \boldsymbol{D_B}((\hat{\mathrm{Y}}, \hat{\mathrm{X}}_t), \Theta_{\mathrm{DB}})\right\|_2 - 1\right)^2 \tag{4}$$

where $\boldsymbol{D_B}(\cdot)$ and Θ_{DB} represent the backward discriminator network function and its respective trainable parameters, respectively. Similar to Eq. (2), the first two terms are the normal critic losses, and the third term is the gradient penalty.

3.2.3. Network structure

The proposed model is made up of two generator networks with identical structures and two discriminator networks with the same architectures. Fig. 4 depicts the network structure, using the forward generator and discriminator as examples. As shown in Fig. 4(a), the generator network begins with a convolutional (Conv) layer to transform the multi-source inputs (size of m × n) into the feature domain. Subsequently, two convolutional layers with a stride of 2 (S2) are employed for the feature encoding, reducing the feature scale to 1/4 while increasing the feature number by four times. Six residual blocks (He et al., 2016) are then utilized to extract deep features while simultaneously preserving shallow information and facilitating multi-

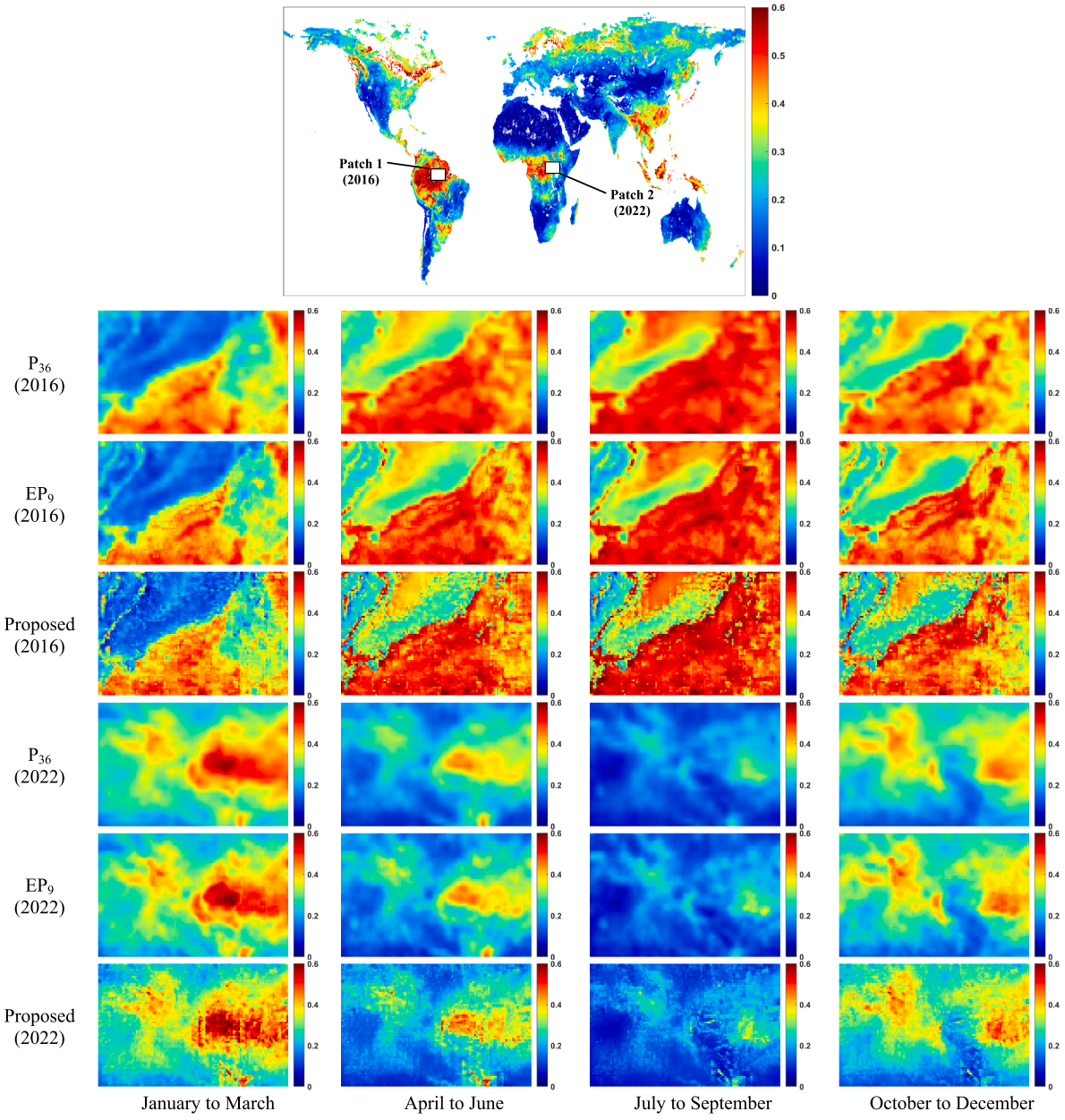
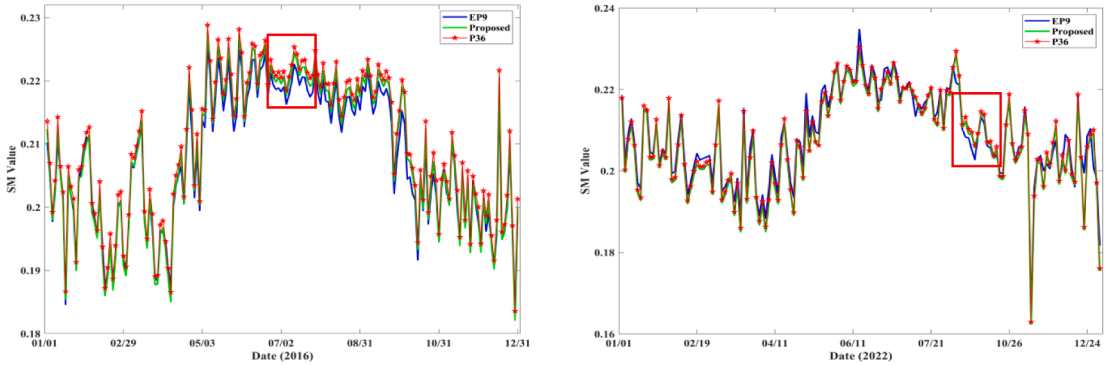


Fig. 11. Downscaling results for 2016 and 2022.



(a) SM curves for 2016 (b) SM curves for 2022

Fig. 12. Diurnal SM variation curves for 2016 and 2022.

