https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

Cross-temporal downscaling and fusion for hourly 0.01° precipitation estimation: A case study in Youxian District, China

Wei He[1], Jingjuan Li[1], Dian Wen[2], Jianbin Su[3], Zhengxian Zhang[4], Xiaogang Wang[1,*]

[1] Department Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China
[2] PowerChina Chengdu Engineering Co., Ltd, Chengdu 610072, China
[3] Key Laboratory of Virtual Geographic Environment (Ministry of Education of PRC), Nanjing Normal University, Nanjing 210023, China
[4] Collaborative Innovation Center of Sustainable Forestry in Southern China of Jiangsu Province, Nanjing Forestry University, Nanjing 210037, China

Correspondence to: Xiaogang Wang (xgwang@nhri.cn)

Abstract. Reliable precipitation data are essential for fine-scale hydrological applications at the regional level. Consequently, numerous studies have sought to generate high-resolution and high-accuracy precipitation products through spatial downscaling of satellite-based precipitation estimates and bias correction using ground observations. However, few such studies have considered the sub-daily scale, which holds greater application value. In this study, a cross-temporal "downscaling-fusion" framework, termed CTDF, is proposed. Both stages employ extreme gradient boosting (XGBoost) modeling: the first stage spatially downscales daily GPM precipitation from 0.1° to 0.01° using various high-resolution environmental factors, while the second stage fuses the downscaled GPM, cloud properties, and rain gauge observations to generate the final hourly precipitation estimates. With Youxian District, China as the study area, the performance of CTDF was compared against five alternative schemes, and the spatial distribution of the generated precipitation was analyzed. Results indicate that: (1) CTDF exhibits the best overall performance (CC = 0.81, MAE = 0.88 mm, RMSE = 1.95 mm, Bias = 0.4 %), mitigating the systematic underestimation inherent in the original GPM product, while omitting either stage results in performance degradation; (2) CTDF demonstrates more robust performance across different precipitation intensities and diurnal conditions; (3) CTDF substantially enhances the representation of spatial precipitation heterogeneity, increasing the coefficient of variation (CV) of GPM by 170 % and 255 % for convective and stratiform precipitation events, respectively. Overall, the two-stage collaborative design of CTDF achieves spatial refinement and accuracy improvement, providing a viable technical pathway for generating high spatiotemporal resolution precipitation products.

1 Introduction

Precipitation is a fundamental component of the global hydrological cycle, and accurate precipitation data provide a critical foundation for water resources management and flood disaster early warning systems (Hinge et al., 2022; Trenberth et al., 2007). This need is particularly pronounced in densely populated urban areas and small- to medium-sized catchments, where hydrological responses are rapid and spatial variability is significant. Such settings require precipitation data that combine both

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

high spatiotemporal resolution and high accuracy to support refined hydrological analysis and management (Saharia et al., 2021).

At present, precipitation information is primarily obtained from three sources: ground-based rain gauge observations, weather radar estimates, and satellite-based remote sensing retrievals. Rain gauge observations generally offer high measurement accuracy and are often regarded as "ground truth"; however, they represent point-based measurements whose spatial representativeness depends strongly on station density and network configuration (Kidd et al., 2017). Weather radar systems can provide precipitation estimates with relatively high spatial and temporal resolution over regional scales, but their deployment and maintenance are costly, and network coverage is constrained by terrain and infrastructure conditions (Sokol et al., 2021). In contrast, satellite-based precipitation products offer advantages in terms of broad spatial coverage, high observation frequency, and relatively weak dependence on terrain conditions, and have been extensively applied in hydrological modeling and eco-environmental studies (Ashouri et al., 2015; Huffman et al., 2007).

Despite their importance, satellite-based precipitation products still face limitations in spatial resolution and estimation accuracy, which restrict their applicability in fine-scale regional studies (Lasser et al., 2019; Woods et al., 2023). Spatial downscaling aims to disaggregate coarse-resolution data onto finer grids by exploiting relationships between the target variable and high-resolution auxiliary variables such as the digital elevation model (DEM), normalized difference vegetation index (NDVI), and land surface temperature (LST) (Kofidou et al., 2023). Given the complex and nonlinear relationships between precipitation and environmental factors, machine learning-based methods have attracted increasing attention in recent years (Gong et al., 2025; Zhu et al., 2023). It should be noted that downscaling primarily refines the spatial distribution of precipitation and does not inherently improve estimation accuracy (Zhan et al., 2018). Consequently, ground rain gauge observations are typically introduced after downscaling for bias correction, establishing a technical framework of "downscaling first, calibration second"(Duan and Bastiaanssen, 2013). As demands for greater flexibility and robustness in calibration methods have increased, machine learning approaches have gradually been incorporated into the calibration process (Chen et al., 2021).

However, the effectiveness of calibration strongly depends on rain gauge density and spatial representativeness. To address information gaps in sparsely gauged regions, many studies have increasingly adopted multi-source data fusion frameworks to generate high-quality precipitation products. Such studies typically first harmonize input datasets to a target spatiotemporal resolution through interpolation or downscaling, and then construct fusion models based on geostatistical methods, machine learning, or deep learning using ground observations as the target variable. Compared with the use of a single satellite product, fusing multiple satellite-based precipitation datasets generally leads to substantial improvements in estimation accuracy (Zhao, 2023). In regions with well-established radar networks, some studies have further developed three-source fusion frameworks that integrate radar, satellite, and gauge observations (Chen et al., 2020). Although environmental variables do not directly measure precipitation, they can provide important physical constraints, such as vegetation indices characterizing precipitation-vegetation response relationships (Zhu et al., 2024) and cloud fraction data improving precipitation estimation (Zhang et al., 2024).

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

From a temporal perspective, most studies on generating high-resolution precipitation products from remote sensing data focus on daily or longer time scales. However, applications such as regional hydrological simulation and urban flood modeling have a more urgent demand for hourly or even sub-hourly precipitation data (Cristiano et al., 2017). Hourly precipitation is often derived through temporal disaggregation strategies, but subsequent calibration using ground observations is still required to improve accuracy (Wu et al., 2025; Zhuang et al., 2023). Only a limited number of studies have attempted to directly perform multi-source data fusion at the hourly scale (Zhang et al., 2022; Zhu et al., 2022). These approaches are primarily driven by multi-source precipitation observations, and even when auxiliary variables are introduced, they are mostly static factors (e.g., DEM or geographic coordinates), which offer limited explanatory power for spatiotemporal precipitation variability at the hourly scale. In contrast, cloud properties are dynamic variables that are closely linked to precipitation formation processes, and the corresponding data products are available at high temporal resolutions consistent with satellite precipitation observations. Ma et al. (2020) demonstrated the effectiveness of cloud properties in enhancing the spatial resolution of hourly precipitation, although their linear disaggregation approach has limited capability in capturing the nonlinear characteristics of precipitation processes.

