A DETAILED LIST OF THE RESPONSES TO REVIEWER #2

Anonymous Reviewer #2 comments

General comments:

The study presents a novel and interesting approach to developing high-resolution gridded precipitation data (CHM_PRE v2.0) by integrating station data, multiple covariate factors, and machine learning techniques. Given the significant spatial and temporal variability of precipitation, the development of reliable and credible gridded precipitation datasets is crucial for hydroclimatological research. The authors have clearly put considerable effort into this work, particularly by incorporating covariate factors beyond traditional spatial interpolation methods. This dataset is likely to have broad applicability in the field. However, several issues need to be addressed to improve the clarity, rigor, and impact of the manuscript.

Response: We greatly appreciate your careful reading of the manuscript, insightful 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. The introduction mentions that the authors previously developed CHM_PRE v1.0. It is unclear how much innovation or improvement has been achieved in v2.0 compared to v1.0. The authors should provide a detailed explanation of the differences between the two versions and justify why a new release (v2.0) is necessary instead of simply updating v1.0. This is critical for readers to understand the added value of this new version.

Response: Thank you for the valuable comment. We have revised the introduction to better highlight the differences between CHM_PRE V2 and V1, as well as the significance of CHM_PRE V2. The corresponding revisions are as follows:

"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.

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 50–72)

2. The authors fused data from 2,816 stations to produce 0.1-degree gridded precipitation data. However, the rationale for choosing 0.1-degree resolution over finer resolutions (e.g., 0.05-degree or 1 km) is not explained. Given the availability of high-resolution precipitation datasets in China, including sub-daily data, the authors should discuss why 0.1-degree resolution was selected and whether finer resolutions were considered.

Response: Thank you for pointing out the issue regarding the choice of resolution. Spatial resolution is a very important attribute for gridded datasets, and in this study, consistency with our previous datasets was the primary factor in selecting the spatial resolution. Our previous datasets, CHM_Drought (Zhang et al., 2025) and CHM_PRE V1 (Han et al., 2023), both have a spatial resolution of 0.1°. Therefore, CHM_PRE V2 was also set at this resolution to ensure compatibility with other datasets in the CHM family. In addition, we believe that a 0.1° resolution provides a good balance between accuracy and computational efficiency at large scales. We have added some explanations in the introduction to better clarify this point:

"The spatial resolution of the dataset is set to 0.1° to maintain consistency with our previous dataset (Han et al., 2023; Zhang et al., 2025)." (Lines 73–74)

3. The terms "Spatiotemporal correlated data" and "Physically correlated data" are introduced but not clearly defined. A more detailed explanation of these terms is necessary to ensure readers fully understand the methodology and its theoretical basis.

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)

