

Dataset Development of the Forecasting Global Surface Soil Moisture Using Multi-scenario Integration Methodology (2015–2100)

Yang, F. Liu, Y. X. Y.*

Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Abstract: Soil moisture is a key land surface element to express the effects of global climate change. In order to develop a reliable global future multi-scenario surface soil moisture fusion dataset, this study firstly utilized the Enhanced Triple Collocation (ETC) to evaluate the accuracy of 22 CMIP6 (Coupled Model Intercomparison Project Phase 6) soil moisture datasets, and obtained the random error standard deviation (RESL) and correlation coefficient (CC) to select the qualified Earth System Model datasets. Secondly, 9 of the Earth System Model datasets were fused based on the normalized weighting of RESL and CC. Finally, the accuracy of the fused data was verified by the evaluation of the measured data at the stations, and the results showed that the fused soil moisture data could effectively describe the pattern of global surface soil moisture. The dataset includes: (1) global monthly 0.5° resolution soil moisture data of SSP1-2.6, SSP2-4.5, and SSP5-8.5. (2) *in situ* measurements from 4 networks, which are NAQU, REMEDHUS, SMOSMANIA, and TWENTE. The dataset is archived in .tif, .shp and .csv formats, and consists of 3,124 data files with data size of 829 MB (compressed into 4 files with data size of 770 MB).

Keywords: surface soil moisture; future multi-scenario; global; fusion

DOI: <https://doi.org/10.3974/geodp.2025.02.03>

Dataset Availability Statement:

The dataset supporting this paper was published and is accessible through the *Digital Journal of Global Change Data Repository* at: <https://doi.org/10.3974/geodb.2024.11.10.V1>.

1 Introduction

Soil moisture refers to water content stored in the unsaturated soil zone, and is a physical quantity that indicates the degree of dryness or wetness of the soil. Soil moisture is the link between the conversion of atmospheric water, surface water, plant water and groundwater,

Received: 03-12-2024; **Accepted:** 13-05-2025; **Published:** 25-06-2025

Foundations: Ministry of Science and Technology of P. R. China (2022YFF0711603); National Natural Science Foundation of China (42101475)

***Corresponding Author:** Liu, Y. X. Y., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, lyxy@lreis.ac.cn

Data Citation: [1] Yang, F., Liu, Y. X. Y. Dataset development of the forecasting global surface soil moisture using multi-scenario integration methodology (2015–2100) [J]. *Journal of Global Change Data & Discovery*, 2025, 9(2): 155–162. <https://doi.org/10.3974/geodp.2025.02.03>.

[2] Yang, F., Liu, Y. X. Y. Forecasting global surface soil moisture dataset using multi-scenario integration methodology (2015–2100) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2024. <https://doi.org/10.3974/geodb.2024.11.10.V1>.

and the carrier of material transfer, playing an indispensable role in water and material cycles^[1-3]. Meanwhile, in the context of global warming caused by greenhouse gas emissions from human activities, extreme weather events (e.g., high-temperature heat waves, floods, and droughts) are frequently occurred. Permafrost in high-latitude and high-altitude regions is ablated and degraded. And the spatial and temporal distribution pattern of soil moisture changes significantly, thus affecting the spatial and temporal evolution of the climate and ecosystems^[4,5]. Therefore, obtaining reliable soil moisture data on a global scale and over a long time series is of great scientific value and strategic significance for climate change research, water cycle analysis, vegetation growth monitoring, early warning of droughts and floods, and food security^[6-8].

The Coupled Model Intercomparison Project Phase 6 (CMIP6), initiated by the World Climate Research Program's Working Group on Coupled Modelling, has provided a rich set of coupled climate models for understanding future climate change^[9,10]. The CMIP6 climate models have now become the basis for the United Nations Intergovernmental Panel on Climate Change (IPCC) to prepare a report on future climate change^[11]. A series of scenarios depicting different future socio-economic development patterns and atmospheric greenhouse gas concentrations, called Shared Socioeconomic Pathways (SSPs), have been constructed around CMIP6.

In order to enhance the trend representativeness of the data and reduce anomaly bias, mainstream studies usually simply calculate the ensemble mean of multiple future scenarios to characterize the global future spatial-temporal distribution of soil moisture, but they also face the risk of being dominated by the data with the largest errors. In addition, although Earth System Models (ESM) continue to undergo optimization iterations, systematic errors and uncertainties still exist at present^[12]. In addition, there is an urgent need to replace simple ensemble averaging with rational complementary and optimal combinatorial fusion reconstruction of multi-source information to improve the data reliability of soil moisture for future scenarios^[13,14].