Based on the above considerations, this study proposes a cross-temporal "downscaling-fusion" framework, termed CTDF. The framework adopts differentiated modeling strategies across temporal scales: in the downscaling stage, environmental variables are used to spatially refine daily GPM precipitation, whereas in the fusion stage, downscaled GPM precipitation, cloud properties, and rain gauge observations are jointly incorporated to construct a multi-source fusion model at the hourly scale. The resulting product provides hourly precipitation estimates at a spatial resolution of 0.01°. Using Youxian District in southwestern China as a case study, the primary objective of this study is to evaluate the effectiveness and robustness of the CTDF framework, and to provide a viable technical pathway for generating hourly, high-resolution precipitation products suitable for fine-scale hydrological applications.

2 Materials and methods

2.1 Study area

Youxian District (Fig. 1) is under the jurisdiction of Mianyang City, Sichuan Province, located in the northwestern part of the Sichuan Basin (104°42′–105°09′ E, 31°21′–31°34′ N), covering an area of approximately 1,020 km². To the east, it borders the Fu River and is adjacent to Fucheng District. The region experiences a humid subtropical monsoon climate, with an annual mean temperature of 16.5°C and an average annual precipitation of approximately 1,000 mm, more than 70 % of which occurs during the summer monsoon period (June–September). The topography of the district is dominated by low hills and plains, with elevations ranging from 412 to 715 m, primarily between 500 and 600 m. The terrain generally slopes gently from the northwest to the southeast. Youxian lies within the Fu River basin, with major tributaries including Furong Creek, Weiliu River, and Xudong River. As a transitional zone from the mountainous areas of southwestern China to the Sichuan Basin, the district is subject to frequent extreme precipitation events due to the combined influence of monsoon climate and complex

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

terrain, often triggering flash floods and associated geological hazards. This context underscores the urgent need for high spatiotemporal resolution precipitation products. Historically, before 2023, the rain gauge network in the district was sparse, and real-time automated data transmission was limited, constraining the ability to make informed flood control decisions based on meteorological early warning information. Recurrent extreme precipitation events in recent years have caused substantial economic losses and posed serious threats to agricultural production and public safety.

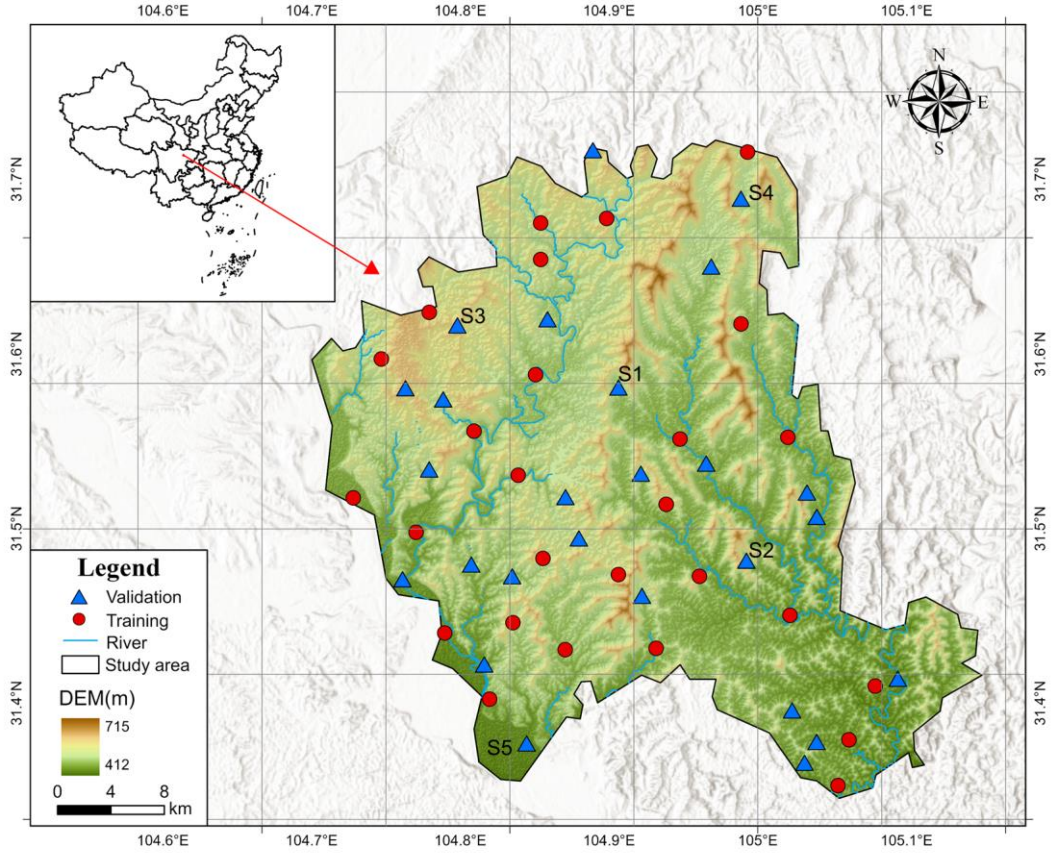


Figure 1. Topography, rain gauge distribution, and geographical location of Youxian District.

2.2 Data

In addition to ground-based observations, the datasets used in this study are listed in Table 1.

Table 2. Topography, rain gauge distribution, and geographical location of Youxian District.

Data	Dataset	Resolution
Satellite precipitation	GPM IMERG	30 min, 0.1°

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

Cloud properties	Himawari-8	10 min, 0.05°
LST	TRIMS LST	Daily, 1 km
NDVI	MOD13Q1	16 days, 250 m
DEM	SRTM DEM	-, 90 m

2.2.1 The GPM IMERG precipitation dataset

The Global Precipitation Measurement (GPM) mission, jointly launched by NASA and the Japan Aerospace Exploration Agency (JAXA) in 2014, represents the next-generation global precipitation observation system with an expanded coverage to near-global latitudes (60°S–60°N) (Huffman et al., 2014). The integrated multi-satellite retrievals for GPM (IMERG) algorithm produces globally continuous precipitation estimates with a spatial resolution of 0.1° and a temporal resolution of 30 minutes. IMERG provides three product variants—early, late, and final runs—distinguished primarily by their latency and calibration level. The final run product employs monthly ground observations for retrospective calibration, offering superior data quality and accuracy compared with near-real-time products. In this study, the 30-minute resolution IMERG final run product was utilized.