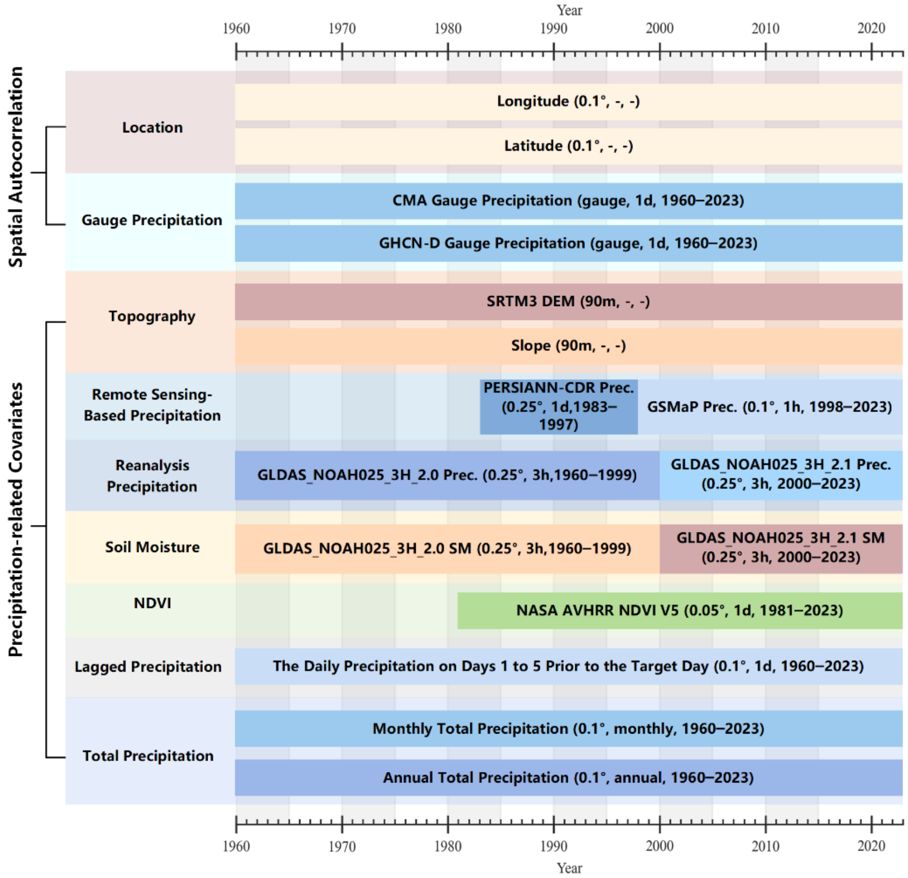

Figure 1. The data used for precipitation retrieval.

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

4. The manuscript contains numerous abbreviations, which hinder the readability and flow of the text. The authors should minimize the use of abbreviations or provide a glossary for reference.

Response: Thank you for pointing out the issue regarding the abbreviations. Following your suggestion, we have summarized all abbreviations used in the manuscript and added them to the supplementary materials (**Table S1**). The corresponding revisions are as follows:

"For clarity, a list of abbreviations used throughout this paper is presented in **Table S1** in the supplementary materials." (Lines 77–78)

Table S1. List of abbreviations used throughout this paper.

Abbreviation	**Full Term**
LGBM	Light gradient boosting machine
PRISM	Parameter-elevation regression on independent slopes model
IDW	Inverse distance weighting
CMA	China Meteorological Administration
GHCND	Global Historical Climatology Network-Daily
NCDC	National Climatic Data Center
NOAA	National Oceanic and Atmospheric Administration
SRTM	Shuttle Radar Topography Mission
DEM	Digital Elevation Model
CGIAR-CSI	Consortium for Spatial Information, Consultative Group for International Agricultural Research
NASA	National Aeronautics and Space Administration
GSMaP	Global Satellite Mapping of Precipitation
PERSIANN-CDR	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
JAXA	Japan Aerospace Exploration Agency
GLDAS NOAH	Global Land Data Assimilation System Noah Land Surface Model
NDVI	Normalized Difference Vegetation Index
IMERG	Integrated Multi-satellitE Retrievals for GPM
CMA-HD	High-density automatic rain gauge stations across Chinese mainland
NEC	North East China
NC	North China
SCC	South and Central China
IM	Inner Mongolia
NWC	North West China
SWC	South West China
QT	Qinghai-Tibet Plateau
CDD	Correlation decay distance
ADW	Adaptive distance weighting

GBDT	Gradient-boosted decision tree
AE	Absolute error
KGE	Kling-Gupta efficiency
RSD	Relative standard deviation
HSS	Heidke skill score
FAR	False alarm ratio
POD	Probability of detection

Thank you again for highlighting this issue, which has helped us make the manuscript more readable.

5. In CHM_PRE v1.0, the authors used ADW (Anisotropic Distance Weighting) interpolation, but in v2.0, they reverted to IDW (Inverse Distance Weighting). The rationale for this change is not explained. The authors should clarify why IDW was chosen for v2.0 and how it compares to ADW in terms of performance.

Response: Thank you for your valuable comment. In fact, regarding the process of interpolating gauge observations to generate gridded precipitation, the core method used in both CHM_PRE V2 and V1 is the same — an inverse distance weighting (IDW) method with a correlation decay distance (CDD). Based on previous studies (Han et al., 2023; Shen et al., 2010; Xie et al., 2007) and our extensive testing, this method is capable of generating high-accuracy gridded precipitation datasets. In this study, there are three main differences in the interpolation of gauge precipitation compared with CHM_PRE V1:

(1) CHM_PRE V2 restricts the interpolation to the nearest 10 stations when more than 10 stations are available within CDD1, in order to reduce the overestimation of precipitation events in densely gauged areas in eastern China.