This dataset is weighted and fused based on the accuracy spatial distribution characteristics on the basis of evaluating multiple CMIP6 soil moisture. And it can provide scientific data support for the study and analysis of the spatial-temporal evolution pattern of the future surface water cycle.

2 Metadata of the Dataset

The metadata of Forecasting global surface soil moisture dataset using multi-scenario integration methodology (2015–2100)^[15] is summarized in Table 1. It includes the dataset full name, short name, authors, year of the dataset, temporal resolution, spatial resolution, data format, data size, data files, data publisher, and data sharing policy, etc.

3 Methods

3.1 Data Sources

The CMIP6 surface (0–10 cm) soil moisture data^[11] used in this dataset consisted of 3 main SSPs, which are SSP1-2.6 (Sustainable Pathway), SSP2-4.5 (Medium Pathway), and SSP5-8.5 (Fossil Fuel Burning Pathway). The soil moisture data involved in the evaluation and fusion were derived from 22 Earth system models, namely ACCESS-CM2, BCC-CSM2-MR, CAMS-CSM1-0, CanESM5-CanOE, CESM2, CMCC-CM2-SR5, CMCC-ESM2, CNRM-CM6-1, and CNRM-CM6-1-HR, CNRM-ESM2-1, EC-Earth3-Veg-LR,

Table 1 Metadata summary of the Forecasting global surface soil moisture dataset using multi-scenario integration methodology (2015–2100)

Items	Description
Dataset full name	Forecasting global surface soil moisture dataset using multi-scenario integration methodology (2015–2100)
Dataset short name	MonthlyinsituData
Authors	Yang, F., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, yangf@igsnr.ac.cn Liu, Y. X. Y., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, lyxy@lreis.ac.cn
Geographical region	Global (90°N–60°S)
Year	2015–2100
Temporal resolution	monthly
Spatial resolution	0.5°×0.5°
Data format	.tif, .shp, .csv
Data size	770 MB (compressed)
Data files	(1) Global monthly 0.5° resolution soil moisture data if SSP1-2.6, SSP2-4.5, and SSP5-8.5; (2) <i>In situ</i> measurements from 4 networks, which are NAQU, REMEDHUS, SMOSMANIA, and TWENTE
Foundation	National Natural Science Foundation of China (42101475)
Data publisher	Global Change Research Data Publishing & Repository, http://www.geodoi.ac.cn
Address	No. 11A, Datun Road, Chaoyang District, Beijing 100101, China
Data sharing policy	(1) <i>Data</i> are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use <i>Data</i> subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute <i>Data</i> subject to written permission from the GCdataPR Editorial Office and the issuance of a <i>Data</i> redistribution license; and (4) If <i>Data</i> are used to compile new datasets, the “ten percent principal” should be followed such that <i>Data</i> records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset ^[16]
Communication and searchable system	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC

GFDL-ESM4, IPSL-CM6A-LR, KACE-1-0-G, MIROC6, MIROC-ES2L, MPI-ESM1-2-LR, MRI-ESM2-0, NorESM2-LM, NorESM2-MM, TaiESM1, UKESM1-0-LL.

In addition, this study used SMAP (Soil Moisture Active Passive) soil moisture data^[17], ERA5-Land (Medium-Range Weather Forecast Reanalysis v5-Land) surface soil moisture data^[18], and International Soil Moisture Monitoring Network data^[19] as auxiliary data for evaluation and validation.

3.2 Technological Route

This study is aimed to obtain stable and reliable CMIP6 soil moisture fusion data products. First, all raster data were converted to TIF format, with the spatial resolution unified to 0.5°×0.5° and the temporal resolution unified to the monthly scale. Second, the SMAP and ERA5-Land soil moisture data were used to perform a quality evaluation of the CMIP6 soil moisture data using the ETC method (Enhanced Triple Collocation) to obtain the Random Error Standard Deviation (RESD) and Correlation Coefficient (CC). Third, based on the quality evaluation results, CMIP6 Earth system model data with better accuracy were screened out, and the spatial distribution of the accuracy of each data was combined to calculate the summation and normalization weights, and then fusion was carried out on the basis of the summation and normalization weights. Fourth, the accuracy level of the fused data was verified and evaluated by the ground-based data. Bias, Root Mean Square Error (RMSE) and goodness of fit (*R*) are selected as the error parameters to systematically verify

the accuracy of the fused data. At the same time, the evaluation and verification results of the simple weighted average fused data are calculated as a reference. The specific data development process is shown in Figure 1.