2.2.2 The cloud properties dataset

Under suitable thermodynamic conditions, cloud microphysical characteristics and associated processes directly govern precipitation formation and development (Zhao et al., 2023). Himawari-8, launched by the Japan Meteorological Agency (JMA) in 2014, provides continuous observations over East Asia and the western Pacific region with a temporal resolution of 10 minutes (Bessho et al., 2016). Official cloud products from the JAXA Himawari Monitor P-Tree system are limited to daytime coverage and are therefore insufficient for all-weather precipitation estimation. To address this, this study employed the deep learning model developed by Nie et al. (2025) to generate all-weather cloud property data, achieving nighttime retrievals with accuracy comparable to daytime observations. The extracted cloud property variables include cloud-top temperature (CTT), cloud-top height (CTH), cloud optical thickness (COT), cloud effective radius (CER), and cloud type (CTYPE), with a spatial resolution of 0.05° and a temporal resolution of 10 minutes. CTYPE is classified according to the ISCCP system (Rossow and Schiffer, 1991) into nine categories based on cloud-top pressure and optical thickness.

2.2.3 Other auxiliary datasets

Land surface temperature exhibits a notable correlation with precipitation (Trenberth and Shea, 2005). The China land and surrounding regions daily all-weather LST dataset (TRIMS LST), released by the National Tibetan Plateau Science Data Center, was utilized in this study, with a spatial resolution of 1 km and four daily observations (Tang et al., 2024).

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

Vegetation cover has been shown to significantly influence precipitation variability (Feldman et al., 2024). In this study, the NDVI was employed as the vegetation indicator. NDVI data were obtained from the MODIS Terra surface vegetation index product (MOD13Q1 Collection 6.1), which generates 16-day composites at 250 m spatial resolution (Didan et al., 2015). Topography exerts a first-order control on the spatial distribution of precipitation (Tesfa et al., 2020). Digital elevation data for this study were derived from the Shuttle Radar Topography Mission (SRTM) DEM, with a spatial resolution of 3 arc-seconds (~90 m) (Yang et al., 2011).

2.2.4 Ground observation data

Ground-based precipitation observations were provided by the Youxian District Water Resources Bureau, covering hourly precipitation records from 1 April to 31 October 2024. The rain gauges have a sensitivity of 0.5 mm h^{-1} and provide coverage across the main areas of the study region. A systematic quality control procedure was applied, including outlier detection, spatial consistency check, and temporal consistency check. After quality control, 53 rain gauge stations were retained for subsequent analysis (Fig. 1), with all retained stations exhibiting a data availability rate exceeding 75 %.

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

2.3 Methodology

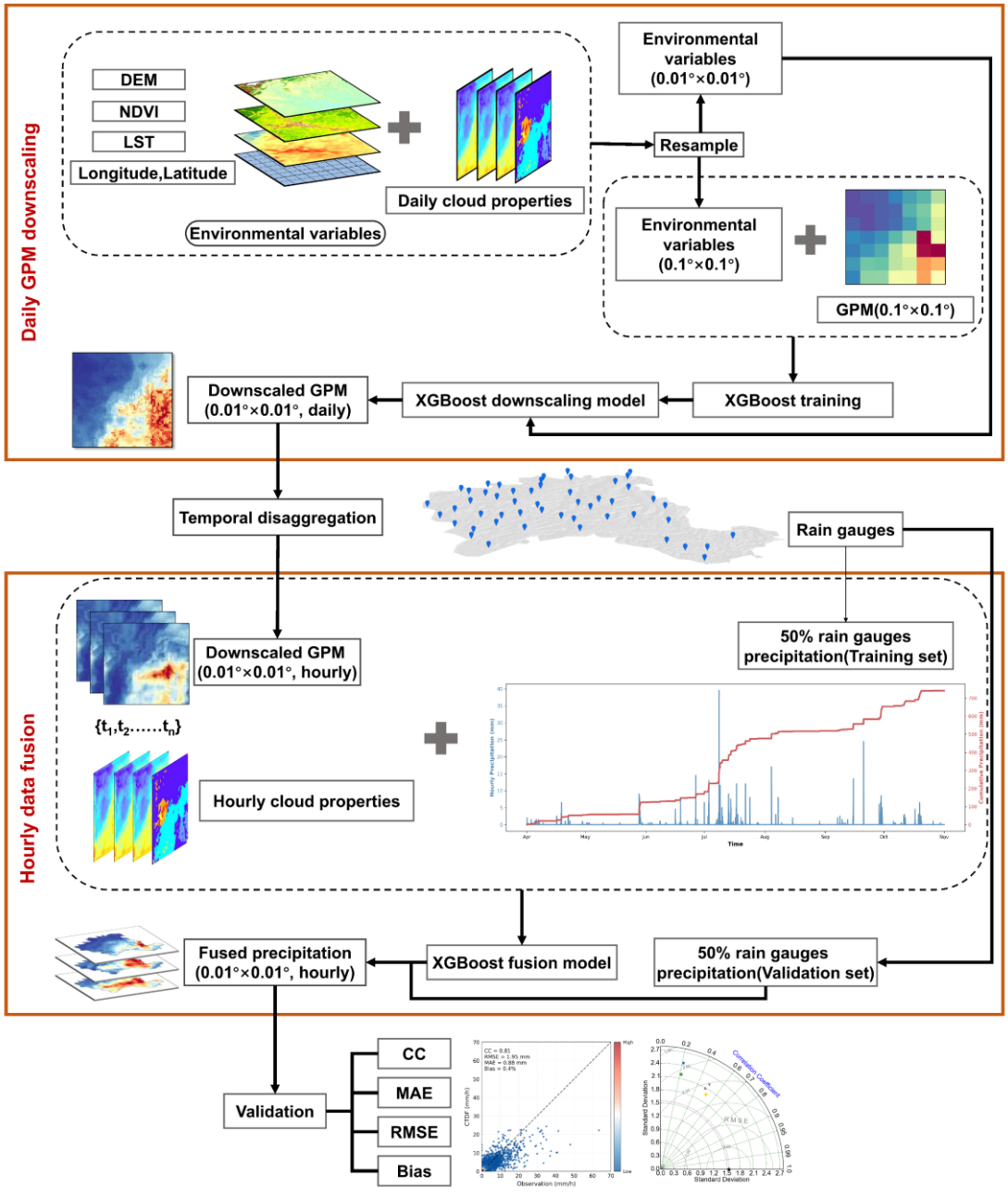


Figure 2. Workflow of the CTDF method for generating hourly precipitation gridded data at 0.01° resolution.