(2) In CHM_PRE V1, interpolation was performed by interpolating the ratio of a station's daily precipitation to its daily climatology and then multiplying by the daily climatology. However, for some stations with very low precipitation, this approach could produce extremely large ratios, resulting in unrealistic high precipitation in arid regions. CHM_PRE V2 interpolates the anomalies relative to the climatology instead (He et al., 2020; Zhang et al., 2025) to address this issue.

(3) CHM_PRE V1 used the parameter-elevation regression on independent slopes model (PRISM) climatology data (Daly et al., 1994; Daly et al., 2002) to account for local topographic effects. In contrast, CHM_PRE V2 incorporates local topographic influences such as elevation and slope through a data-driven modeling approach (Section 3.3). Therefore, daily and monthly gridded climatologies were directly calculated and interpolated from gauge observations.

We hope to introduce the production process of the CHM_PRE V2 dataset to its users in a concise manner. After careful consideration, we have rewritten parts of the introduction to better highlight the necessity and significance of upgrading CHM_PRE V1 to V2. However, in the interpolation section based on gauge observations (Section 3.2), we have maintained the original structure without adding detailed comparisons with CHM_PRE V1, in order to avoid overburdening the readers. The revisions made to the introduction are as follows:

"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.
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 50–72)

6. The term "CMA-HD" is used but not defined. The authors should provide a clear explanation of what this term refers to.
Response: We apologize for this issue. We have added the full form of the abbreviation in the revised manuscript. Additionally, a list of abbreviations has been included in the supplementary materials (**Table S1**) to improve the readability of the manuscript. The corresponding revisions are as follows:

"To further validate the reliability of precipitation data, we obtained daily precipitation observations from 72,901 high-density automatic rain gauge stations across the Chinese mainland (hereafter we refer to it as CMA-HD), provided by the National Meteorological Information Center of CMA (Li et al., 2018)." (Lines 144–146)

Table S1. List of abbreviations used throughout this paper.

Abbreviation	**Full Term**
LGBM	Light gradient boosting machine
PRISM	Parameter-elevation regression on independent slopes model
IDW	Inverse distance weighting
CMA	China Meteorological Administration
GHCND	Global Historical Climatology Network-Daily
NCDC	National Climatic Data Center
NOAA	National Oceanic and Atmospheric Administration
SRTM	Shuttle Radar Topography Mission
DEM	Digital Elevation Model
CGIAR-CSI	Consortium for Spatial Information, Consultative Group for International Agricultural Research
NASA	National Aeronautics and Space Administration
GSMaP	Global Satellite Mapping of Precipitation
PERSIANN-CDR	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
JAXA	Japan Aerospace Exploration Agency
GLDAS NOAH	Global Land Data Assimilation System Noah Land Surface Model
NDVI	Normalized Difference Vegetation Index
IMERG	Integrated Multi-satellitE Retrievals for GPM
CMA-HD	High-density automatic rain gauge stations across Chinese mainland
NEC	North East China
NC	North China
SCC	South and Central China
IM	Inner Mongolia
NWC	North West China
SWC	South West China
QT	Qinghai-Tibet Plateau
CDD	Correlation decay distance
ADW	Adaptive distance weighting
GBDT	Gradient-boosted decision tree
AE	Absolute error
KGE	Kling-Gupta efficiency
RSD	Relative standard deviation
HSS	Heidke skill score
FAR	False alarm ratio
POD	Probability of detection

7. The authors evaluated the dataset using 63,397 station data points. However, instead of interpolating these station data to 0.1-degree grids using IDW or ADW, they averaged the station values within each 0.1-degree grid for accuracy assessment. The rationale for this approach should be explained, as interpolation might provide a more consistent comparison.

Response: Thank you for your comment regarding the accuracy evaluation. We fully agree that interpolating the validation stations can generate gridded data with spatial relationships more consistent with CHM_PRE V2. At the same time, directly comparing station observations with the corresponding grid values introduces the issue of missing accuracy assessments for grid cells without validation stations. However, considering the inevitable uncertainties introduced by interpolation—particularly in regions with sparse station coverage such as western China—we believe that comparing raw station observations with their corresponding grid values yields more reliable accuracy assessments.