Table 7
Quantitative evaluation of auxiliary parameter selection on the validation set.

Strategy	Idea data	ST (benchmark)	ST + TB	ST + NDVI	ST + LST	ST + Ref	ST + DEM	ST + LonLat	ST + TB + NDVI + Ref
R	1	0.7919	0.7947	0.7960	0.7912	0.7940	0.2916	0.7536	**0.8008**
bias	0	0.0029	0.0026	**0.0005**	0.0044	0.0027	0.0569	0.0082	0.0023
RMSE	0	0.0418	0.0409	0.0408	0.0415	0.0408	0.1073	0.0458	**0.0399**
ubRMSE	0	0.0405	0.0397	0.0396	0.0402	0.0395	0.0815	0.0431	**0.0386**

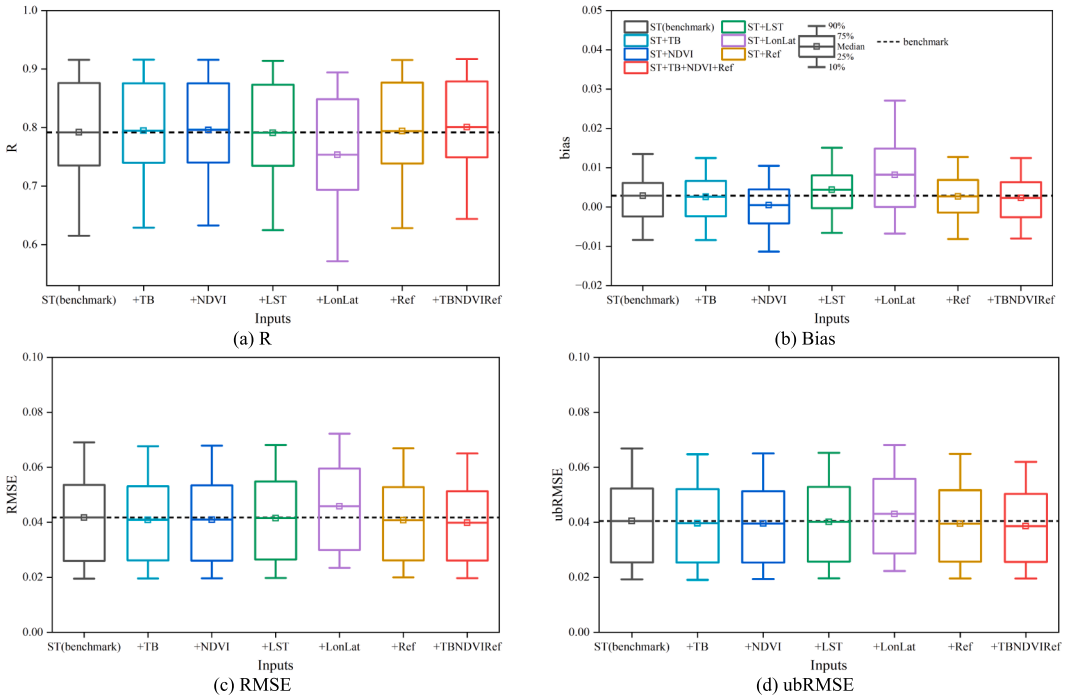


Fig. 13. Boxplots of the average quantitative evaluation results listed in Table 7.

dimensional information interaction. Subsequently, to revert the features to the initial image scale, two deconvolutional (DeConv) layers with a stride of 2 are employed for the feature decoding. Finally, a convolutional layer is utilized to transform the features back to the image domain, outputting the downscaling result.

Similar to Pan et al. (2021) and Jiang et al. (2022b), the generator network employs 7 × 7 convolutional kernels for transformation between the image and feature domains, 3 × 3 convolutional kernels for feature encoding and extraction, and 4 × 4 convolutional kernels for feature decoding. It incorporates batch normalization (Batch-norm) for feature normalization, and utilizes a rectified linear unit (ReLU) activation function in the preceding layers, followed by tanh activation function in the final layer.

As illustrated in Fig. 4(b), the discriminator network primarily relies on 4 × 4 convolutions with a stride of 2 for the feature encoding (Isola et al., 2017, Pan et al., 2021). The final layer employs a sigmoid activation function to transform the outputs into the value range of 0 to 1, thus facilitating the discrimination between real and fake samples. To ensure thorough optimization of the network parameters, no normalization layers are used in the discriminator network, to preserve the gradient of each input in a batch independently (Gulrajani et al., 2017).

Table 8
Quantitative evaluation of the different fusion strategies on the T1 test set.

Fusion strategy	Inputs	R	bias	RMSE	ubRMSE
I	P_{36} + TB + NDVI + Ref	0.8646	0.0033	0.0840	0.0839
II	$P_{36} + AP_9^{base} + P_{36}^{base}$	0.9307	0.0029	0.0610	0.0609
III	$P_{36} + AP_9^{base} + P_{36}^{base}$ + TB + NDVI + Ref	**0.9368**	**−0.0014**	**0.0583**	**0.0582**

3.2.4. Loss function of the generator networks

Deep learning networks calculate errors through loss functions and propagate these errors backward to update the trainable parameters. In the proposed network, the two discriminator networks optimize their parameters using the adversarial loss functions specified in Eq. (2) and Eq. (4). The two generator networks are trained together using a combined loss function. This combined loss function includes the regular adversarial loss between the generators and discriminators to steer the data distribution of the outputs (Gulrajani et al., 2017), as well as the content loss to ensure the pixel-level accuracy in the output results (Jiang et al., 2022b). The combined loss function can be formulated as follows:

$$L_G = L_{adv} + \alpha L_{num} + \beta L_{cyc} \tag{5}$$

where L_G represents the total loss function for the generator networks. L_{adv} denotes the adversarial loss between each generator and discriminator pair, which can be written as:

$$L_{adv} = -\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{D_F}(\mathrm{X}_d) - \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{D_B}\left(\mathrm{Y}^*, \mathrm{X}_t^*\right) \tag{6}$$

L_{num} and L_{cyc} in Eq. (5) are the content loss functions, which can be formulated as follows:

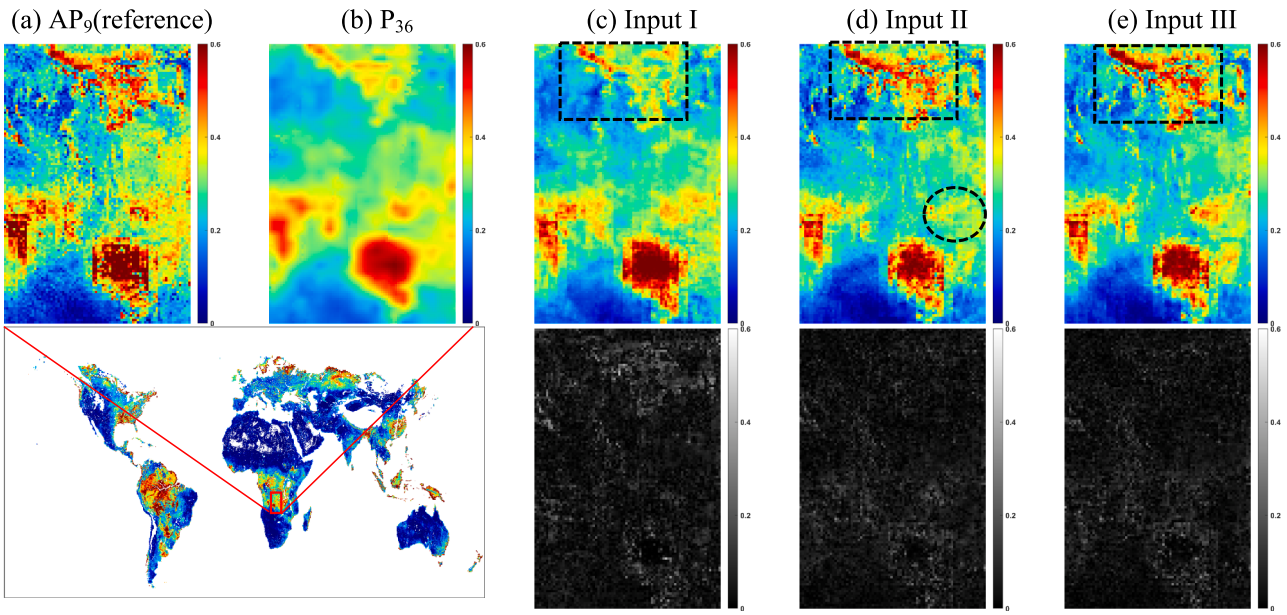


Fig. 14. The local composed SM downscaling results and the corresponding absolute residual images from 2015/04/15 to 2015/04/22.

Table 9
Effectiveness analysis of the proposed model components and loss functions on the validation set.

	Ideal data	w/o (7)	w/o (8)	w/o (7) & (8)	w/o discriminators	w/o backward stage	Proposed
R	1	0.4325	0.7985	0.1717	0.7987	0.7983	**0.8008**
ubRMSE	0	0.0698	0.0390	0.0910	0.0388	0.0389	**0.0386**

Table 10
Sensitivity analysis of weight parameters α and β on the validation set.

Effect of weight parameter α on the validation set with $\beta = 10$											
α	Ideal data	100	200	300	400	500	600	700	800	900	1000
R	1	0.7881	0.7976	0.7989	0.7994	**0.8008**	0.7992	0.7992	0.7974	0.8004	0.7988
ubRMSE	0	0.0399	0.0392	0.0388	0.0388	**0.0386**	0.0389	0.0389	0.0392	0.0388	0.0388
Effect of weight parameter β on the validation set with $\alpha = 500$											
β	Ideal data	5	10	15	20	25	30	35	40	45	50
R	1	0.8007	**0.8008**	0.7984	0.7995	0.7980	0.7951	0.7976	0.7947	0.7948	0.7894
ubRMSE	0	0.0388	**0.0386**	0.0388	0.0388	0.0390	0.0391	0.0391	0.0391	0.0392	0.0394

$$L_{num} = \frac{1}{N}\sum_{i=1}^{N} \|\mathrm{X}_d - \mathrm{X}\|_1 \tag{7}$$

$$L_{cyc} = \frac{1}{N}\sum_{i=1}^{N} \left\|(\mathrm{Y}^*, \mathrm{X}_t^*) - (\mathrm{Y}, \mathrm{X}_t)\right\|_1 \tag{8}$$

Eq. (7) denotes the numerical consistency loss term between the downscaling result and the HR SM label data, ensuring pixel-level consistency between them. Eq. (8) represents the spatio-temporal cycle-consistency loss between the outputs of the backward stage and the inputs of the forward stage, thus ensuring that the crucial spatio-temporal information from the input data is preserved in the downscaling result. α and β are two adjustable parameters that balance these terms, which are set as $\alpha = 500$ and $\beta = 10$.

The overall training process of the STC-CycleGAN network is summarized in Algorithm 1. After training the network, the global SM products and the corresponding auxiliary parameters are fed into the trained forward fusion generator to produce the global daily downscaling results.

Table 11
Effect of the kernel size on the validation set.

Kernel size	k1 = 3	k1 = 5	k1 = 9	k2 = 5	k2 = 7	k3 = 5	k3 = 7	k4 = 3	k5 = 3	k5 = 5	Proposed (k1,2,3,4,5 = 7,3,3,4,4)
R	0.7967	0.8005	0.7999	0.7987	0.7971	0.8001	0.7953	0.7983	0.8002	0.7973	**0.8008**
ubRMSE	0.0392	0.0389	0.0390	0.0389	0.0389	0.0389	0.0390	0.0389	0.0388	0.0390	**0.0386**

*Note: k1 represents the kernel size in the domain transformation layers, k2 represents the kernel size in the feature encoding layers, k3 represents the kernel size in the residual blocks, k4 represents the kernel size in the deconvolutional layers, and k5 represents the kernel size in the discriminator network.

Algorithm 1: STC-CycleGAN. We use default value of α=500, β=10, λ=10, N=64

for number of training iterations **do**

Select N input patches $\{(\mathrm{Y},\mathrm{Y}_t,\mathrm{X}_t,\mathrm{Z})^{(1)},\cdots(\mathrm{Y},\mathrm{Y}_t,\mathrm{X}_t,\mathrm{Z})^{(N)}\}$ from training data

Generate X_d by Eq. (1)

Generate $\mathrm{Y}^*,\mathrm{X}_t^*$ by Eq. (3)

$$L_G \leftarrow \frac{1}{N}\sum_{i=1}^{N}\left(-\boldsymbol{D_F}\left(\mathrm{X}_d\right)-\boldsymbol{D_B}\left((\mathrm{Y}^*,\mathrm{X}_t^*)\right)+\alpha\left\|\mathrm{X}_d-\mathrm{X}\right\|_1+\beta\left\|(\mathrm{Y}^*,\mathrm{X}_t^*)-(\mathrm{Y},\mathrm{X}_t)\right\|_1\right)$$

$$\Theta_{\mathrm{GF}},\Theta_{\mathrm{GB}} \leftarrow \mathrm{Adam}\left(\nabla_{\Theta_{\mathrm{GF}},\Theta_{\mathrm{GB}}}L_G,\Theta_{\mathrm{GF}},\Theta_{\mathrm{GB}}\right)$$