Figure 2 illustrates the workflow of the proposed CTDF framework. The methodology consists of four main steps:

1. Daily GPM Spatial Downscaling. First, daily GPM precipitation at 0.1° resolution is spatially downscaled using XGBoost regression models trained with daily environmental variables at the same resolution, including cloud properties, NDVI,

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

DEM, LST, and geographic coordinates. The trained model is then applied to the corresponding 0.01° resolution environmental variables to generate high-resolution daily precipitation grids (0.01° × 0.01°).

2. Temporal Disaggregation. To obtain hourly precipitation estimates, the downscaled daily precipitation is disaggregated based on the temporal variation patterns of the original half-hourly GPM data, producing hourly downscaled GPM grids at 0.01° × 0.01° resolution.
3. Hourly Multi-Source Data Fusion. Rain gauge stations are randomly split 1:1 into training and validation sets (Fig. 1). Using the observed hourly precipitation as the target variable, the corresponding hourly downscaled GPM results and cloud property features at each station are used as predictors to train an XGBoost fusion model. Once trained, the model is applied to the entire 0.01° grid to generate the final hourly precipitation product (0.01° × 0.01°).
4. Accuracy Assessment. The generated precipitation estimates are evaluated using the validation set. Four performance metrics are computed: correlation coefficient (CC), mean absolute error (MAE), root mean square error (RMSE), and bias (Bias).

2.3.1 Data preprocessing

The multi-source datasets used in this study differ in their original spatial and temporal resolutions. Moreover, the CTDF framework involves two modeling stages: daily downscaling and hourly fusion. Therefore, systematic data preprocessing is required to generate a consistent set of input variables suitable for both stages, including temporal aggregation and spatial resampling operations.

To align the daily GPM precipitation and hourly ground observations, the cloud property data at 10-minute temporal resolution were aggregated to generate eight derived features (Table 2). LST data include both daytime and nighttime observations, and the daily mean LST was computed to represent thermal conditions at the daily scale. The DEM, as a time-invariant topographic feature, remains constant throughout the study period. The NDVI is derived from 16-day composite products and is assumed to remain constant within each compositing period.

Spatial preprocessing was performed following a "training-prediction" resolution separation strategy. During the training of the GPM downscaling model, high-resolution environmental variables and cloud properties were downsampled to the original GPM resolution of 0.1° using pixel averaging. During spatial prediction, all variables were resampled to 0.01° using bilinear interpolation (with nearest-neighbor interpolation applied for cloud type). For the second-stage fusion model, corresponding pixel values at 0.01° resolution were extracted based on rain gauge coordinates and paired with station observations. All spatial operations were conducted in the WGS84 coordinate system.

Table 2. Daily/hourly derived cloud-property features.

Cloud properties (10 min)	Derived cloud features (daily/hourly)	Aggregation method
COT	COT_{mean}	Daily/hourly mean

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

CTH	CTH_{mean}	Daily/hourly mean
CER	CER_{mean}	Daily/hourly mean
CTT	CTT_{mean}	Daily/hourly mean
COT	COT_{std}	Daily/hourly standard deviation
CTH	CTH_{max}	Daily/hourly maximum
CTT	CTT_{min}	Daily/hourly minimum
CTYPE	$CTYPE_{mode}$	Daily/hourly mode

2.3.2 XGBoost model with Optuna-based hyperparameter optimization

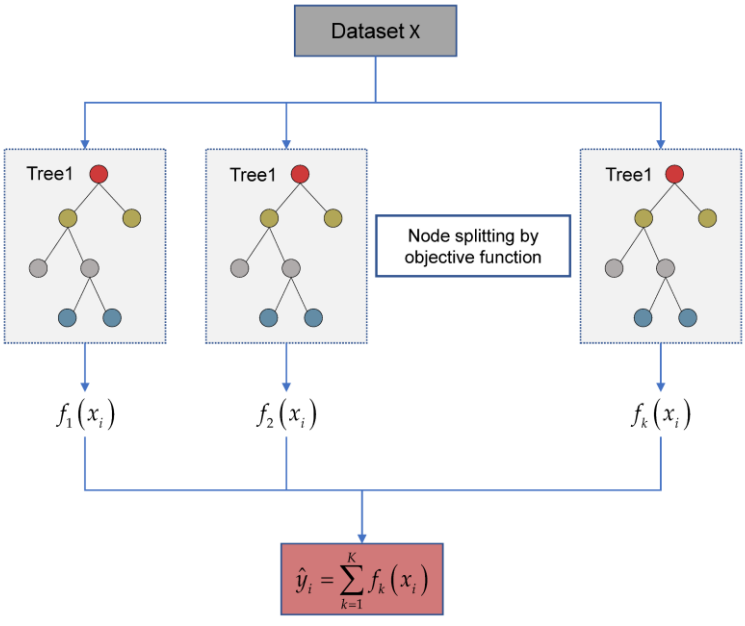


Figure 3. Schematic diagram of the XGBoost model.

XGBoost (Extreme Gradient Boosting) (Fig. 3), proposed by Chen and Guestrin (2016), is an efficient ensemble learning algorithm based on the Gradient Boosting Decision Tree (GBDT) framework. The algorithm iteratively constructs multiple weak learners, typically decision trees, and combines them linearly to form a strong predictive model. XGBoost has demonstrated excellent performance in modeling complex nonlinear relationships and handling high-dimensional data. Assume the training dataset consists of n samples, represented as $\{(x_1, y_1), (x_2, y_2) \cdots (x_n, y_n)\}$. After k iterations, the model prediction can be expressed as:

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \quad (1)$$

Here, f_k represents the k-th regression tree, $\hat{y}_i$ and denotes the predicted value for sample i. Each tree f_k belongs to the space of regression trees, which includes all possible tree structures and their corresponding leaf weights. At the t-th iteration, XGBoost optimizes an objective function that consists of a loss term and a regularization term:

$$\mathrm{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (2)$$

Here, $l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right)$ denotes the loss function, which measures the discrepancy between the predicted value and the true target. $\hat{y}_i^{(t-1)}$ represents the prediction from the previous $t-1$ iterations, and $f_t(x_i)$ is the contribution of the t-th tree to the prediction for sample i. $\Omega(f_t)$ is the regularization term, which represents one of the key improvements of XGBoost over traditional GBDT. It is used to control the complexity of the model and prevent overfitting.