Thank you again for your valuable comment. To better clarify this issue, we have added the following explanation to Section 3.4:

"There are two approaches to using station observations to validate the accuracy of gridded precipitation data. The first approach involves interpolating the station data—using methods such as IDW—to generate gridded data at the same spatial resolution as the dataset being validated. This method can produce spatially consistent results with the target gridded dataset. However, as previously mentioned, interpolation methods have some limitations and inevitably introduce interpolation-related uncertainties (McMillan et al., 2018; Wagner et al., 2012). Moreover, the uneven spatial distribution of stations makes the validation results in sparsely monitored areas less reliable. The second approach is to directly compare the station observations with the corresponding grid cell values in the dataset being validated. Although this method only provides validation results for grid cells that contain observation stations, it avoids the uncertainties introduced by interpolation and ensures the reliability of the accuracy assessment. In this study, we adopted the second approach for the validation. To align with the 0.1° gridded precipitation data, station observations were mapped onto a 0.1° grid, and the average precipitation of all stations within each grid cell was regarded as the true precipitation value for that grid cell." (Lines 268–278)

8. The units of variables in Equations 1-8 are not provided. The authors should include the units to ensure clarity and reproducibility.

Response: Thank you for bringing this issue to our attention. We have added units for all variables in the revised manuscript. The relevant sentences are as follows:

"where $d(G, P_i)$ represents the distance (km) between grid cell G and gauge station P_i, and p is the distance weighting exponent." (Lines 191–192)

"where y and $\hat{y}$ represent the observed precipitation values and the gridded precipitation values (mm/day), respectively; μ denotes the mean value, σ signifies the standard deviation" (Lines 285–286)

9. The relative importance plot in Figure 5 is not well explained. Specifically, it is unclear how the relative importance values were calculated and what "2nd-day prior Prec." and "5th-day prior Prec." represent. Are these cumulative values? A more detailed explanation is needed.

Response: Thank you for your valuable comment. The previous analysis of relative importance was based on feature importance derived from the LGBM method by the node splitting, as described at the end of Section 3.4 in the previous manuscript. In this revision, we have re-evaluated the use of feature importance and concluded that it may not be sufficiently reliable for explaining the contributions of different variables to precipitation retrieval. Therefore, we have removed the related content on relative importance (**Figure 5**(c)) in the latest manuscript to ensure the manuscript's rigor. The updated **Figure 5** is as follows:

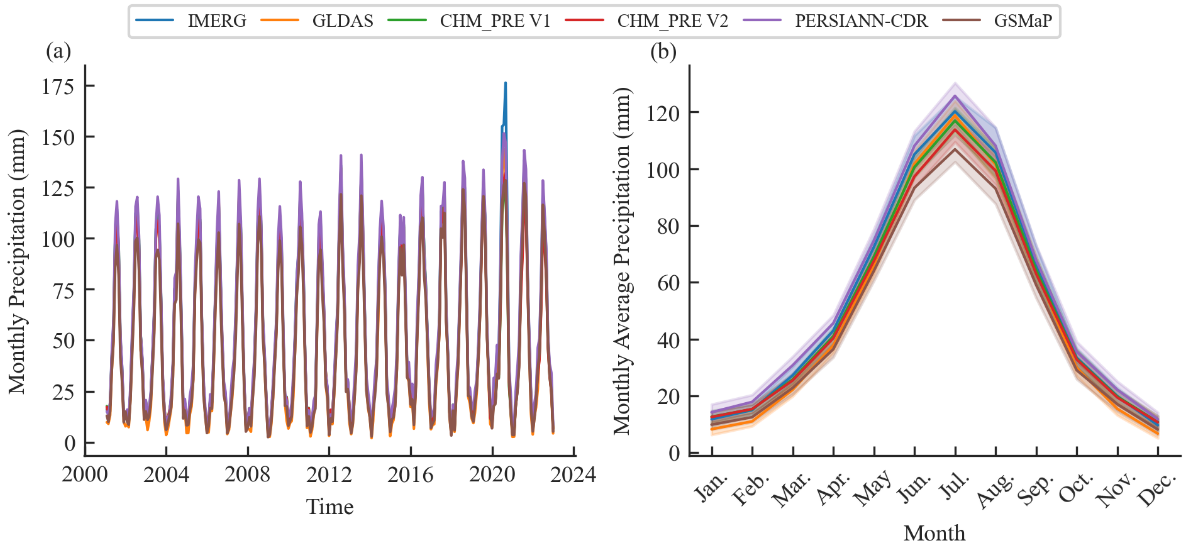

Figure 5. (a) time series of monthly precipitation; (b) multi-year mean monthly precipitation from 2001 to 2020.

We also sincerely apologize for not providing a sufficiently clear explanation of the modeling variables. The variable "2nd-day prior Prec." refers to the daily precipitation two days before the current day—it represents only the value of that specific day, not an accumulated amount over multiple days. In the revised manuscript, we have clarified the meaning of the time-related covariates used in the modeling in Section 3.2. We have also added a comprehensive list of variables used in precipitation retrieval (**Table S3** in the supplementary materials) to better illustrate the modeling details. 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)

Table S3. The variables used in the precipitation retrieval.

Variable Type	**Variable Name**	**Description**
	Lat	Latitude of the grid center

Spatial autocorrelation variables	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

10. Figure 6 shows notably high absolute errors in the NC and SCC regions. The authors should discuss the potential reasons for these high errors and whether they are related to regional characteristics or methodological limitations.

Response: As you rightly pointed out, both CHM_PRE V2 and V1 exhibit larger absolute errors in regions such as NC, SCC, and QT compared to other regions. This is mainly attributed to the higher precipitation amounts in these regions, which naturally lead to greater absolute errors. To further analyze the error characteristics of CHM_PRE V2 across different regions, we calculated the relative error for each dataset (**Figure R1**(a)) as well as the relative error difference between CHM_PRE V2 and V1 in each region (**Figure R1**(b)). The relative error here is defined as the absolute error of each precipitation event divided by the true daily precipitation value, expressed as a percentage (%).

Figure R1(a) shows that CHM_PRE V2 has lower relative errors compared to other datasets. **Figure R1**(b) indicates that CHM_PRE V2 exhibits only minor differences in relative error across different regions and performs better than CHM_PRE V1. Therefore, we conclude that CHM_PRE V2 maintains generally stable error levels across regions.

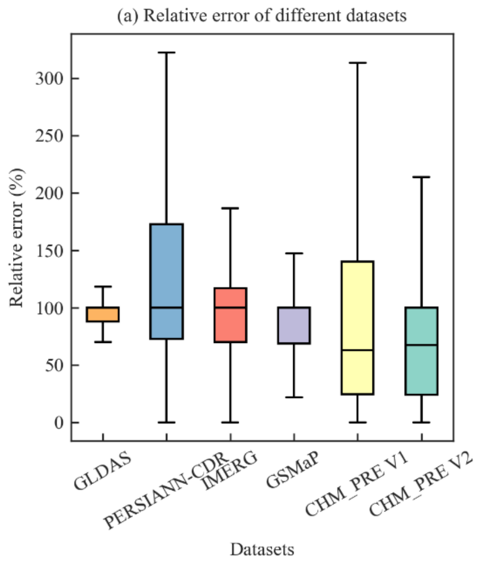

Figure R1. (a) relative error for different precipitation datasets on the testing dataset CMA-HD; (b) relative error difference between CHM_PRE V2 and V1 in each region.

In contrast, accuracy metrics that are unaffected by the magnitude of the variable—such as the Kling-Gupta Efficiency (KGE; **Figure 6**(e)) and the Relative Standard Deviation (RSD; **Figure 6**(f))—demonstrate better regional consistency. The variability of KGE and RSD is relatively higher in SWC and QT, which may be attributed to the sparse distribution of precipitation observation stations and the high spatiotemporal variability of precipitation in these areas (Li et al., 2015; Liu et al., 2019). We have added some discussion to Section 4.2, as follows:

"Specifically, **Figure 6**(d) shows that both the CHM_PRE V2 and V1 datasets exhibit larger absolute errors in regions such as NC, SCC, and QT compared to other areas. This is mainly attributed to the higher precipitation amounts in these regions, which naturally lead to greater absolute errors. In contrast, accuracy metrics that are not affected by the magnitude of the variables, such as KGE (**Figure 6**(e)) and RSD (**Figure 6**(f)), demonstrate better stability across different regions. The KGE and RSD in SWC and QT exhibit relatively greater variability, which could possibly be explained by the sparse distribution of precipitation observation stations and the high spatiotemporal variability of precipitation in these regions (Li et al., 2015; Liu et al., 2019)." (Lines 335–341)

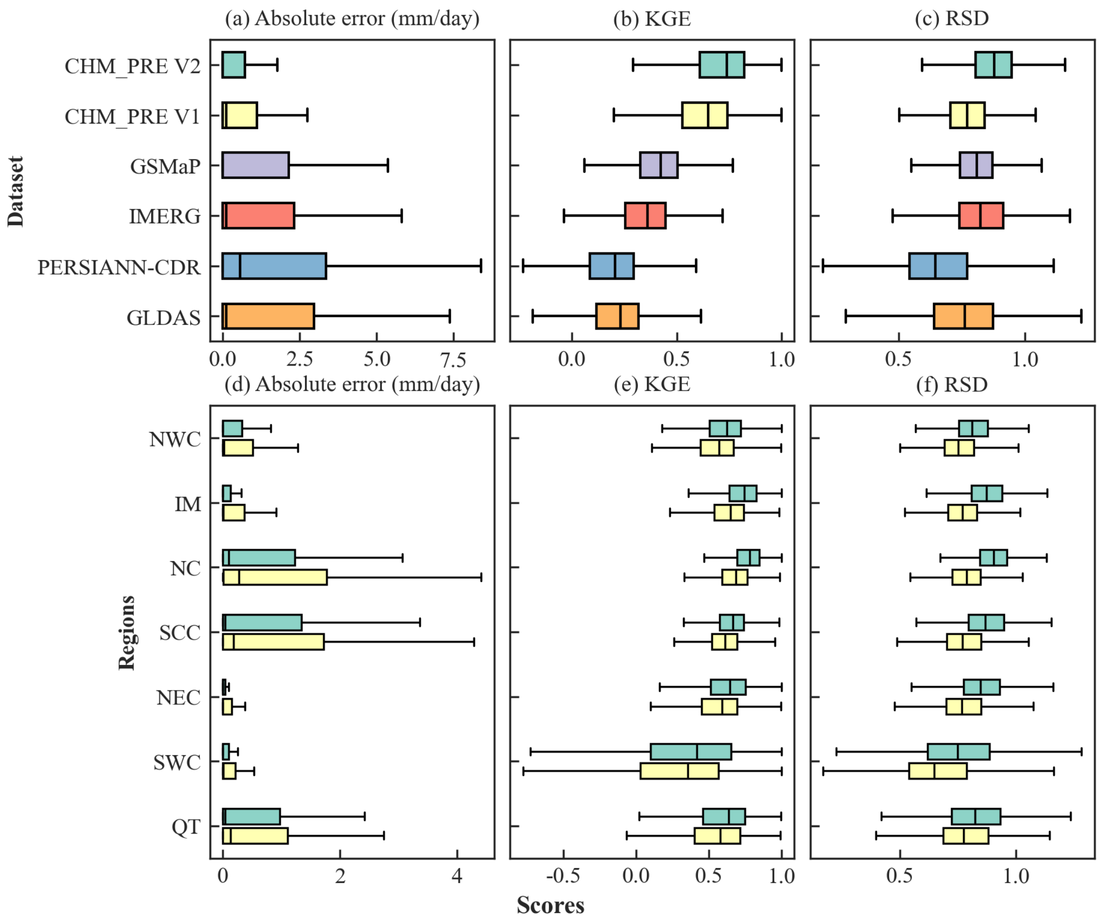

Figure 6. Accuracy of different precipitation datasets on the testing dataset CMA-HD. The green and yellow boxes in subfigures (d-f) represent CHM_PRE V2 and CHM_PRE V1, respectively. The ideal values for absolute error, KGE, and RSD are 0 mm/day, 1.0, and 1.0, respectively.