Select N fusion patches $\{\mathrm{X}_d^{(1)},\cdots\mathrm{X}_d^{(N)}\}$ from $\boldsymbol{G_F}$, N label patches $\{\mathrm{X}^{(1)},\cdots\mathrm{X}^{(N)}\}$

$$L_{DF} \leftarrow \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{D_F}\left(\mathrm{X}_d\right)-\boldsymbol{D_F}\left(\mathrm{X}\right)+\lambda\left(\left\|\nabla_{\hat{\mathrm{X}}}\boldsymbol{D_F}(\hat{\mathrm{X}})\right\|_2-1\right)^2\right)$$

$$\Theta_{\mathrm{DF}} \leftarrow \mathrm{Adam}\left(\nabla_{\Theta_{\mathrm{DF}}}L_{DF},\Theta_{\mathrm{DF}}\right)$$

Select N reverse-generated input patches, N input patches

$$L_{DB} \leftarrow \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{D_B}\left((\mathrm{Y}^*,\mathrm{X}_t^*)\right)-\boldsymbol{D_B}\left((\mathrm{Y},\mathrm{X}_t)\right)+\lambda\left(\left\|\nabla_{\hat{\mathrm{Y}},\hat{\mathrm{X}}_t}\boldsymbol{D_B}((\hat{\mathrm{Y}},\hat{\mathrm{X}}_t)\right\|_2-1\right)^2\right)$$

$$\Theta_{\mathrm{DB}} \leftarrow \mathrm{Adam}\left(\nabla_{\Theta_{\mathrm{DB}}}L_{DB},\Theta_{\mathrm{DB}}\right)$$

end

4. Experiments

For the hyperparameters of the network training, we initialized the learning rate at 0.0005, keeping this constant for the initial 75 epochs and then decaying to 0 over the final 75 epochs linearly. The Adam optimizer (Kingma and Ba, 2015) was employed to optimize the trainable parameters. In the following, the experimental datasets are first introduced. The qualitative and quantitative experimental results for the different periods and the in-situ site validations are then provided. Finally, a comparison with the SMAP enhanced 9-km SM (EP_9) product is presented.

4.1. Training, validation, and test sets and in-situ datasets

Table 3 details the experimental datasets used for training, validation, testing, and in-situ site verification. Given that the synthesis of the baseline HR-LR SM pair mainly relied on the AP_9 and P_{36} products from June 30, 2015, to July 7, 2015, and the prevalent issue of missing data in the daily SMAP SM products, we generated 27,920 blocks for the network training from May 1, 2015, to June 29, 2015. We then generated 1095 blocks for the validation from April 15, 2015, to April 30, 2015. As listed in Table 3, each training block was 40 × 40 in size and had a certain limit on the data missing rate. In the testing, to comprehensively verify the performance of the proposed approach at a global scale, we referred to Jiang et al. (2019) and divided the dates into T1 with AP_9 available (2015/04/15 to 2015/07/07) and T2 with AP_9 unavailable (2015/07/08 to 2015/10/31). Correspondingly, the in-situ SM validation was performed on the in-situ measurements from the various site networks during T1 and T2. Moreover, to thoroughly validate the robustness and long-temporal transferability, the results of the proposed approach were compared to the EP_9 product for the years 2016 and 2022.

4.2. Experiments on the T1 test set with AP_9 available

For the T1 test set with available AP_9 data from 2015/04/15 to 2015/07/07, four representative metrics are utilized here for the quantitative assessment: the correlation coefficient (R, unitless), bias (m^3/m^3), the root-mean-square error (RMSE, m^3/m^3), and the unbiased RMSE (ubRMSE, m^3/m^3) (Entekhabi et al., 2010b). To evaluate the performance of the proposed method, we compared the ANN (Liu et al., 2020), RF, ResNet (Zhao et al., 2022), and light gradient boosting machine (LGBM) (Luo et al., 2023) models from the generalized spatial-spectral fusion category and the STFM (Jiang et al., 2019), extended super-resolution convolutional neural network (ESRCNN) (Shao et al., 2019), VIPSTF (Yang et al., 2022), and GAN-based STFM (GANSTFM) (Tan et al., 2022) methods from the spatio-temporal fusion category. Fig. 5 presents the R and ubRMSE values for these methods. As shown in the figure, LGBM achieved a relatively high R value and lower ubRMSE value within the generalized spatial-spectral fusion category, while ESRCNN performed well in the spatio-temporal fusion category. Therefore, for brevity, LGBM and ESRCNN were selected as representative methods from each category for a more detailed comparison and analysis with the proposed method.

Table 4 presents the average quantitative evaluation values across 16 sets of test data for T1 with AP_9 as the reference, with the best performance marked in bold. As can be observed in Table 4, the generalized spatial-spectral fusion (GSSF)-based downscaling method achieves slightly higher accuracy than the P_{36} product, while the spatio-temporal fusion (STF)-based downscaling method outperforms the GSSF method across all metrics. The proposed generalized spatio-temporal-spectral integrated fusion-based downscaling method combines the respective advantages of these two methods and yields the best results across all metrics.

Fig. 6 shows the composed SM downscaling results on the T1 test set. Visually, the GSSF method produces a downscaling result that is spatially similar to the P_{36} product but with less detail than the AP_9 product. It also exhibits local underestimation of high SM values, as depicted by the rectangle in Fig. 6(c). The STF method produces richer spatial structures. However, it results in obvious local overestimation of the SM values, as shown in the red rectangle in Fig. 6(d). The proposed method mines the complementary information between the multiple auxiliary parameters and the multi-temporal SM products, yielding results that are closest to the reference, as shown in Fig. 6(e). Regarding the R and ubRMSE metrics, the 16-day composite operation during the T1 period enhances the correlation between the downscaling results and

the reference data, and reduces the performance gaps among the various downscaling methods. However, the proposed method still achieves the best performance in both the R and ubRMSE metrics by a significant margin, which is consistent with its visual performance. Fig. 7 shows the bias between the various downscaling results and the reference in Fig. 6. As in Fig. 7, both GSSF and STF result in notable local overestimation and underestimation of SM, especially in South America and Central Africa. In contrast, the proposed method exhibits minimal bias both globally and locally, demonstrating its superiority.

