A DETAILED LIST OF THE RESPONSES
TO REVIEWER #3

Anonymous Reviewer #3 comments

General comments:

This study developed a daily precipitation dataset (CHM_PRE_V2) for China spanning 1960–2023, demonstrating significantly improved accuracy compared to previous version. This dataset holds substantial importance for advancing hydrological and climatic research in China, particularly in regions with sparse ground observations. The authors have invested considerable effort, and the work is suitable for publication in ESSD. Below are the review comments to further enhance the manuscript.

Response: We greatly appreciate your careful reading of the manuscript, positive comments, and valuable suggestions. Your thoughtful review has enhanced our paper considerably. The manuscript has been revised thoroughly according to your comments, with our point-by-point responses detailed below.

Specific comments:

1. Caution in using the term "spatiotemporal and physical correlations." Physical correlations typically encompass: (1) Static factors (e.g., latitude, longitude, elevation, slope, vegetation); (2) Meteorological dynamic factors (e.g., humidity, wind speed, available precipitation amount); (3) Land-atmosphere interactions (e.g., soil moisture, vegetation indices, sea surface temperature anomalies); (4) Cloud and precipitation microphysics (e.g., cloud-top temperature, precipitation particle scattering).

As shown in Figure 5c （Figure 1）, two-thirds of the selected "physical factors" are precipitation-related variables from different temporal scales or data sources, which conflicts with conventional understanding. The authors should either revise these selections or provide explicit justification.

Additionally, the contributions of factors listed in Figure 5c—whether they correspond to monthly or daily precipitation—require clarification. The relationships between physical factors and precipitation are strongly time-scale-dependent, especially for discontinuous daily precipitation. The authors should consider tailoring factor selection to specific temporal scales (e.g., climatological vs. daily precipitation) and regional variations.

Response: We sincerely appreciate your insightful comments. As you pointed out, the explanation of "spatiotemporal and physical correlations" in the previous manuscript was unclear. Upon careful consideration, we concluded that introducing these definitions is not necessary. Therefore, in the latest manuscript, we have updated "spatial correlation" to "spatial autocorrelation" to more precisely express the dependence of precipitation at a location on surrounding areas (Chen et al., 2010, 2016; Fan et al., 2021; Huff and Shipp, 1969). Meanwhile, "temporal and physical correlations" have been revised to "precipitation-related covariates." We have made corresponding revisions throughout the manuscript wherever correlations were mentioned to ensure the rigor of the manuscript. The major revisions are summarized as follows:

"An upgraded high-precision gridded precipitation dataset for the Chinese mainland considering spatial autocorrelation and covariates" (Title)

"Precipitation is a critical driver of the water cycle, profoundly influencing water resources, agricultural productivity, and natural disasters. However, existing gridded precipitation datasets exhibit markable deficiencies in capturing the spatial autocorrelation and associated environmental and climatic influences—here referred to collectively as precipitation-related covariates—which limits their accuracy, particularly in regions with sparse meteorological stations. To address these challenges, this study proposes a completely new gridded precipitation generation scheme that integrates long-term daily observations from 3,746 gauges with 11 key precipitation-related covariates." (Lines 12–17)

"In summary, a key limitation of existing datasets is that they tend to focus on either spatial autocorrelation or a limited set of precipitation-related covariates, but rarely incorporate multiple types of information simultaneously. However, precipitation is influenced not only by spatial autocorrelation—that is, the dependence of precipitation at a given location on surrounding areas (Chen et al., 2010, 2016; Fan et al., 2021; Huff and Shipp, 1969; Tang et al., 2020)—but also by a wide array of covariates, such as elevation, land surface conditions, atmospheric parameters, and recent precipitation events (Adler et al., 2008; Ham et al., 2023; Ravuri et al., 2021; Trucco et al., 2023). This lack of comprehensive consideration for multiple covariates constrains the accuracy of these datasets, particularly in regions with sparse meteorological stations, such as western China (Jiang et al., 2023). Moreover, existing methods tend to generate excessive minor precipitation, leading to an overestimation of precipitation events, which will have considerable impacts on hydrologic modelling (Dong et al., 2020; Kang et al., 2024; Wei et al., 2022).
To address the aforementioned issues, this study introduces a new high-precision, long-term daily gridded precipitation dataset for the Chinese mainland (a member of the China Hydro-Meteorology datasets, hereinafter called CHM_PRE V2). Building on CHM_PRE V1, CHM_PRE V2 integrates precipitation gauges, remote sensing observations, reanalysis data, and various precipitation-related factors. Through the use of advanced spatial interpolation and machine learning algorithms, our method captures spatial autocorrelation while jointly modelling multiple covariates to enhance precipitation accuracy." (Lines 58–73)

"**2 Data**
The CHM_PRE V2 dataset was developed using extensive precipitation gauge observations, supplemented with a diverse array of ancillary datasets that serve as precipitation covariates. These covariates include satellite-derived products, land surface model outputs, and various geophysical and meteorological variables, aiming to enhance the characterization of precipitation, particularly in regions with sparse observational coverage. This integration of multi-source information is designed to improve the spatial continuity and accuracy of the precipitation estimates across the Chinese mainland. **Figure 1** illustrates details of the various datasets utilized in CHM_PRE V2 construction, including dataset names, original spatial and temporal resolutions, and coverage periods. In total, 16 datasets from 11 distinct categories were incorporated. These datasets collectively provide critical information on land surface

properties, atmospheric conditions, and recent precipitation patterns that influence precipitation generation and distribution. In addition, the CHM_PRE V2 dataset is designed to represent precipitation characteristics across the Chinese mainland, excluding Taiwan, Hong Kong, Macau, and other Chinese islands. In the following sections, we will provide a detailed introduction to the data sources employed in the construction of the CHM_PRE V2 dataset.