11. From the perspective of RSD (Relative Standard Deviation), it appears that GSMaP might have higher accuracy than CHM_PRE v1.0. The authors should address this observation and discuss how CHM_PRE v2.0 compares to GSMaP in terms of performance.

Response: Thank you for pointing out this issue. We carefully examined the RSD values across different precipitation datasets and found that CHM_PRE V2 achieved the best performance among all datasets (RSD takes values in the range (0, +∞), and the optimal value is 1). As shown in **Figure 6**(c), CHM_PRE V2 demonstrates superior accuracy in terms of RSD, with the median RSD across stations being closer to 1.0 compared to other datasets.

Also, we calculated the overall accuracy of precipitation values for each dataset (**Table S5**) and found that CHM_PRE V2 attained an overall RSD of 0.88, which is approximately 4.76% better than the second-best dataset, IMERG (RSD = 0.84). In fact, CHM_PRE V2 outperforms all other datasets in terms of overall MAE, KGE, and RSD. The only exception is the Bias metric, where CHM_PRE V2 (1.05) is slightly worse than GSMaP (1.04).

Table S5. Precipitation accuracy of different datasets validated by high-density gauge data. The bolded numbers in the column represent the optimal accuracy values for that metric.

Dataset Name	MAE (mm/day)	KGE	Bias	RSD
CHM_PRE V2	**1.48**	**0.79**	1.05	**0.88**
CHM_PRE V1	1.67	0.70	1.12	0.78
GSMaP	2.94	0.48	**1.04**	0.80
IMERG	3.27	0.44	1.12	0.84
PERSIANN-CDR	3.70	0.29	1.12	0.70
GLDAS	3.69	0.31	1.04	0.79

To better clarify this point, we have added the following content to the revised manuscript:

"CHM_PRE V2 achieved an overall MAE, KGE, and RSD of 1.48 mm/day, 0.79, and 0.88, respectively, outperforming other datasets by 12.84%, 12.86%, and 4.76% (Table S5 in the supplementary material)." (Lines 328–330)