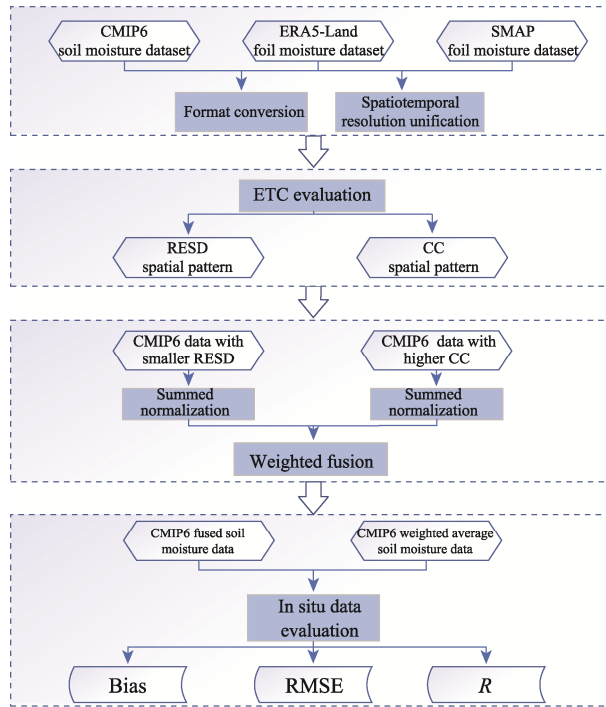


Figure 1 Flowchart of the dataset development

As shown in Table 2, the mean value of soil moisture data evaluation of each Earth system model was first calculated. ACCESS-CM2, IPSL-CM6A-LR, MIROC-ES2L, MPI-ESM1-2-LR, TaiESM1 with smaller RESD, and CanESM5-CanOE, CMCC-CM2-SR5, CNRM-CM6-1-HR, and KACE-1-0-G with higher CC were screened out. Finally, 9 Earth system model soil moisture datasets were involved in the fusion.

Table 2 Evaluated mean values of soil moisture data for each Earth system model

ESM	RESD (m^3/m^3)	CC	ESM	RESD (m^3/m^3)	CC
ACCESS-CM2	0.026	0.465	GFDL-ESM4	0.032	0.479
BCC-CSM2-MR	0.030	0.457	IPSL-CM6A-LR	0.026	0.486
CAMS-CSM1-0	0.033	0.460	KACE-1-0-G	0.069	0.515
CanESM5-CanOE	0.048	0.501	MIROC-ES2L	0.029	0.473
CESM2	0.048	0.495	MIROC6	0.030	0.486
CMCC-CM2-SR5	0.044	0.523	MPI-ESM1-2-LR	0.026	0.458
CMCC-ESM2	0.044	0.524	MRI-ESM2-0	0.036	0.479
CNRM-CM6-1	0.036	0.499	NorESM2-LM	0.047	0.491
CNRM-CM6-1-HR	0.036	0.501	NorESM2-MM	0.047	0.498
CNRM-ESM2-1	0.036	0.503	TaiESM1	0.030	0.499
EC-Earth3-Veg-LR	0.046	0.470	UKESM1-0-LL	0.039	0.494

4 Data Results and Validation

4.1 Dataset Composition

The main contents of the datasets include: (1) Global soil moisture data, with a temporal range from January 2015 to December 2100 and a spatial range covering the globe, with a temporal resolution on a monthly scale and a spatial resolution of $0.5^\circ \times 0.5^\circ$. The data unit is m^3/m^3 and the data value range is 0 to 1. The file naming scheme is *SSP***_yyyy-mm.tif*. (2) *In situ* measurements of NAQU, REMEDHUS, SMOSMANIA, and TWENTE networks. The above-mentioned datasets are archived in .tif, .shp, and .csv, which have 3,124 files in total.