2.1 Spatial autocorrelation data

CHM_PRE V2 incorporates comprehensive daily precipitation gauge data to support spatial autocorrelation modelling. The primary daily precipitation gauge data sourced from the China Meteorological Administration (CMA; http://data.cma.cn, last access: January 2024) spans the entire Chinese mainland, encompassing records from 2,816 stations between 1960 and 2023. Daily precipitation is defined as the cumulative precipitation recorded between 20:00 on one day and 20:00 on the following day (local time in Beijing), with all data subjected to rigorous quality control (Zhang et al., 2020). To mitigate the limit of boundary effects (Ahrens, 2006), additional precipitation gauges near the Chinese mainland were obtained from the Global Historical Climatology Network-Daily Version 3 (GHCND) dataset. The GHCND is a reliable and globally comprehensive climate dataset, and maintained by the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA) (Durre et al., 2008, 2010; Menne et al., 2012). The GHCND dataset was sourced from NOAA (https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily) on September 11, 2024.

To ensure data quality, only stations with more than 70% effective days (over 255 days) in a year were retained for dataset construction. **Figure 2(a)** illustrates the spatial distribution of both CMA and GHCND stations, while **Figure 2(b)** shows their annual availability. Over time, the number of available CMA stations increased from 1,992 in 1960 to 2,767 in 2023, improving spatial coverage considerably. In contrast, the number of accessible GHCND stations in the region declined from 674 in 1960 to 264 in 2023.

2.2 Precipitation-related covariate data

The Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) dataset was utilized to characterize the influence of elevation on precipitation and to generate slope data. In this study, we used the SRTM DEM V4 acquired from the Consortium for Spatial Information, Consultative Group for International Agricultural Research (CGIAR-CSI, https://srtm.csi.cgiar.org/) on August 8, 2024, with a spatial resolution of 3 arc-seconds (approximately 90 meters near the equator). The SRTM DEM V4 was generated based on National Aeronautics and Space Administration (NASA) SRTM DEM V1, and has undergone post-processing of the NASA data to “fill in” the no data voids, such as water bodies (lakes and rivers), areas with snow cover and in mountainous regions (e.g., the Himalayas), resulting in seamless elevation for the globe.

To enhance the spatial and temporal detail of precipitation estimation, two satellite-based precipitation products—the Global Satellite Mapping of Precipitation (GSMaP) and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN-CDR) dataset—were incorporated as covariates. GSMaP V8 data spans from 1998 to the present with 0.1° spatial and 1-hour temporal resolution (Kubota et al., 2020). We acquired the GSMaP data from Japan Aerospace Exploration Agency (JAXA; https://sharaku.eorc.jaxa.jp) on September 9, 2024, and used the data from 1998 to 2023.

PERSIANN-CDR data spans from 1983 to the present (Ashouri et al., 2015), and the data from 1983 to 1997 was used for the retrieval.
The precipitation and soil moisture from the Global Land Data Assimilation System Noah Land Surface Model (GLDAS NOAH) (Rodell et al., 2004) were also used for the retrieval. The data spans from 1960 to 1999 and the data spans from 2000 to 2023 were acquired from the GLDAS Noah L4 V2.0 and GLDAS Noah L4 V2.1 datasets. The NOAA Climate Data Record (CDR) of AVHRR Normalized Difference Vegetation Index (NDVI) (Vermote and NOAA CDR Program, 2019) was utilized to depict the vegetation characteristics, and the data from 1981 to 2023 was used.
In addition to spatial and environmental variables, precipitation temporal features were also introduced as covariates. Two types of temporal indicators were constructed: (1) the cumulative precipitation of the current month and year, representing broader-scale precipitation conditions; and (2) daily lagged precipitation values from the previous five days, capturing short-term fluctuations. Each of these five recent days was treated as a separate variable. For example, the variable named "1st-day prior Prec." refers to precipitation one day before the current date, while "5th-day prior Prec." corresponds to five days prior." (Lines 79–135)