Three representative sub-regions in Fig. 6(a) are zoomed in on in Fig. 8. In these patches, the GSSF method exhibits pronounced underestimation of the high SM values. In addition, because of the absence of SM spatio-temporal information and the common pixel-wise learning mode in the generalized spatial-spectral fusion approaches, the GSSF result shows grainy spatial details. These details are only slightly richer than those in the P_{36} product, highlighting a notable disparity when compared to the AP_9 product. The STF method, which relies solely on the spatio-temporal information of SM, effectively preserves the spatial characteristics of SM. However, it is susceptible to potential changes in SM values between the target and baseline dates, leading to localized distortion, as shown in Patch 3. The proposed method incorporates both the spatio-temporal features of SM and the generalized spectral features from auxiliary parameters, allowing it to obtain SM data with rich details and stable quality.

4.3. Experiments on the T2 test set with AP_9 unavailable

To compare the temporal extrapolation capabilities of the different methods, experiments were conducted on the T2 test set. This time, due to the unavailability of AP_9 data, we downsampled the fusion results to 36 km and used P_{36} as a reference to quantitatively evaluate the performance at the low resolution. In addition, we included two metrics: mean (m^3/m^3) and spatial frequency (SF, cycles/km), which do not require a reference, to assess the numerical distribution and spatial structure of the downscaling results at the 9-km resolution.

Table 5 lists the average quantitative evaluation results for 115 test data pairs for the T2 period, with the best-performing values highlighted in bold for each metric. It can be observed that, when using P_{36} as a reference for the quantitative evaluation, the performance gaps among the three methods are smaller than those in Table 4, with GSSF showing the best performance. The lower quantitative results of the proposed method when compared to the GSSF method may be due to the inherent accuracy differences between the AP_9 and P_{36} products. Nevertheless, the proposed results maintain a high correlation with both AP_9 and P_{36}. For the mean and SF metrics, all three methods produce mean values that are lower than that of the P_{36} product, while the proposed method has a mean that is closest to P_{36}, and the highest SF value.

Fig. 9 displays the composed downscaling results for the last 16 days of the T2 period (2015/10/16 to 2015/10/31). As depicted in the figure, the GSSF downscaling result is generally consistent with the P_{36} product visually, but the mean and SF values are lower than those of P_{36}, indicating some underestimation of SM values and a significant deficiency in spatial resolution enhancement. The STF method inputs multi-temporal SM data and produces downscaling results that are rich in spatial details. However, it tends to overestimate SM values in regions such as South America and underestimate them in areas such as North America, resulting in an overall low mean SM value. The proposed method fully mines the spatio-temporal characteristics from the SM products and the real-time surface state characteristics from the auxiliary parameter data, and produces visually superior results, as shown in Fig. 9(d).

Fig. 10 further analyzes three representative patches selected from Fig. 9. Consistent with the global result, the results of the GSSF method lack spatial details and underestimate high SM values, resulting in low mean and SF values. Although the STF method achieves the highest SF values in some patches, it shows significant overestimation in Patch 1 and considerable underestimation in Patch 3. In the results of the proposed method, the SM distribution aligns with that of the P_{36} product, and the proposed method generates rich and reliable spatial details, demonstrating its favorable temporal extrapolation capability.

4.4. In-situ site validation for the T1 and T2 periods

Considering the limited availability of the AP_9 product, and to comprehensively evaluate the performance of the proposed method, quantitative assessments were conducted for both the T1 and T2 periods using the in-situ SM datasets listed in Table 2. Due to the disparities in the sensing devices and measurement depths between the satellite and in-situ SM data, as well as the spatial limitation of the in-situ station coverage not completely matching the corresponding satellite SM pixels, we focus on the temporal variations in SM that are less impacted by these factors. Furthermore, since a satellite SM pixel may encompass multiple in-situ sites, the arithmetic mean of these sites was calculated to assess the satellite pixels.

In Table 6, the left side presents the in-situ site validation results for the T1 period, while the right side is the T2 period. The best-performing results are indicated in bold, with the second-best results underlined. To simplify the tables and highlight the performance of the proposed method across different site networks, we grouped the ISMN networks with fewer sites, except for USCRN and SCAN, into the 'Others' category. As listed in Table 6, P_{36} performs the best in most metrics, while AP_9 is inferior to P_{36} in all the metrics during the T1 period. The accuracy gap between AP_9 and P_{36} stems from the inherent differences in their estimations (Entekhabi et al., 2010a). Due to the lower accuracy of AP_9 compared to P_{36}, and as the proposed network was trained using the AP_9 product as label data, the in-situ validation metrics of the proposed method are slightly inferior to those of P_{36} but significantly better than those of AP_9. This indicates that the proposed method not only improves the spatial resolution of P_{36} but also effectively maintains its high accuracy.

Since the R and ubRMSE metrics are less affected by the SM value on individual dates, they can better reflect the correlation between the temporal series of downscaling results and the in-situ measurements. As shown in Table 6, for most site networks, the R values of the proposed method are slightly lower than those of the P_{36} product (0.021 lower on average during the T1 period and 0.009 lower on average during the T2 period), but higher than those of the other methods. The ubRMSE values of the proposed method are only about 0.001 to 0.002 higher than those of the P_{36} product across the different site networks during both T1 and T2 periods, demonstrating the reliability of the proposed method.

4.5. Comparison with the EP_9 product

To validate the capability of the proposed approach in generating long-term products, the result of the proposed method was compared with the EP_9 product in 2016 and 2022. The EP_9 product was developed by the SMAP team to address the absence of the AP_9 product. It was generated using the Backus-Gilbert interpolation algorithm based on the 36-km passive microwave products (Chan et al., 2018). During the proposed product production, residual correction (Jiang et al., 2019) was performed to alleviate the influence of the long intervals between the target and the baseline dates and maintain the high accuracy of the P_{36} product.

Fig. 11 presents a representative patch of the P_{36} product, the EP_9 product, and the downscaling results of the proposed method, composed at three-month intervals for the years 2016 and 2022. As shown in the figure, the EP_9 product appears close to the P_{36} product, without a discernible increase in spatial details. The proposed method can effectively maintain the overall SM distribution of the P_{36} product while increasing its spatial detail information. Moreover, the SM in Patch 1 increases from spring to autumn in 2016, with a slight decrease in winter, whereas the SM in Patch 2 undergoes a transition from high values to low values and back to high values in 2022. These SM changes

are clearly presented in the results of the proposed method, showcasing the reliable temporal transferability of the proposed method. Fig. 12 presents the daily global average SM curves for the proposed method, the EP_9 product, and the P_{36} product. It is clear that the curves of the proposed method align more closely with the P_{36} curve than the EP_9 curves do, demonstrating the high accuracy of the proposed method.