The XGBoost model involves numerous hyperparameters (Table 3), which directly affect its fitting ability and generalization performance. To efficiently identify the optimal hyperparameter combination, this study employed automated hyperparameter optimization using the Optuna framework, which utilizes the Tree-structured Parzen Estimator (TPE) algorithm (Akiba et al., 2019). The optimal hyperparameter configuration was determined through iterative evaluation using XGBoost training combined with cross-validation.

Table 3. Main hyperparameters of the XGBoost model.

Parameters	Description
learning_rate	Controls the contribution (weight) of each tree
n_estimators	Number of boosting rounds (trees)
max_depth	Maximum depth of each tree, controlling model complexity
min_child_weight	Minimum sum of instance weights required in a child node, controlling node splitting
subsample	Fraction of training samples randomly sampled for each tree

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

2.3.3 Temporal disaggregation

To convert daily precipitation to an hourly scale, this study employed a proportional allocation method based on the temporal structure of the original GPM data(Ma et al., 2022). This approach assumes that the spatial downscaling process does not alter the intra-day temporal distribution of precipitation. The calculation can be expressed as follows:

$$P_{\text{hourly}}(x,y,h) = P_{\text{daily}}(x,y) \times R(h) \tag{3}$$

$$R(h) = \frac{\sum P_{GPM}^{0.1^\circ}(h)}{\sum P_{GPM}^{0.1^\circ}(day)} \tag{4}$$

In the equation, $P_{\text{hourly}}(x,y,h)$ represents the downscaled precipitation (mm) at hour h for the 0.01° grid cell (x,y), and $P_{\text{daily}}(x,y)$ is the corresponding downscaled daily precipitation (mm) for that grid cell. $R(h)$ is the temporal allocation ratio calculated from the original GPM product. Specifically, $P_{GPM}^{0.1^\circ}(h)$ denotes the total precipitation (mm) of the original 0.1° GPM product within the study area at hour h, obtained by summing the two half-hourly observations within that hour. $P_{GPM}^{0.1^\circ}(day)$ is the total daily precipitation (mm) corresponding to that hour, calculated by summing all 0.1° grids in the study area.

2.3.4 Validation metrics

Using an independent validation set (50 % of rain gauge observations), the accuracy of the fused precipitation result was evaluated with four statistical metrics (Yin et al., 2024): correlation coefficient (CC), mean absolute error (MAE), root mean square error (RMSE), and relative bias (Bias). The calculation formulas are as follows:

$$CC = \frac{\sum_{i=1}^{n}(G_i - \bar{G})(F_i - \bar{F})}{\sqrt{\sum_{i=1}^{n}(G_i - \bar{G})^2} \times \sqrt{\sum_{i=1}^{n}(F_i - \bar{F})^2}} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|F_i - G_i| \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - G_i)^2} \tag{7}$$

$$Bias = \frac{\sum_{i=1}^{n}(F_i - G_i)}{\sum_{i=1}^{n} G_i} \times 100\% \tag{8}$$

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

In the equations, G_i denotes the ground-observed precipitation (mm) at the i-th validation station, and F_i represents the estimated precipitation (mm) from the fused precipitation result at the grid cell corresponding to the station location. $\bar{G}$ and $\bar{F}$ are the mean values of the observed and estimated precipitation, respectively, and n is the total number of validation samples.

3 Results and discussion

3.1 Selection of cloud property features

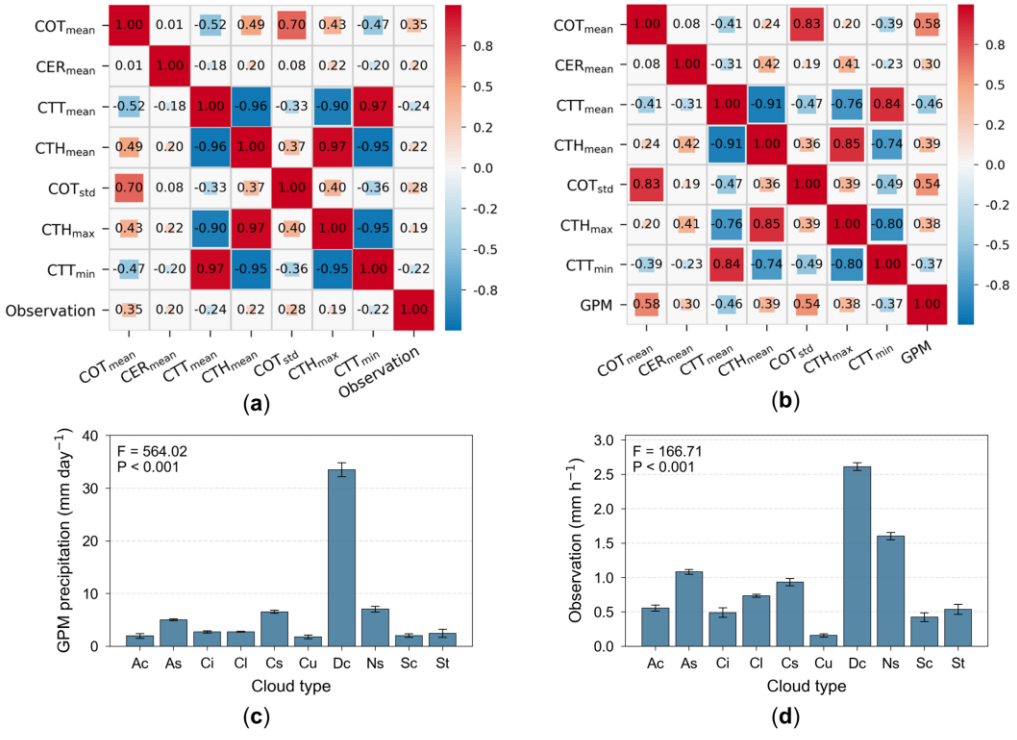


Figure 4. Cloud-property feature analysis at daily (a, c) and hourly (b, d) scales: (a, b) Pearson correlation matrices between cloud-property features and precipitation; (c, d) mean precipitation by cloud type with ANOVA results. Cloud type abbreviations: Ac = altocumulus, As = altostratus, Ci = cirrus, Cl = clear sky, Cs = cirrostratus, Cu = cumulus, Dc = deep convective, Ns = nimbostratus, Sc = stratocumulus, St = stratus.