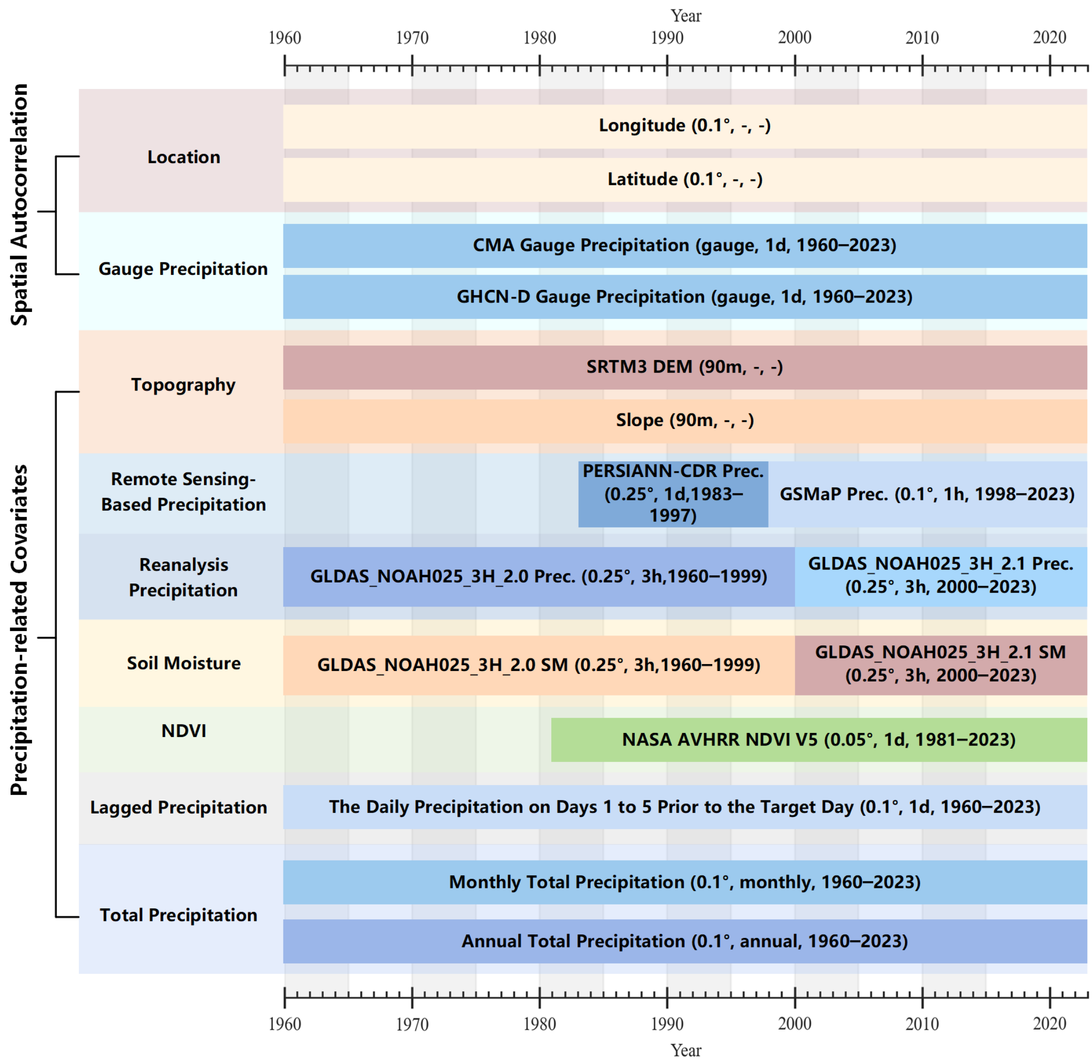

Figure 1. The data used for precipitation retrieval.

Thank you again for your thoughtful comments and support, which have helped us significantly improve the rigor of our manuscript.

2. Clarify the data sources for GLDAS 2.0 and 2.1 precipitation. The authors should specify whether GLDAS 2.0 and 2.1 precipitation are reanalysis, remote sensing, or fused products. The term "Data Assimilation precipitation" is imprecise and should be revised in the text.

Response: Thank you for pointing out this issue. In the revised manuscript, we have explicitly clarified that GLDSA precipitation refers to reanalysis precipitation, in order to better address this issue. The corresponding revisions have been made to Figure 1 and the relevant text, as follows:

"Precipitation datasets derived from gauge-based interpolation (CHM_PRE V1 and CHM_PRE V2) demonstrate significantly higher accuracy compared to those based on remote sensing (GSMaP, IMERG, and PERSIANN-CDR) and reanalysis (GLDAS), as evidenced by lower absolute error, higher KGE) and RSD (Figure 6(a-c))." (Lines 325–328)