5. Discussions

In this section, we further analyze the performance of the proposed method through auxiliary parameter selection, fusion strategy comparison, and sensitivity analysis.

5.1. Auxiliary parameter selection

Unlike the multi-parameter fusion (i.e., generalized spatial-spectral fusion)-based downscaling methods that rely on a certain number of auxiliary parameters to guarantee a good performance, the proposed method additionally inputs the spatio-temporal SM information and mitigates the requirement for auxiliary parameters. Consequently, it is essential to screen the auxiliary parameters to identify the positive ones. To this end, we conducted ablation experiments on the validation set using the spatio-temporal fusion input (P_{36}, P_{36}^{base}, AP_9^{base}) as the benchmark. Table 7 lists the quantitative results for each auxiliary parameter, highlighting the best performance in bold for each metric. It can be seen that, except for DEM, the auxiliary parameters considered in this study do not significantly reduce the experimental performance. Fig. 13 presents boxplots of the quantitative evaluation results, except for (ST + DEM). It is evident that TB, NDVI, and Ref exhibit positive effects across all metrics, and their combination (ST + TB + NDVI + Ref) outperforms any single auxiliary parameter in the R, RMSE, and ubRMSE metrics. Therefore, this combination was adopted in the proposed method, while all the auxiliary parameters were utilized in the GSSF comparison method.

5.2. Comparison of different fusion strategies

In order to explore the performance of the proposed STC-CycleGAN network under different input combinations, we considered the following fusion strategies: (I) generalized spatial-spectral fusion with the input of the LR SM and HR auxiliary parameters (i.e., P_{36} + TB + NDVI + Ref); (II) spatio-temporal fusion with only the SM products as input (i.e., $P_{36} + AP_9^{base} + P_{36}^{base}$); and (III) generalized spatio-temporal-spectral fusion with both the auxiliary parameters and SM products as input (i.e., $P_{36} + AP_9^{base} + P_{36}^{base}$ + TB + NDVI + Ref).

Table 8 lists the quantitative evaluation results on the T1 test set under the various inputs. As listed in the table, both Inputs I and II yield acceptable results, and their combination (i.e., III) exhibits the best performance across all the metrics. This indicates the complementarity between the auxiliary parameters and the spatio-temporal information of SM and demonstrates the superiority of the proposed generalized spatio-temporal-spectral integrated fusion strategy.

Fig. 14 shows the composed SM downscaling results and the absolute residual images between the reference and the downscaling results from 2015/04/15 to 2015/04/22. As depicted in Fig. 14, Input I produces downscaling results that are consistent with P_{36}, but with limited details, as in the black rectangle. Input II generates downscaling results with rich spatial details but some localized distortions, as shown in the black oval. This phenomenon is likely due to the SM changes between the target and baseline date, and would be more significant with an increased date interval. Input III combines the respective advantages of the above strategies and produces downscaling results that are closest to the reference.

5.3. Effectiveness and sensitivity analysis

The proposed model comprises multiple generator and discriminator networks, along with various loss functions. To verify their contributions, ablation experiments were conducted on the validation set using the R and ubRMSE metrics. Table 9 presents the average quantitative evaluation results, indicating that the content loss, which is made up of the numerical consistency loss (Eq. (7) and the spatio-temporal cycle-consistency loss (Eq. (8), significantly impacts the model performance, with the numerical consistency loss (Eq. (7) being the most important. Removing either the discriminator networks or the backward spatio-temporal constraint stage reduces the quantitative accuracy, while using both in the proposed method yields the best performance.

In addition, the loss function includes three adjustable weight parameters: α, β, and λ, which are crucial for the proposed downscaling task. Since the gradient penalty weight λ is typically empirically set to 10 for stable training performance (Gulrajani et al., 2017), a sensitivity analysis of α and β was conducted on the validation set. Specifically, following Pan et al. (2021), we fixed β at 10 and varied α within the range [100, 1000] in increments of 100. Table 10 shows that the model performs best with $\alpha = 500$. Similarly, $\beta = 10$ is found to achieve the best performance. Therefore, α was set to 500 and β to 10 in this study. Furthermore, Table 11 illustrates the effect of the kernel size on the validation set. This shows that, except for k1, which is used for domain transformation in the first and last layers of the generator network, smaller kernel sizes (3 or 4) for the other kernels (k2, k3, k4, k5) generally lead to a better performance. The proposed settings achieve the best results.

6. Conclusions

In this paper, by broadly defining the spectral characteristics of geographic objects, we have proposed a generalized spatio-temporal-spectral integrated fusion framework to integrate the generalized spectral features of auxiliary parameters with the spatio-temporal features of the target parameter to achieve high-precision downscaling of the target surface parameter. Specifically, we used the SM parameter as a case study and developed the STC-CycleGAN network to downscale SM using this integrated fusion strategy. The developed STC-CycleGAN network is made up of a forward integrated fusion stage and a backward spatio-temporal constraint stage, combined with a spatio-temporal cycle-consistency loss function. Qualitative, quantitative, and in-situ site evaluations showed that the proposed method can integrate the complementary characteristics from the multi-source auxiliary parameters and SM data to obtain downscaling results that are consistent with the LR SM data and rich in spatial details.

In the future, the proposed method could be further improved and developed from the following aspects. From the perspective of data, it will be of great significance to import more data sources. For example, on the one hand, active microwave SM data with a higher resolution, such as Sentinel synthetic aperture radar (SAR) data, could be introduced to achieve higher-resolution downscaling for fine regional monitoring. On the other hand, the assimilated SM data of dynamic models with higher continuity could be incorporated to achieve seamless downscaling of SM for spatio-temporally continuous monitoring. From the perspective of the network, it would be feasible to add real-time P_{36} products to the network training through unsupervised or fine-tuning strategies, to further enhance the stability and accuracy of the downscaling model. From an application perspective, there is huge potential to extend the proposed generalized spatio-temporal-spectral integrated fusion framework to other surface parameters, such as LST and precipitation.

CRediT authorship contribution statement

Menghui Jiang: Writing – original draft, Methodology. **Huanfeng**

Shen: Writing – review & editing, Conceptualization. **Jie Li:** Methodology. **Liangpei Zhang:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 42301531, Grant 42130108 and Grant 62071341, the Key R&D Program of Hubei Province, China, under Grant 2023BCA003, and the China Postdoctoral Science Foundation (No. 2023M732684).