4.2 Data Products

Figure 2 shows the soil moisture data of the fused CMIP6 under the 3 shared socio-economic path modes of SSP1-2.6, SSP2-4.5, and SSP5-8.5. The January, April, July, and October of 2030 are used as the samples of the fused soil moisture product of CMIP6 to present 4 seasons of winter, spring, summer, and autumn. As can be seen from the figure, the spatial and temporal distribution pattern of CMIP6 fused soil moisture is characterized by a more consistent climate seasonal rhythm cycle. The data were filtered using water mask to exclude rivers, lakes, glaciers and other terrestrial water bodies to guarantee the reasonableness.

4.3 Data Validation

In order to quantitatively evaluate the accuracy of CMIP6 soil moisture fusion data, this study selected and collected data from 4 ground-based soil moisture monitoring networks with long-term monitoring capabilities, including NAQU (located in the Nagqu region of the Qinghai-Xizang Plateau), REMEDHUS (located in Spain), SMOSMANIA (located in France) and TEWNTTE (located in the Netherlands), to evaluate the accuracy of the fusion data during 2015–2024. Given that ground station measured data mostly provide hourly surface soil moisture data, the hourly data were first synthesized into daily scale data and then into monthly scale data. In order to guarantee the data quality, stability and representativeness, it was set that monitoring data of not less than 12 hours in a day could be weighted to calculate daily-scale data, and monitoring data of not less than 15 days in a month could be weighted to calculate monthly-scale data. At the same time, the weighted average of 22 CMIP6 soil moisture datasets were calculated as a reference.

As shown in Table 3, the CMIP6 soil moisture fusion data has significant advantages in numerical accuracy (Bias and RMSE) compared with the weighted average data, revealing the accuracy improvement effect of normalized fusion based on the spatial distribution data of ETC accuracy. The CMIP6 soil moisture fusion data can be comparable to the weighted average data in terms of the accuracy of the fusion data in terms of *R*. It could be able to reasonably depict the textural characteristics of the spatial and temporal distribution of soil moisture on the ground, and reasonably fit the measured soil moisture values on the ground.

5 Discussion and Conclusion

Soil moisture is an important component of the surface water cycle system, and the