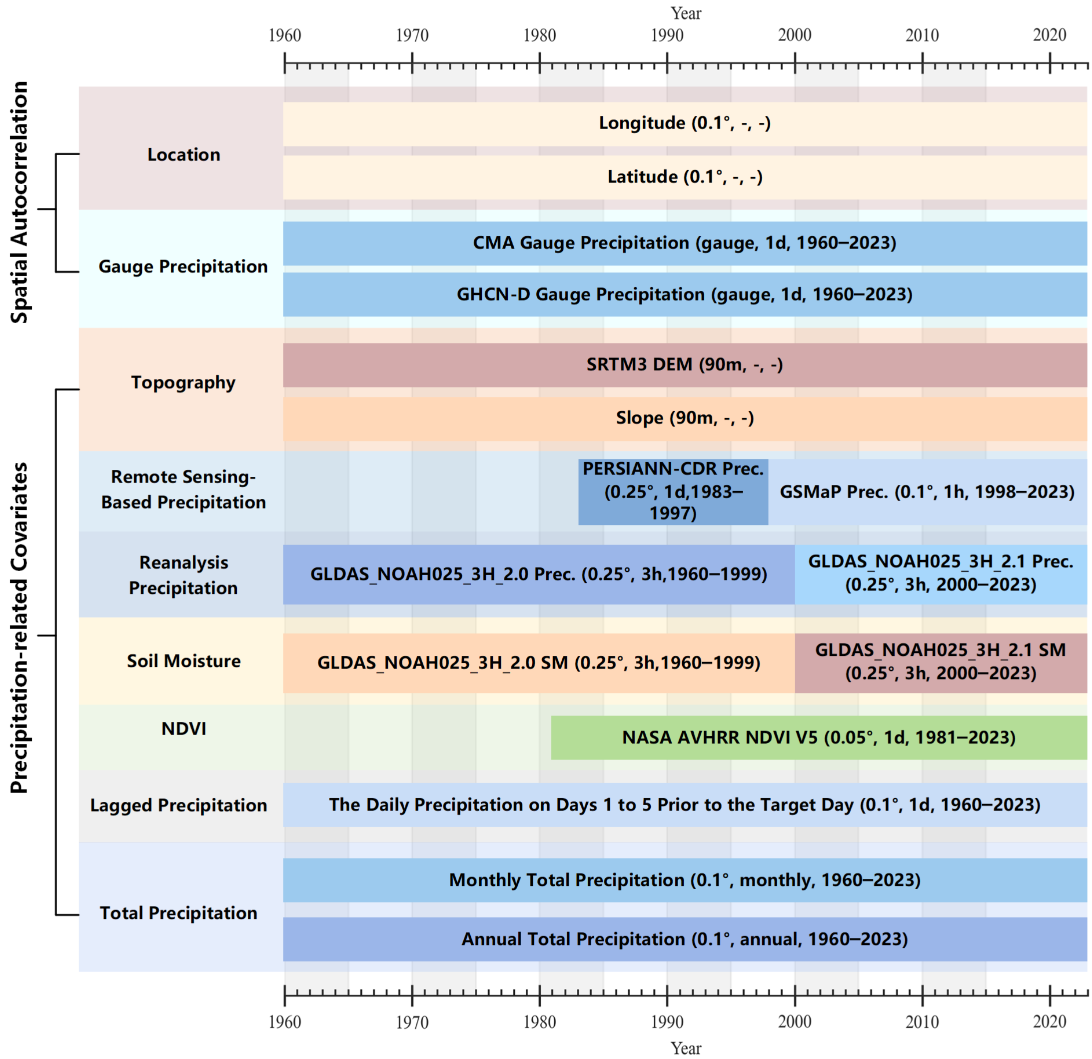

Figure 1. The data used for precipitation retrieval.

3. Highlight key improvements of updated data. A table or summary explicitly comparing critical differences between CHM_PRE_V2 and its predecessor (e.g., input data, methodology, validation metrics) is strongly recommended. This will underscore the dataset's advancements and novelty.

Response: We fully agree with your suggestion to further highlight the comparison between CHM_PRE V2 and V1. In the revised manuscript, we have added Section 4.4 titled "Improvements compared to the previous CHM_PRE V1 dataset" to emphasize the advancements and novelty of CHM_PRE V2. Similarly, we have also strengthened the comparison with CHM_PRE V1 in the Abstract, Introduction, and other relevant sections. The corresponding revisions are as follows:

"Building upon the improved inverse distance weighting interpolation method used in our previous dataset CHM_PRE V1, we integrated a machine learning algorithm—light gradient boosting machine (LGBM)—to incorporate precipitation-related covariates in a data-driven manner." (Lines 17–20)

"Our previous study developed a gridded precipitation dataset for the Chinese mainland (a member of the China Hydro-Meteorology datasets, hereinafter called CHM_PRE V1) based on inverse-distance weighting interpolation method and parameter-elevation regression on independent slopes model (PRISM) (Daly et al., 1994, 2002), using data from 2,839 gauges. The CHM_PRE V1 demonstrates overall high accuracy across the Chinese mainland (Han et al., 2023), and has received widespread attention and extensive use, benefiting a large number of hydro-meteorological related studies (Hu et al., 2024; Wan and Zhou, 2024; Yin et al., 2025). However, interpolation-based precipitation datasets rely heavily on ground meteorological gauges, performing poorly in areas with sparse station distribution or missing data." (Lines 50–56)