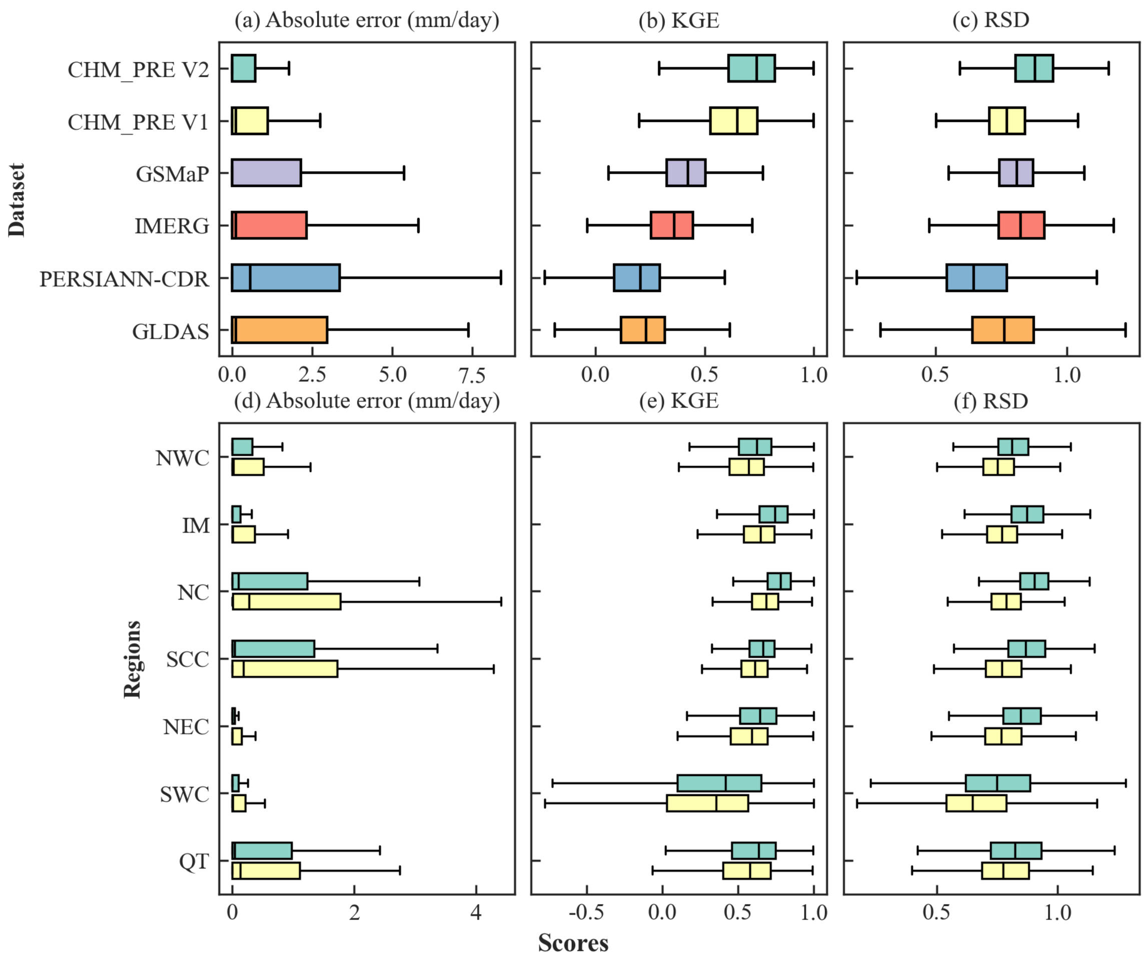

Figure 6. Accuracy of different precipitation datasets on the testing dataset CMA-HD. The green and yellow boxes in subfigures (d-f) represent CHM_PRE V2 and CHM_PRE V1, respectively. The ideal values for absolute error, KGE, and RSD are 0 mm/day, 1.0, and 1.0, respectively.

We are sincerely grateful for your insightful comments, which have significantly 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.

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.

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.

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.

Fan, C., Yin, S., and Chen, D.: Spatial correlations of daily precipitation over mainland China, International Journal of Climatology, 41, 6350–6365, https://doi.org/10.1002/joc.7199, 2021.

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.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., and Li, X.: The first high-resolution meteorological forcing dataset for land process studies over China, Scientific Data, 7, 25, https://doi.org/10.1038/s41597-020-0369-y, 2020.

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.

Li, R., Wang, K., and Qi, D.: Validating the Integrated Multisatellite Retrievals for Global Precipitation Measurement in Terms of Diurnal Variability With Hourly Gauge Observations Collected at 50,000 Stations in China, Journal of Geophysical Research: Atmospheres, 123, 10423–10442, https://doi.org/10.1029/2018JD028991, 2018.

Li, Y.-G., He, D., Hu, J.-M., and Cao, J.: Variability of extreme precipitation over Yunnan Province, China 1960–2012, International Journal of Climatology, 35, 245–258, https://doi.org/10.1002/joc.3977, 2015.

Liu, J., Shangguan, D., Liu, S., Ding, Y., Wang, S., and Wang, X.: Evaluation and comparison of CHIRPS and MSWEP daily-precipitation products in the Qinghai-Tibet Plateau during the period of 1981–2015, Atmospheric Research, 230, 104634, https://doi.org/10.1016/j.atmosres.2019.104634, 2019.

McMillan, H. K., Westerberg, I. K., and Krueger, T.: Hydrological data uncertainty and its implications, WIREs Water, 5, e1319, https://doi.org/10.1002/wat2.1319, 2018.

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.

Shen, Y., Feng, M., Zhang, H., and Gao, F.: Interpolation Methods of China Daily Precipitation Data, Journal of Applied Meteorological Science (In Chinese), 21, 279–286, 2010.

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.

Wagner, P. D., Fiener, P., Wilken, F., Kumar, S., and Schneider, K.: Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions, Journal of Hydrology, 464–465, 388–400, https://doi.org/10.1016/j.jhydrol.2012.07.026, 2012.

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.

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., and Liu, C.: A Gauge-Based Analysis of Daily Precipitation over East Asia, Journal of Hydrometeorology, 8, 607–626, https://doi.org/10.1175/JHM583.1, 2007.

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, 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.

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.