Prior to model training, Pearson correlation analysis and one-way analysis of variance (ANOVA) were employed to evaluate the association between derived cloud property features and precipitation variables. Considering scale dependency and differences in target variables, independent feature selection was conducted for each modeling stage (Fig. 4).

At the daily scale, COT_{mean} exhibits the strongest positive correlation with GPM precipitation ($r = 0.58$, $p < 0.001$), while CTT_{mean} shows a significant negative correlation ($r = -0.46$, $p < 0.001$). Strong intercorrelations exist among some features

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

(e.g., r = 0.83 between COT_{mean} and COT_{std}, r = −0.91 between CTT_{mean} and CTH_{mean}), indicating information redundancy. Based on a correlation threshold of $|r| \geq 0.30$ and preferential selection of features with stronger target correlations within highly correlated groups, three numerical features—COT_{mean}, CTT_{mean}, and CER_{mean}—were retained. At the hourly scale, correlations are generally weaker, with COT_{mean} decreasing by approximately 40 % relative to the daily scale. Given this complexity, a relaxed threshold ($|r| \geq 0.20$) was adopted. COTstd was retained despite its high correlation with COT_{mean}, as hourly precipitation may be more sensitive to cloud spatial heterogeneity. Ultimately, COT_{mean}, COT_{std}, and CTT_{mean} were selected for the hourly fusion model.

The ANOVA results indicate highly significant precipitation differences among cloud types at both scales (daily: F = 564.02, $p < 0.001$; hourly: F = 166.71, $p < 0.001$), with deep convective clouds (Dc) associated with the highest precipitation. Despite reduced discriminative capability at the hourly scale, $CTYPE_{mode}$ was incorporated as a categorical feature in both stages.

3.2 Results of daily GPM downscaling

Although the daily GPM downscaling stage does not introduce new observations to correct systematic biases, it provides prior information with fine spatial structure for subsequent fusion.

The dataset was randomly divided into training and testing sets at a ratio of 3:1, and five-fold cross-validation was used for hyperparameter optimization. Figure 5 presents the prediction performance on both sets. The training set achieved $R^2 = 0.911$, RMSE = 3.13 mm, and MAE = 1.41 mm using out-of-fold (OOF) prediction. The independent testing set yielded comparable results ($R^2 = 0.916$, RMSE = 3.09 mm, MAE = 1.38 mm), indicating good generalization ability without overfitting.

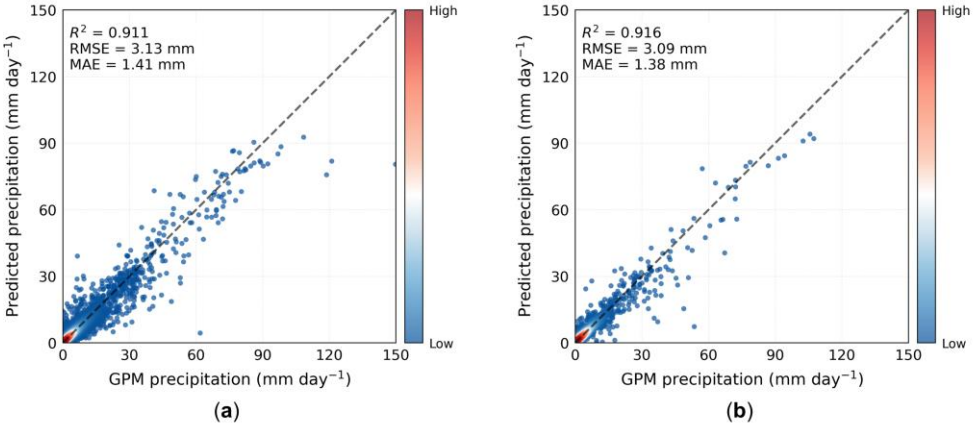


Figure 5. Scatter plots of daily GPM precipitation downscaling results: (a) out-of-fold (OOF) predictions for the training set, (b) predictions for the independent test set.

Figure 6 compares the spatial distributions of the original GPM precipitation (0.1°) and the downscaled precipitation (0.01°) for four representative dates covering different precipitation intensities. For light precipitation events (20 April and 18 October 2024), the downscaled results preserve the overall spatial pattern while exhibiting a smooth precipitation field. For heavy

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

precipitation events (8 July and 9 August 2024), the high-precipitation regions and pronounced spatial gradients are well preserved, with more detailed spatial variability revealed within each original 0.1° grid cell, although some blocky artifacts are observed.

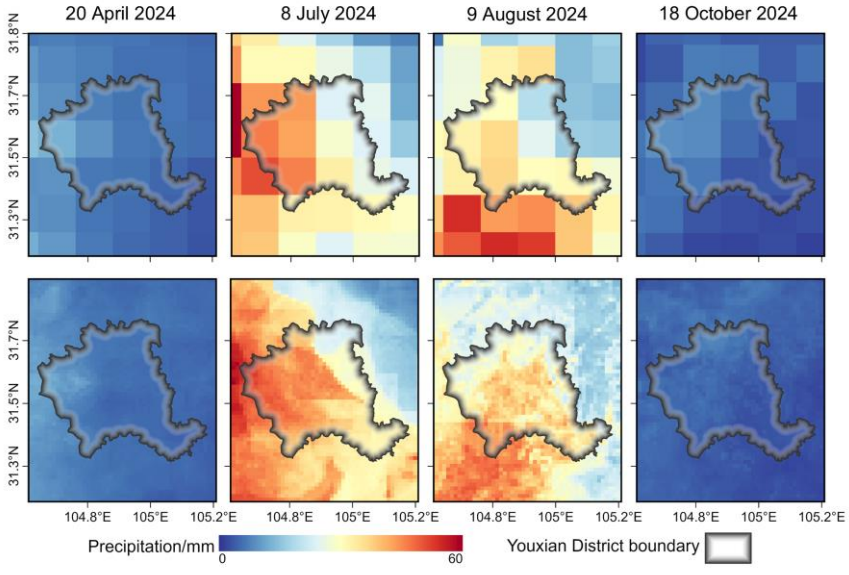


Figure 6. Spatial comparison between original GPM and downscaled GPM precipitation for four representative rainfall days (20 April 2024, 8 July 2024, 9 August 2024, and 18 October 2024): (a) original GPM precipitation, (b) downscaled GPM precipitation.