References

Aggarwal, A., Mittal, M., Battineni, G., 2021. Generative adversarial network: An overview of theory and applications. Int. J. Inf. Manag. Data Insights. 1 (1), 100004.
Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks. Proc. Int. Conf. Mach. Learn. 70, 214–223.
Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J., Anderson, C., 2007. Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). Geophys. Res. Lett. 34, L20401.
Chan, S., Bindlish, R., O'Neill, P.E., Jackson, T.J., Njoku, E., Dunbar, R.S., Chaubell, J., Peipmeier, J., Yueh, S., Entekhabi, D., Colliander, A., Chen, F., Cosh, M.H., Caldwell, T., Walker, J., Berg, A., McNairn, H., Thibeault, M., Martinez-Fernandez, J., Udall, F., Seyfried, M.S., Bosch, D.D., Starks, P., Holyfield Collins, C., Prueger, J.H., 2018. Development and assessment of the SMAP enhanced passive soil moisture product. Remote Sens. Environ. 204, 931–941.
Chauhan, N.S., Miller, S., Ardanuy, P., 2003. Spaceborne soil moisture estimation at high resolution: a microwave-optical/ir synergistic approach. Int. J. Remote Sens. 24 (22), 4599–4622.
Chen, N., He, Y., Zhang, X., 2017. Nir-red spectra-based disaggregation of SMAP soil moisture to 250 m resolution based on Oznet in Southeastern Australia. Remote Sens. 9 (1), 51.
Cheng, Q., Liu, H., Shen, H., Wu, P., Zhang, L., 2017. A Spatial and temporal nonlocal filter-based data fusion method. IEEE Trans. Geosci. Remote Sens. 55 (8), 4476–4488.
Das, N.N., Entekhabi, D., Njoku, E.G., 2010. An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval. IEEE Trans. Geosci. Remote Sens. 49, 1504–1512.
Dorigo, W.A., Gruber, A., de Jeu, R., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R., 2015. Evaluation of the ESA CCI soil moisture product using ground-based observations. Remote Sens. Environ. 162, 380–395.
Dorigo, W., Xaver, A., Vreugdenhil, M., Gruber, A., HegyiováA, A., Sanchis-Dufau, A., 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 J. 12 (3).
Draper, C., Reichle, R., 2015. The impact of near-surface soil moisture assimilation at subseasonal, seasonal, and inter-annual timescales. Hydrol. Earth Syst. Sci. 19 (12), 4831.
Engman, E.T., Chauhan, N., 1995. Status of microwave soil moisture measurements with remote sensing. Remote Sens. Environ. 51 (1), 189–198.
Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., et al., 2010a. The soil moisture active passive (smap) mission. Proc. IEEE 98 (5), 704–716.
Entekhabi, D., Reichle, R.H., Koster, R.D., Crow, W.T., 2010b. Performance metrics for soil moisture retrievals and application requirements. J. Hydrometeorol. 11, 832–840.
Fang, B., Lakshmi, V., Jackson, T.J., Bindlish, R., Colliander, A., 2019. Passive/active microwave soil moisture change disaggregation using SMAPVEX12 data. J. Hydrol. 574, 1085–1098.
Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the landsat and MODIS surface reflectance: Predicting daily landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44, 2207–2218.
Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C., 2017. Improved training of wasserstein GANs, in Proc. 30st Int. Conf. Neural Inf. Process. Syst., 5769–5779.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 770–778.
Hirschi, M., Seneviratne, S.I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orlowsky, B., Stepanek, P., 2011. Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. Nat. Geosci. 4, 17–21.
Hoehn, D.C., Niemann, J.D., Green, T.R., Jones, A.S., Grazaitis, P.J., 2017. Downscaling soil moisture over regions that include multiple coarse-resolution grid cells. Remote Sens. Environ. 199, 187–200.
Hu, F., Wei, Z., Zhang, W., Dorjee, D., Meng, L., 2020. A spatial downscaling method for SMAP soil moisture through visible and shortwave-infrared remote sensing data. J. Hydrol. 590, 125360.
Huang, S., Zhang, X., Wang, C., Chen, N., 2023. Two-step fusion method for generating 1 km seamless multi-layer soil moisture with high accuracy in the Qinghai-Tibet plateau. ISPRS J. Photogramm. Remote Sens. 197, 346–363.
Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. Proc. IEEE Conf. Comput. Vis. Pattern Recognit.1125–1134.
Jiang, M., Shen, H., Li, J., 2022a. Cycle GAN Based Heterogeneous Spatial-Spectral Fusion for Soil Moisture Downscaling, in Proc. IEEE Int. Geosci. Remote Sens. Symp. 4819-4822.
Jiang, H., Shen, H., Li, X., Zeng, C., Liu, H., Lei, F., 2019. Extending the SMAP 9-km soil moisture product using a spatio-temporal fusion model. Remote Sens. Environ. 231, 111224.
Jiang, M., Shen, H., Li, J., 2022b. Deep-learning-based spatio-temporal-spectral integrated fusion of heterogeneous remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–15.
Jing, Y., Lin, L., Li, X., Li, T., Shen, H., 2022. An attention mechanism based convolutional network for satellite precipitation downscaling over China. J. Hydrol. 613, 128388.
Jing, Y., Shen, H., Li, X., Wu, J., Qiu, Z., 2024. Spatial Downscaling of Remote Sensing Parameters from the Perspective of Data Fusion. Geomatics Inf. Sci. Wuhan Univ. 49 (2), 175–189. In Chinese.
Justice, C., Townshend, J.R., Vermote, E., Masuoka, E., Wolfe, R., Saleous, N., Roy, D., Morisette, J., 2002. An overview of MODIS Land data processing and product status. Remote Sens. Environ. 83, 3–15.
Kerr, Y.H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F.O., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., 2010. The SMOS mission: new tool for monitoring key elements of the global water cycle. Proc. IEEE. 98 (5), 666–687.
Kim, S., Balakrishnan, K., Liu, Y., Johnson, F., Sharma, A., 2017. Spatial disaggregation of coarse soil moisture data by using high-resolution remotely sensed vegetation products. IEEE Geosci. Remote Sens. Lett. 14, 1604–1608.
Kingma, D., Ba, J., 2015. Adam: A method for stochastic optimization. Proc. Int. Conf. Learn. Represent. 1–41.
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. Adv. Water Resour. 141, 103601.
Luo, Q., Liang, Y., Guo, Y., Liang, X., Ren, C., Yue, W., Zhu, B., Jiang, X., 2023. Enhancing Spatial Resolution of GNSS-R Soil Moisture Retrieval through XGBoost Algorithm-Based Downscaling Approach: A Case Study in the Southern United States. Remote Sens. 15 (18), 4576.
Mahfouf, J.F., Bergaoui, K., Draper, C., Bouyssel, F., Taillefer, F., Taseva, L., 2009. A comparison of two off-line soil analysis schemes for assimilation of screen level observations. J. Geophys. Res. Atmos. 114.
Martínez-Fernández, J., González-Zamora, A., Sánchez, N., Gumuzzio, A., Herrero-Jiménez, C.M., 2016. Satellite soil moisture for agricultural drought monitoring: assessment of the SMOS derived Soil Water Deficit Index. Remote Sens. Environ. 177, 277–286.
Narayan, U., Lakshmi, V., Jackson, T.J., 2006. High-resolution change estimation of soil moisture using L-band radiometer and radar observations made during the SMEX02 experiments. IEEE Trans. Geosci. Remote Sens. 44 (6), 1545–1554.
Pan, J., Dong, J., Liu, Y., Zhang, J., Ren, J., Tang, J., Tai, Y.W., Yang, M.H., 2021. Physicsbased generative adversarial models for image restoration and beyond. IEEE Trans. Pattern Anal. Mach. Intell. 43 (7), 2449–2462.
Peng, J., Loew, A., Zhang, S., Wang, J., Niesel, J., 2016. Spatial downscaling of satellite soil moisture data using a vegetation temperature condition index. IEEE Trans. Geosci. Remote Sens. 54, 558–566.
Peng, J., Loew, A., Merlin, O., Verhoest, N.E.C., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. Rev. Geophys. 55, 341–366.
Petropoulos, G.P., Ireland, G., Barrett, B., 2015. Surface soil moisture retrievals from remote sensing: current status, products & future trends. Phys. Chem. Earth Parts a/b/c 83–84, 36–56.
Piles, M., Entekhabi, D., Camps, A., 2009. A change detection algorithm for retrieving high-resolution soil moisture from SMAP radar and radiometer observations. IEEE Trans. Geosci. Remote Sens. 47 (12), 4125–4131.
Ranney, K.J., Niemann, J.D., Lehman, B.M., Green, T.R., Jones, A.S., 2015. A method to downscale soil moisture to fine resolutions using topographic, vegetation, and soil data. Adv. Water Resour. 76, 81–96.
Reichle, R.H., Entekhabi, D., McLaughlin, D.B., 2001. Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach. Water Resour. Res. 37, 2353–2364.
Renzullo, L.J., van Dijk, A.I.J.M., Perraud, J.-M., Collins, D., Henderson, B., Jin, H., McJannet, D.L., 2014. Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment. J. Hydrol. 519, 2747–2762.
Sabaghy, S., Walker, J.P., Renzullo, L.J., Jackson, T.J., 2018. Spatially enhanced passive microwave derived soil moisture: capabilities and opportunities. Remote Sens. Environ. 209, 551–580.
Sahoo, A.K., Lannoy, G.J.D., Reichle, R.H., Houser, P.R., 2013. Assimilation and downscaling of satellite observed soil moisture over the little river experimental watershed in Georgia, USA. Adv. Water Resour. 52, 19–33.
Sánchez-Ruiz, S., Piles, M., Sánchez, N., Martínez-Fernández, J., Vall-Llossera, M., Camps, A., 2014. Combining SMOS with visible and near/shortwave/thermal infrared satellite data for high resolution soil moisture estimates. J. Hydrol. 516, 273–283.