"**4.4 Improvements compared to the previous CHM_PRE V1 dataset**

CHM_PRE V2 is a continuation and improvement of our previously published CHM_PRE V1. Therefore, we further summarize the differences between CHM_PRE V2 and CHM_PRE V1 in **Table 1**, and highlight the improvements of CHM_PRE V2 over the previous version by using bold font. It can be observed that CHM_PRE V2 shares the same spatiotemporal resolution and coverage with V1 (except for the extended time range up to 2023), mainly to maintain consistency with other datasets in the CHM family (Zhang et al., 2025). The spatial interpolation method used in CHM_PRE V2 is largely consistent with that in V1, but it incorporates precipitation-related covariates in a data-driven manner by integrating the LGBM method. Eleven precipitation-related variables were considered, including topographic features (elevation and slope), satellite-derived precipitation estimates, reanalysis-based precipitation products, soil moisture, NDVI, recent daily precipitation records, and aggregate precipitation metrics. The inclusion of these covariates allows for a better representation of the spatiotemporal variability of precipitation (Gu et al., 2023; Ma et al., 2025), resulting in improved precipitation accuracy (with MAE and KGE reaching 1.48 mm/day and 0.79, representing improvements of approximately 12.84% and 12.86% compared to CHM_PRE V1, respectively). In addition, the capability of detecting precipitation events is a critical indicator of the accuracy of precipitation datasets (Dong et al., 2020; Kang et al., 2024). CHM_PRE V2 applies a two-stage modelling approach to distinguish and correct precipitation events, which reduces overestimation of such precipitation events and improves event detection accuracy (with FAR and HSS reaching 0.24 and 0.68, respectively, reflecting improvements of approximately 54.17% and 17.24% over CHM_PRE V1). Overall, CHM_PRE V2 demonstrates obvious improvements over CHM_PRE V1 and serves as a high-accuracy daily gridded precipitation dataset for the Chinese mainland.

Table 1. Comparison between CHM_PRE V2 and CHM_PRE V1.

Category	**Item**	**CHM_PRE V1**	**CHM_PRE V2**
Metadata	Spatial resolution	0.1°	0.1°
	Temporal resolution	Daily	Daily

	Spatial coverage	18°N–54°N, 72°E–136°E	18°N–54°N, 72°E–136°E
	Time Span	1961–2022	1960–2023
Method	Spatial autocorrelation considered	✓	✓
	Interpolation method	Improved IDW method	**Improved IDW method**
	Precipitation-related covariates	Only PRISM climatology data	**11 precipitation covariates**
	Covariate modelling approach	✗	**LGBM**
	Precipitation event considered	✗	✓
Accuracy of precipitation value	MAE (mm/day)	1.67	**1.48**
	KGE	0.70	**0.79**
	RSD	0.78	**0.88**
Accuracy of precipitation event	HSS	0.58	**0.68**
	Accuracy score	0.79	**0.85**
	FAR	0.37	**0.24**

” (Lines 389–409)

Once again, we sincerely thank you for your valuable comments, which helped us better highlight the novelty and significance of this study.

4. Clarify model configurations in Section 3.3. The manuscript states that two models were developed for “precipitation event retrieval” and “precipitation value retrieval.” Please clarify whether these models share identical predictors and weighting schemes. Detailed descriptions of input variables, parameters, and training protocols for each model are essential.
Response: Thanks for your constructive suggestion, which has helped us improve the clarity of our methodology. Following your suggestion, we have thoroughly rewritten Sections 2.2 and 3.3 to better describe our modeling approach. In addition, we have added Table S3 in the supplementary materials to more clearly present the variables used for precipitation retrieval. The corresponding revisions are as follows:

“In addition to spatial and environmental variables, precipitation temporal features were also introduced as covariates. Two types of temporal indicators were constructed: (1) the cumulative precipitation of the current month and year, representing broader-scale precipitation conditions; and (2) daily lagged precipitation values from the previous five days, capturing short-term fluctuations. Each of these five recent days was treated as a separate variable. For example, the variable named “1st-day prior Prec.” refers to precipitation one day before the current date, while “5th-day prior Prec.” corresponds to five days prior.” (Lines 131–135)

“3.3 Precipitation retrieval based on covariates
Except spatial autocorrelation, precipitation is influenced by a range of meteorological factors that vary over space and time. However, most existing gridded precipitation datasets tend to

model these aspects in isolation, often focusing solely on spatial autocorrelation or meteorological inputs, which may constrain the accuracy and generalizability of the datasets, especially in regions with sparse gauge coverage. To address this limitation, we propose a novel framework that integrates multiple precipitation covariates into a unified machine learning-based retrieval system, thereby enhancing the fidelity of precipitation estimates. To model spatial autocorrelation, we employed gridded precipitation data derived from gauge-based interpolation in Section 3.2, along with geographic coordinates (longitude and latitude). Precipitation covariates were drawn from various sources, including topographic features (elevation and slope), satellite-derived precipitation estimates, reanalysis-based precipitation products, soil moisture, and the normalized difference vegetation index (NDVI). Recent daily precipitation records and aggregate precipitation metrics were also incorporated to capture the temporal variability and underlying climatological patterns. The details of the retrieval data can be found in Figure 1.