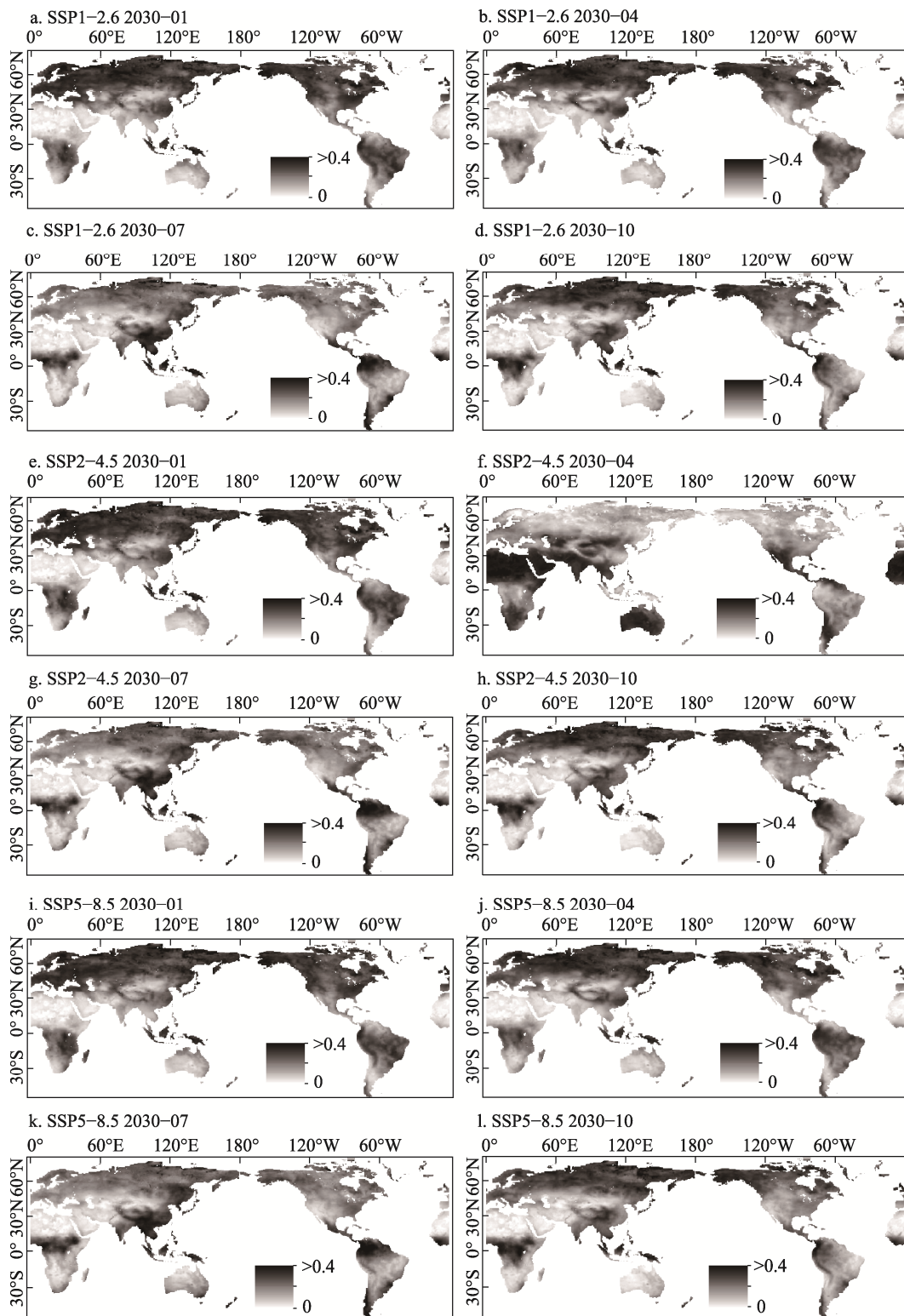


Figure 2 Maps of soil moisture integration data for SSP1-2.6, SSP2-4.5, and SSP5-8.5 in January, April, July, and October 2030

Table 3 CMIP6 soil moisture fusion data accuracy evaluation results

Soil moisture monitoring network	Evaluation metrics	Fusion data			Weighted average data		
		SSP1-2.6	SSP2-4.5	SSP5-8.5	SSP1-2.6	SSP2-4.5	SSP5-8.5
NAQU	Bias (m ³ /m ³)	0.043	0.042	0.040	0.081	0.087	0.083
	RMSE (m ³ /m ³)	0.055	0.055	0.054	0.091	0.096	0.092
	<i>R</i>	0.779	0.766	0.757	0.685	0.703	0.737
REMEDUHS	Bias (m ³ /m ³)	0.107	0.107	0.104	0.140	0.138	0.136
	RMSE (m ³ /m ³)	0.117	0.119	0.116	0.147	0.146	0.145
	<i>R</i>	0.747	0.708	0.710	0.731	0.714	0.712
SMOSMANIA	Bias (m ³ /m ³)	0.032	0.029	0.027	0.054	0.051	0.050
	RMSE (m ³ /m ³)	0.084	0.084	0.083	0.092	0.092	0.091
	<i>R</i>	0.681	0.682	0.673	0.696	0.679	0.688
TWENTE	Bias (m ³ /m ³)	0.168	0.163	0.169	0.214	0.211	0.214
	RMSE (m ³ /m ³)	0.174	0.170	0.176	0.218	0.215	0.218
	<i>R</i>	0.587	0.593	0.545	0.570	0.602	0.561

acquisition of reliable global soil moisture data of long time series is a key support for understanding the water cycle pattern. In the context of global climate change, obtaining reliable soil moisture data for future scenarios is essential to accurately study and explore the evolutionary characteristics of the spatial and temporal distribution of water resources. In this study, from the perspective of data fusion, multiple sets of CMIP6 future soil moisture data were normalized and weighted with the corresponding precision spatial distribution data on the basis of ETC evaluation, and got a global monthly-scale 0.5°×0.5° resolution surface soil moisture dataset from 2015 to 2100. As verified by the measured data at the site, the fusion data has significantly lower Bias and RMSE than the simple weighted average data, indicating that the fusion data can effectively improve the accuracy and reliability.

This dataset, as a global surface future soil moisture dataset, can be used as a reference basis for the study of climate change, ecological risk and in-depth understanding of the spatial and temporal evolution of soil moisture. Meanwhile, it provides a scientific data basis for revealing the simulation of the migration of soil moisture on the surface, and auxiliary decision-making support for understanding and coping with the evolution of the imbalanced distribution of surface water resources, and for sustainable development of surface water resources.