Shangguan, Y., Min, X., Shi, Z., 2023. Inter-comparison and integration of different soil moisture downscaling methods over the Qinghai-Tibet plateau. J. Hydrol. 617, 129014.
Shao, Z., Cai, J., Fu, P., Hu, L., Liu, T., 2019. Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. Remote Sens, Environ, p. 235.
Shen, H.F., Huang, L.W., Zhang, L.P., Wu, P.H., Zeng, C., 2016. Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: a 26-year case study of the city of Wuhan in China. Remote Sens. Environ. 172, 109–125.
Srivastava, P.K., 2017. Satellite soil moisture: review of theory and applications in water resources. Water Resour. Manag. 31, 3161–3176.
Tan, Z., Gao, M., Li, X., Jiang, L., 2022. A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network. IEEE Trans. Geosci, Remote Sens, p. 60.
Tang, Y.J., Wang, Q.M., Tong, X.H., Atkinson, P.M., 2021. Integrating spatio-temporalspectral information for downscaling Sentinel-3 OLCI images. ISPRS J. Photogramm. Remote Sens. 180, 130–150.
Tewes, A., Thonfeld, F., Schmidt, M., 2015. Using RapidEye and MODIS data fusion to monitor vegetation dynamics in semi-arid rangelands in South Africa. Remote Sens. 7, 6510–6534.
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 Sens. Environ. 225, 30–44.
Wu, A.N., Biljecki, F., 2023. InstantCITY: Synthesising Morphologically Accurate Geospatial Data for Urban Form Analysis, Transfer, and Quality Control. ISPRS J. Photogramm. Remote Sens. 195, 90–104.
Xiao, Y., Zeng, C., Shen, H., 2021. Soil moisture downscaling method combining parameter statistics and spatio-temporal fusion. Remote Sens. Technol. Appl. 36 (5), 1033–1043. In Chinese.
Xu, M., Yao, N., Yang, H., Xu, J., Hu, A., de Goncalves, L.G.G., Liu, G., 2022. Downscaling SMAP soil moisture using a wide & deep learning method over the continental United States. J. Hydrol. 609, 127784.
Yang, H., Wang, Q., Zhao, W., Tong, X., Atkinson, P.M., 2022. Reconstruction of a global 9 km, 8-day SMAP surface soil moisture dataset during 2015–2020 by spatiotemporal fusion. J. Remote Sens. 9871246.
Yang, J., Yang, Q., Hu, F., Shao, J., Wang, G., 2024. A climate-adaptive transfer learning framework for improving soil moisture estimation in the Qinghai-Tibet Plateau. J. Hydrol. 630, 130717.
Zhang, Y., Liang, S., Zhu, Z., Ma, H., He, T., 2022. Soil moisture content retrieval from Landsat 8 data using ensemble learning. ISPRS J. Photogramm. Remote Sens. 185, 32–47.
Zhao, H., Li, J., Yuan, Q., Lin, L., Yue, L., Xu, H., 2022. Downscaling of soil moisture products using deep learning: comparison and analysis on Tibetan plateau. J. Hydrol. 607, 127570.