To synthesize these spatial and covariate-based features, we employed a machine learning regression framework using the light gradient boosting machine (LGBM) algorithm. This model enables the flexible representation of complex nonlinear relationships between precipitation and its associated covariates, surpassing the limitations of conventional linear regression models. While linear regression models are the most commonly used response models, they are limited by their inability to capture nonlinear relationships and their relatively weak fitting capacity (Breiman, 2001; Chen and Guestrin, 2016; Yang et al., 2021). Machine learning-based models, in contrast, offer significant improvements in fitting performance and are more effective in representing nonlinear relationships (Guo et al., 2024; Hu et al., 2023). Among numerous machine learning-based models, LGBM, developed by Microsoft (Ke et al., 2017), is renowned for its high precision and high generalizability. Fundamentally, it employs a series of decision tree models for iterative training, progressively minimizing errors (or residuals) to ultimately generate predictions through a weighted summation. Unlike traditional gradient-boosted decision tree (GBDT) methods, LGBM utilizes a histogram-based technique for data binning, rather than processing each individual data record. This method iterates, calculates gains, and splits data accordingly (Zhang and Gong, 2020). Gradient-based one-side sampling is employed to sample the dataset, assigning greater weights to data points with larger gradients during gain computation. Under equivalent sampling rates, this method often outperforms random sampling (Candido et al., 2021). Owing to these features, LGBM demonstrates exceptional accuracy and generalization, making it widely applicable to various tasks such as classification, regression, and ranking (Bian et al., 2023; Jiang et al., 2024; Zhang et al., 2024). Hu et al. (2023) applied LGBM to the retrieval of suspended sediment concentration in the lower Yellow River and found that LGBM outperformed methods such as partial least squares regression, support vector regression, and random forest in terms of retrieval accuracy. Consequently, we employed the LGBM method to integrate all these variables for precipitation retrieval, effectively accounting for the spatiotemporal and physical correlations of precipitation.

In the precipitation retrieval process, we employed a two-stage strategy: precipitation event classification and precipitation value retrieval. Sixteen variables were used as independent variables in the retrieval process, and all of them are listed in Table S3 in the supplementary materials. For the precipitation event classification model, the variable indicating whether a precipitation event occurred was used as the dependent variable, while the precipitation value was used as the dependent variable in the precipitation value retrieval model. For the convenience of updating and maintaining data every year in the future, we constructed separate models for each year. That is, for each year, the same independent variables were used to develop two different models based on the LGBM method, with precipitation event and precipitation amount as the dependent variables, respectively. One model is used for precipitation event classification, and the other for precipitation value retrieval. From 1960 to 2023, a total of 64 years, 128 different models were generated. Specifically, for a given year, all variables required for retrieval were consolidated and split into training and validation sets at a ratio of 8:2. The training set was utilized to develop a precipitation event classification model based on the LGBM method, while the validation set was used for hyperparameter optimization. Then, the established classification model was applied to all samples to determine whether each sample was a precipitation event. Samples identified as precipitation events were used to train a precipitation value reversal model based on the LGBM method, while non-precipitation samples were excluded from the retrieval process. This approach effectively removed the majority of non-precipitation samples, simplifying the capture of precipitation characteristics and enhancing the accuracy of the reversal model. Additionally, this strategy notably improved the discrimination of precipitation events and mitigated the overestimation of precipitation events commonly associated with traditional interpolation-based methods. Upon completing the retrieval process, the trained precipitation value retrieval models were used to generate the final gridded daily precipitation for the entire Chinese mainland from 1960 to 2023." (Lines 213–263)

Table S3. The variables used in the precipitation retrieval.

Variable Type	**Variable Name**	**Description**
Spatial autocorrelation variables	Lat	Latitude of the grid center
	Lon	Longitude of the grid center
	Interp. Prec.	Gridded precipitation based on gauge interpolation
Precipitation-related covariates	DEM	Average elevation of the grid
	Slope	Average slope of the grid
	GLDAS Prec.	Precipitation of the grid from GLDAS
	Prec. RS	Satellite-derived precipitation of the grid
	GLDAS SM	Soil moisture of the grid from GLDAS
	NDVI	NDVI of the grid
	Annual Prec.	Annual total precipitation of the grid
	Monthly Prec.	Monthly total precipitation of the grid
	1st-day prior Prec.	Daily precipitation one day before the current date
	2nd-day prior Prec.	Daily precipitation two day before the current date
	3rd-day prior Prec.	Daily precipitation three day before the current date
	4th-day prior Prec.	Daily precipitation four day before the current date
	5th-day prior Prec.	Daily precipitation five day before the current date

Once again, we sincerely thank you for your insightful comments, which have greatly enhanced the quality of our manuscript.

References:

Adler, R. F., Gu, G., Wang, J.-J., Huffman, G. J., Curtis, S., and Bolvin, D.: Relationships between global precipitation and surface temperature on interannual and longer timescales (1979–2006), Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2008JD010536, 2008.

Ahrens, B.: Distance in spatial interpolation of daily rain gauge data, Hydrology and Earth System Sciences, 10, 197–208, https://doi.org/10.5194/hess-10-197-2006, 2006.