Author Contributions

Yang, F. did the overall design for the development of the dataset, collected and processed all the data; Liu, Y. X. Y. wrote the data paper.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Zhou, C. H., Yu, J. J. Review and prospect of hydrography research in China [J]. *Journal of Geography*, 2023, 78(7): 1659–1665.
- [2] Liu, Y. X. Y., Yang, Y. P. A review of the progress in fusing large-scale regional soil moisture datasets with multi-source microwave remote sensing [J]. *Frontiers in Data and Computing Development*, 2023, 4(6): 24–33.

- [3] Pan, N., Wang, S., Liu, Y. X., *et al.* Advances in soil moisture remote sensing inversion research [J]. *Journal of Ecology*, 2019, 39(13): 4615–4626.
- [4] Xie, Z. H., Chen, S., Qin, P. H., *et al.* Research on climate feedbacks of human water use activities and their impacts on the terrestrial water cycle: progress and challenges [J]. *Advances in Earth Sciences*, 2019, 34(8): 801–813.
- [5] Liu, C. M., Liu, X. M. Exploring water cycle research from the perspective of Earth system circle mutual feedbacks and geographic synthesis [J]. *Journal of Geography*, 2023, 78(7): 1593–1598.
- [6] Li, Z., Guo, H. D., Shi, J. C. Integrating active and passive microwave data for monitoring soil moisture changes [J]. *Journal of Remote Sensing*, 2002, 6(6): 481–484.
- [7] Ma, Z. G., Fu, Z. B., Xie, L., *et al.* Some issues in the study of soil moisture and climate change relationships [J]. *Advances in Earth Sciences*, 2001, 16(4): 563–568.
- [8] Liu, Y. X. Y., Yang, Y. P., Song, J. Variations in global soil moisture during the past decades: climate or human causes? [J]. *Water Resources Research*, 2023, 59(7): e2023WR034915.
- [9] Zhou, T. J., Zou, L. W., Chen, X. L. Review of the Sixth International Coupled Model Intercomparison Program (CMIP6) [J]. *Advances in Climate Change Research*, 2019, 15(5): 445–456.
- [10] Wang, Y. N., Qiao, L., Zuo, Z. Y. Review of the CMIP6 Land Surface, Snow, and Soil Moisture Model Comparison Program (LS3MIP) [J]. *Advances in Climate Change Research*, 2022, 18(6): 795–800.
- [11] Jiang, D. B., Wang, N. Interpretation of the IPCC AR6 report: changes in the water cycle [J]. *Progress in Climate Change Research*, 2021, 17(6): 699–704.
- [12] Liu, Y. X. Y., Chen, X. N., Bai, Y. Q., *et al.* Evaluation of 22 CMIP6 model-derived global soil moisture products of different shared socioeconomic pathways [J]. *Journal of Hydrology*, 2024, 636: 131241.
- [13] Li, Z. L., Leng, P., Zhou, C. H., *et al.* Soil moisture retrieval from remote sensing measurements: current knowledge and directions for the future [J]. *Earth-Science Reviews*, 2021, 218(1): 1–24.
- [14] Qiao, L., Zuo, Z. Y., Xiao, D. Evaluation of soil moisture in CMIP6 simulations [J]. *Journal of Climate*, 2022, 35(2): 779–800.
- [15] Yang, F., Liu, Y. X. Y. Forecasting global surface soil moisture dataset using multi-scenario integration methodology (2015–2100) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2024. <https://doi.org/10.3974/geodb.2024.11.10.V1>.
- [16] GCdataPR Editorial Office. GCdataPR data sharing policy [OL]. <https://doi.org/10.3974/dp.policy.2014.05> (Updated 2017).
- [17] Spencer, M., Wheeler, K., White, C., *et al.* The Soil Moisture Active Passive (SMAP) mission L-Band radar/radiometer instrument [J]. *IEEE*, 2010: 3240–3243.
- [18] Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., *et al.* ERA5-Land: a state-of-the-art global reanalysis dataset for land applications [J]. *Earth System Science Data*, 2021, 13(9): 4349–4383.
- [19] Dorigo, W. A., Wagner, W., Hohensinn, R., *et al.* The International Soil Moisture Network: a data hosting facility for global *in situ* soil moisture measurements [J]. *Hydrology and Earth System Sciences*, 2011, 15(5): 1675–1698.