3.3 Accuracy validation of hourly 0.01° precipitation results

To validate the performance of the CTDF method, it was systematically compared with five benchmark schemes:

1. Interpolated GPM: the original hourly GPM resampled using bilinear interpolation, serving as an approximate representation of the original GPM data;
2. GMWWDA： an hourly downscaling method based on moving-window weight decomposition of cloud attributes (Ma et al., 2020), serving as a representative benchmark for hour-by-hour downscaling;
3. Downscaled GPM: the GPM product processed only through the first-stage spatial downscaling and temporal disaggregation, to examine the effectiveness of GPM downscaling alone;`
4. Downscaled GPM_GDA: Geographic Differential Analysis (GDA) applied to Downscaled GPM at hourly scale based on training stations, with errors interpolated to 0.01° resolution using Inverse Distance Weighting (IDW) for calibration, representing a traditional and widely adopted "downscaling–calibration" approach;
5. GPM + cloud fusion: the second-stage fusion model applied directly to Interpolated GPM without performing the downscaling stage, to examine the effectiveness of omitting the downscaling step.

All schemes were evaluated using the same set of validation stations.

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

3.3.1 Overall accuracy assessment

Figure 7 presents the scatterplots and validation metrics for the six schemes. The CTDF method achieves the best performance across all metrics (CC = 0.81, RMSE = 1.95 mm, MAE = 0.88 mm, Bias = 0.4 %). Since all three schemes use the original GPM as the downscaling target, Interpolated GPM, GMWWDA, and Downscaled GPM yield highly similar performance metrics (CC ranging from 0.40 to 0.41, RMSE from 3.02 mm to 3.04 mm), and all exhibit substantial underestimation (Bias ≤ −18 %). GMWWDA is suited to hourly-scale conditions where rainfall pixels are limited and regression-based modeling is difficult, whereas Downscaled GPM relies on XGBoost regression and requires a sufficient number of valid rainfall samples. After introducing ground observation-based calibration, the performance of Downscaled GPM_GDA improves markedly (CC = 0.74, RMSE = 2.18 mm, Bias = −4.9 %). The performance of the GPM + cloud fusion scheme is comparable to that of CTDF, and both outperform the other three schemes by a large margin. However, because GPM + cloud fusion skips the first-stage downscaling, its performance remains slightly inferior to that of the complete two-stage framework (a CC difference of 0.02 and an RMSE difference of 0.05 mm), indicating that the first-stage downscaling provides an auxiliary contribution to the final accuracy.

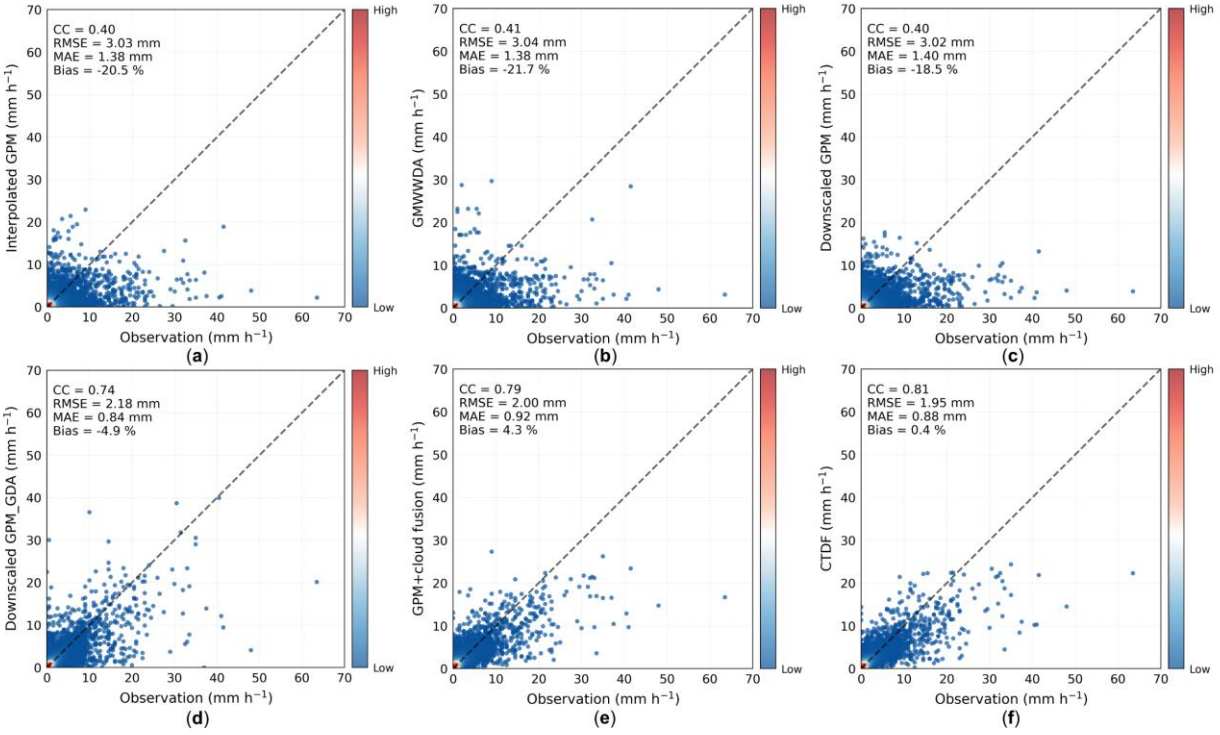


Figure 7. Scatter plots of hourly precipitation estimates at 0.01° resolution generated by different schemes: (a) Interpolated GPM, (b) GMWWDA, (c) Downscaled GPM, (d) Downscaled GPM_GDA, (e) GPM + cloud fusion, (f) CTDF.

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

3.3.2 Accuracy assessment across different precipitation intensity classes

To further evaluate the performance of each scheme under different precipitation intensities, hourly precipitation was classified into three categories: light rain (<2 mm h^{-1}), moderate to heavy rain (2–8 mm h^{-1}), and rainstorm and above (≥8 mm h^{-1}). Table 4 and Fig. 8 present the validation metrics and Taylor diagrams for each precipitation class. The performance of all schemes exhibits clear dependence on precipitation intensity. Light rain events show the largest estimation errors, with low CC values for all methods (0.19–0.45) and pronounced overestimation (Bias > 95 %), which may be related to the measurement limitation of rain gauges with a sensitivity threshold of 0.5 mm h^{-1}. For moderate to heavy rain events, CC values increase markedly (from 0.21 to 0.52), while Bias becomes negative, indicating a tendency to underestimate medium- and high-intensity precipitation (Zhuang et al., 2023). For rainstorm events, RMSE increases sharply, and the degree of underestimation intensifies, constrained by both the reduced sensitivity of satellite sensors to heavy precipitation (Chen et al., 2023) and the limited training sample size for extreme rainfall intensities (Camps-Valls et al., 2025).