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R., and Prat, O. P.: PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies, Bulletin of the American Meteorological Society, 96, 69–83, https://doi.org/10.1175/BAMS-D-13-00068.1, 2015.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Candido, C., Blanco, A. C., Medina, J., Gubatanga, E., Santos, A., Ana, R. S., and Reyes, R. B.: Improving the consistency of multi-temporal land cover mapping of Laguna lake watershed using light gradient boosting machine (LightGBM) approach, change detection analysis, and Markov chain, Remote Sensing Applications: Society and Environment, 23, 100565, https://doi.org/10.1016/j.rsase.2021.100565, 2021.

Chen, D., Ou, T., Gong, L., Xu, C.-Y., Li, W., Ho, C.-H., and Qian, W.: Spatial interpolation of daily precipitation in China: 1951–2005, Advances in Atmospheric Sciences, 27, 1221–1232, https://doi.org/10.1007/s00376-010-9151-y, 2010.

Chen, D., Tian, Y., Yao, T., and Ou, T.: Satellite measurements reveal strong anisotropy in spatial coherence of climate variations over the Tibet Plateau, Scientific Reports, 6, 30304, https://doi.org/10.1038/srep30304, 2016.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Kdd'16: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, New York, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Daly, C., Neilson, R. P., and Phillips, D. L.: A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain, Journal of Applied Meteorology and Climatology, 33, 140–158, https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2, 1994.

Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L., and Pasteris, P.: A knowledge-based approach to the statistical mapping of climate, Climate research, 22, 99–113, https://doi.org/10.3354/cr022099, 2002.

Dong, J., Crow, W. T., and Reichle, R.: Improving Rain/No-Rain Detection Skill by Merging Precipitation Estimates from Different Sources, Journal of Hydrometeorology, 21, 2419–2429, https://doi.org/10.1175/JHM-D-20-0097.1, 2020.

Durre, I., Menne, M. J., and Vose, R. S.: Strategies for Evaluating Quality Assurance Procedures, Journal of Applied Meteorology and Climatology, 47, 1785–1791, https://doi.org/10.1175/2007JAMC1706.1, 2008.

Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., and Vose, R. S.: Comprehensive Automated Quality Assurance of Daily Surface Observations, Journal of Applied

Meteorology and Climatology, 49, 1615–1633, https://doi.org/10.1175/2010JAMC2375.1, 2010.

Gu, J., Ye, Y., Jiang, Y., Dong, J., Cao, Y., Huang, J., and Guan, H.: A downscaling-calibrating framework for generating gridded daily precipitation estimates with a high spatial resolution, Journal of Hydrology, 626, 130371, https://doi.org/10.1016/j.jhydrol.2023.130371, 2023.

Guo, F., Ren, Y., Zhou, Y., Sun, S., Cui, M., and Khim, J.: Machine learning vs. statistical model for prediction modeling and experimental validation: Application in groundwater permeable reactive barrier width design, Journal of Hazardous Materials, 469, 133825, https://doi.org/10.1016/j.jhazmat.2024.133825, 2024.

Ham, Y.-G., Kim, J.-H., Min, S.-K., Kim, D., Li, T., Timmermann, A., and Stuecker, M. F.: Anthropogenic fingerprints in daily precipitation revealed by deep learning, Nature, 622, 301–307, https://doi.org/10.1038/s41586-023-06474-x, 2023.

Han, J., Miao, C., Gou, J., Zheng, H., Zhang, Q., and Guo, X.: A new daily gridded precipitation dataset for the Chinese mainland based on gauge observations, Earth System Science Data, 15, 3147–3161, https://doi.org/10.5194/essd-15-3147-2023, 2023.

Hu, J., Miao, C., Zhang, X., and Kong, D.: Retrieval of suspended sediment concentrations using remote sensing and machine learning methods: A case study of the lower Yellow River, Journal of Hydrology, 627, 130369, https://doi.org/10.1016/j.jhydrol.2023.130369, 2023.

Hu, Y., Wei, F., Fu, B., Wang, S., Xiao, X., Qin, Y., Yin, S., Wang, Z., and Wan, L.: Divergent patterns of rainfall regimes in dry and humid areas of China, Journal of Hydrology, 636, 131243, https://doi.org/10.1016/j.jhydrol.2024.131243, 2024.

Huff, F. A. and Shipp, W. L.: Spatial Correlations of Storm, Monthly and Seasonal Precipitation, Journal of Applied Meteorology and Climatology, 8, 542–550, 1969.

Jiang, Y., Yang, K., Qi, Y., Zhou, X., He, J., Lu, H., Li, X., Chen, Y., Li, X., Zhou, B., Mamtimin, A., Shao, C., Ma, X., Tian, J., and Zhou, J.: TPHiPr: a long-term (1979－2020) high-accuracy precipitation dataset (1／30°, daily) for the Third Pole region based on high-resolution atmospheric modeling and dense observations, Earth System Science Data, 15, 621–638, https://doi.org/10.5194/essd-15-621-2023, 2023.

Kang, X., Dong, J., Crow, W. T., Wei, L., and Zhang, H.: The Conditional Bias of Extreme Precipitation in Multi-Source Merged Data Sets, Geophysical Research Letters, 51, e2024GL111378, https://doi.org/10.1029/2024GL111378, 2024.

Kubota, T., Aonashi, K., Ushio, T., Shige, S., Takayabu, Y. N., Kachi, M., Arai, Y., Tashima, T., Masaki, T., Kawamoto, N., Mega, T., Yamamoto, M. K., Hamada, A., Yamaji, M., Liu, G., and Oki, R.: Global Satellite Mapping of Precipitation (GSMaP) Products in the GPM Era, in: Satellite Precipitation Measurement, vol. 67, edited by: Levizzani, V., Kidd, C., Kirschbaum, D. B., Kummerow, C. D., Nakamura, K., and Turk, F. J., Springer, Cham, 355–373, https://doi.org/10.1007/978-3-030-24568-9_20, 2020.

Ma, Z., Xu, J., Dong, B., Hu, X., Hu, H., Yan, S., Zhu, S., He, K., Shi, Z., Chen, Y., Fang, X., Zhang, Q., Gu, S., and Weng, F.: GMCP: A Fully Global Multisource Merging-and-Calibration Precipitation Dataset (1-Hourly, 0.1°, Global, 2000–the Present), Bulletin of the American Meteorological Society, 106, E596–E624, https://doi.org/10.1175/BAMS-D-24-0051.1, 2025.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G.: An Overview of the Global Historical Climatology Network-Daily Database, Journal of Atmospheric and Oceanic Technology, 29, 897–910, https://doi.org/10.1175/JTECH-D-11-00103.1, 2012.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N., Clancy, E., Arribas, A., and Mohamed, S.: Skilful precipitation nowcasting using deep generative models of radar, Nature, 597, 672–677, https://doi.org/10.1038/s41586-021-03854-z, 2021.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Tang, G., Clark, M. P., Newman, A. J., Wood, A. W., Papalexiou, S. M., Vionnet, V., and Whitfield, P. H.: SCDNA: a serially complete precipitation and temperature dataset for North America from 1979 to 2018, Earth System Science Data, 12, 2381–2409, https://doi.org/10.5194/essd-12-2381-2020, 2020.

Trucco, A., Barla, A., Bozzano, R., Pensieri, S., Verri, A., and Solarna, D.: Introducing Temporal Correlation in Rainfall and Wind Prediction From Underwater Noise, IEEE Journal of Oceanic Engineering, 48, 349–364, https://doi.org/10.1109/JOE.2022.3223406, 2023.

Vermote, E. and NOAA CDR Program: NOAA Climate Data Record (CDR) of AVHRR Normalized Difference Vegetation Index (NDVI) (5), https://doi.org/10.7289/V5ZG6QH9, 2019.

Wan, W. and Zhou, Y.: Spatiotemporal patterns in persistent precipitation extremes of the Chinese mainland (1961–2022) and association with the dynamic factors, Atmospheric Research, 310, 107600, https://doi.org/10.1016/j.atmosres.2024.107600, 2024.

Wei, G., Lü, H., Crow, W. T., Zhu, Y., Su, J., and Ren, L.: Comprehensive Evaluation and Error-Component Analysis of Four Satellite-Based Precipitation Estimates against Gauged Rainfall over Mainland China, Advances in Meteorology, 2022, 9070970, https://doi.org/10.1155/2022/9070970, 2022.

Yang, Y., Huang, T. T., Shi, Y. Z., Wendroth, O., and Liu, B. Y.: Comparing the Performance of an Autoregressive State-Space Approach to the Linear Regression and Artificial Neural Network for Streamflow Estimation, Journal of Environmental Informatics, 37, 36–48, https://doi.org/10.3808/jei.202000440, 2021.

Yin, C., Bai, C., Zhu, Y., Shao, M., Han, X., and Qiao, J.: Future Soil Erosion Risk in China: Differences in Erosion Driven by General and Extreme Precipitation Under Climate Change, Earth's Future, 13, e2024EF005390, https://doi.org/10.1029/2024EF005390, 2025.

Zhang, D. and Gong, Y.: The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure, IEEE Access, 8, 220990–221003, https://doi.org/10.1109/ACCESS.2020.3042848, 2020.

Zhang, Y., Ren, Y., Ren, G., and Wang, G.: Precipitation Trends Over Mainland China From 1961–2016 After Removal of Measurement Biases, Journal of Geophysical Research: Atmospheres, 125, e2019JD031728, https://doi.org/10.1029/2019JD031728, 2020.

Zhang, Y., Feng, X., Zhou, C., Sun, C., Leng, X., and Fu, B.: Aridity threshold of ecological restoration mitigated atmospheric drought via land–atmosphere coupling in drylands, Commun Earth Environ, 5, 1–11, https://doi.org/10.1038/s43247-024-01555-9, 2024.

Zheng, H., Miao, C., Huntingford, C., Tarolli, P., Li, D., Panagos, P., Yue, Y., Borrelli, P., and Van Oost, K.: The Impacts of Erosion on the Carbon Cycle, Reviews of Geophysics, 63, e2023RG000829, https://doi.org/10.1029/2023RG000829, 2025.

-- end line--

In order to make the review of our revision more convenient, we have marked all changes using the “Track Changes” function in Microsoft Word, and have uploaded the “tracked changes” version as Supplementary Material.