Table 4. Performance comparison of hourly precipitation estimates at 0.01° resolution produced by different approaches across rainfall intensity classes.

Intensity Class	Scheme	CC	MAE	RMSE	Bias (%)
Light (<2 mm h^{-1})	Interpolated GPM	0.20	0.72	1.36	159.2
	GMWWDA	0.19	0.71	1.40	151.7
	Downscaled GPM	0.19	0.75	1.43	175.6
	Downscaled GPM_GDA	0.35	0.44	1.10	96.0
	GPM+cloud fusion	0.42	0.55	0.96	149.1
	CTDF	0.45	0.51	0.88	132.7
Moderate-to-heavy (2–8 mm h^{-1})	Interpolated GPM	0.21	2.37	3.01	-40.8
	GMWWDA	0.21	2.41	3.07	-41.7
	Downscaled GPM	0.21	2.33	2.85	-41.9
	Downscaled GPM_GDA	0.50	1.42	2.03	-14.4
	GPM+cloud fusion	0.48	1.44	1.97	-12.4
	CTDF	0.52	1.38	1.85	-14.7
Rainstorm and above (≥8 mm h^{-1})	Interpolated GPM	0.20	10.30	12.44	-73.6
	GMWWDA	0.22	10.21	12.34	-72.5
	Downscaled GPM	0.23	10.32	12.39	-74.1
	Downscaled GPM_GDA	0.43	6.38	8.82	-36.8
	GPM+cloud fusion	0.51	5.95	8.12	-38.0
	CTDF	0.53	6.05	8.10	-38.5

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

Across all precipitation classes, CTDF consistently delivers the best performance, with the Taylor diagrams showing that it is always closest to the observations. The GPM + cloud fusion scheme performs comparably to CTDF for rainstorm events, indicating that hourly data fusion substantially improves the estimation of heavy precipitation. However, its performance across different precipitation classes is less stable than that of CTDF, primarily because the fusion model lacks the constraints from environmental variables that govern spatial precipitation distribution in the downscaling model (Zhu et al., 2023). Downscaled GPM_GDA maintains standard deviations close to those of the observations across different precipitation classes, although its performance also exhibits notable fluctuations because the error structure of precipitation varies nonlinearly with intensity (Chen et al., 2021).

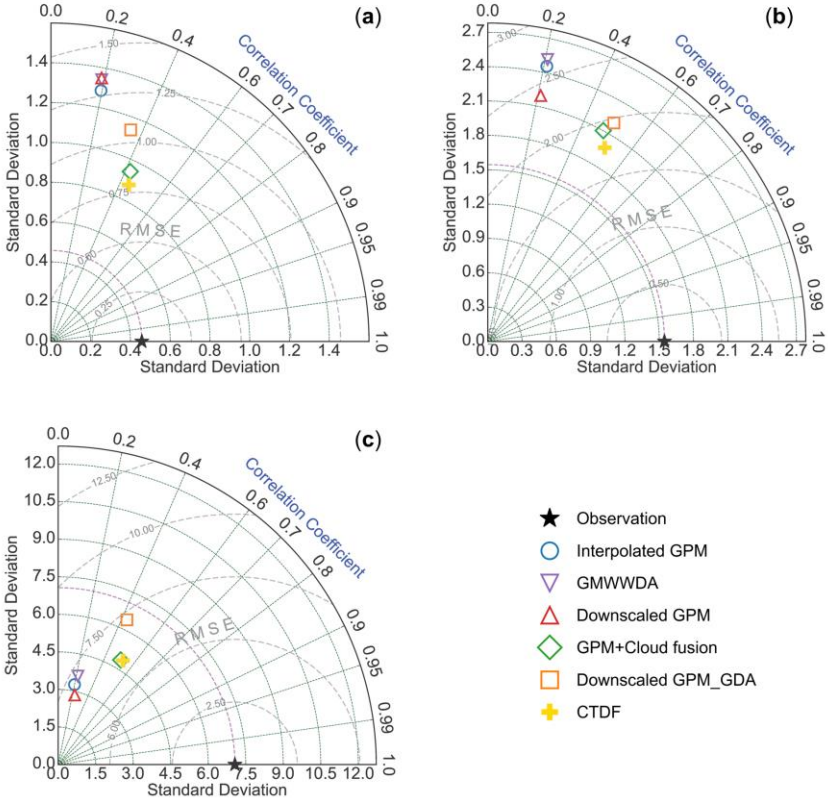


Figure 8. Accuracy comparison of hourly precipitation estimates at 0.01° resolution produced by different schemes across rainfall intensity classes: (a) light rain (<2 mm h^{-1}), (b) moderate to heavy rain (2–8 mm h^{-1}), (c) rainstorm and above (≥8 mm h^{-1}).

3.3.3 Daytime and nighttime accuracy assessment

To evaluate the robustness of each scheme under diurnal variations, the precipitation data were divided into daytime (22:00–10:00 UTC) and nighttime (10:00–22:00 UTC) groups. Figure 9 compares the performance metrics of the six schemes during

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

daytime and nighttime periods. For all schemes, CC values decrease at nighttime, although the magnitude of the decline varies substantially. Interpolated GPM and Downscaled GPM experience CC reductions exceeding 40 %, indicating pronounced performance degradation. In contrast, CTDF exhibits a CC reduction of only 3.8 % (from 0.81 to 0.78), maintaining the highest CC values during both daytime and nighttime. CTDF achieves the lowest RMSE among all schemes in both periods and maintains Bias values close to zero, whereas the Bias of GPM + cloud fusion increases from 2.8 % during the daytime to 7.2 % at nighttime. Previous studies have demonstrated that GPM precipitation products exhibit larger errors at nighttime than during the daytime (Wang et al., 2023). Overall, CTDF demonstrates notably superior performance at nighttime compared to other schemes, minimizing the accuracy gap between daytime and nighttime precipitation estimation.

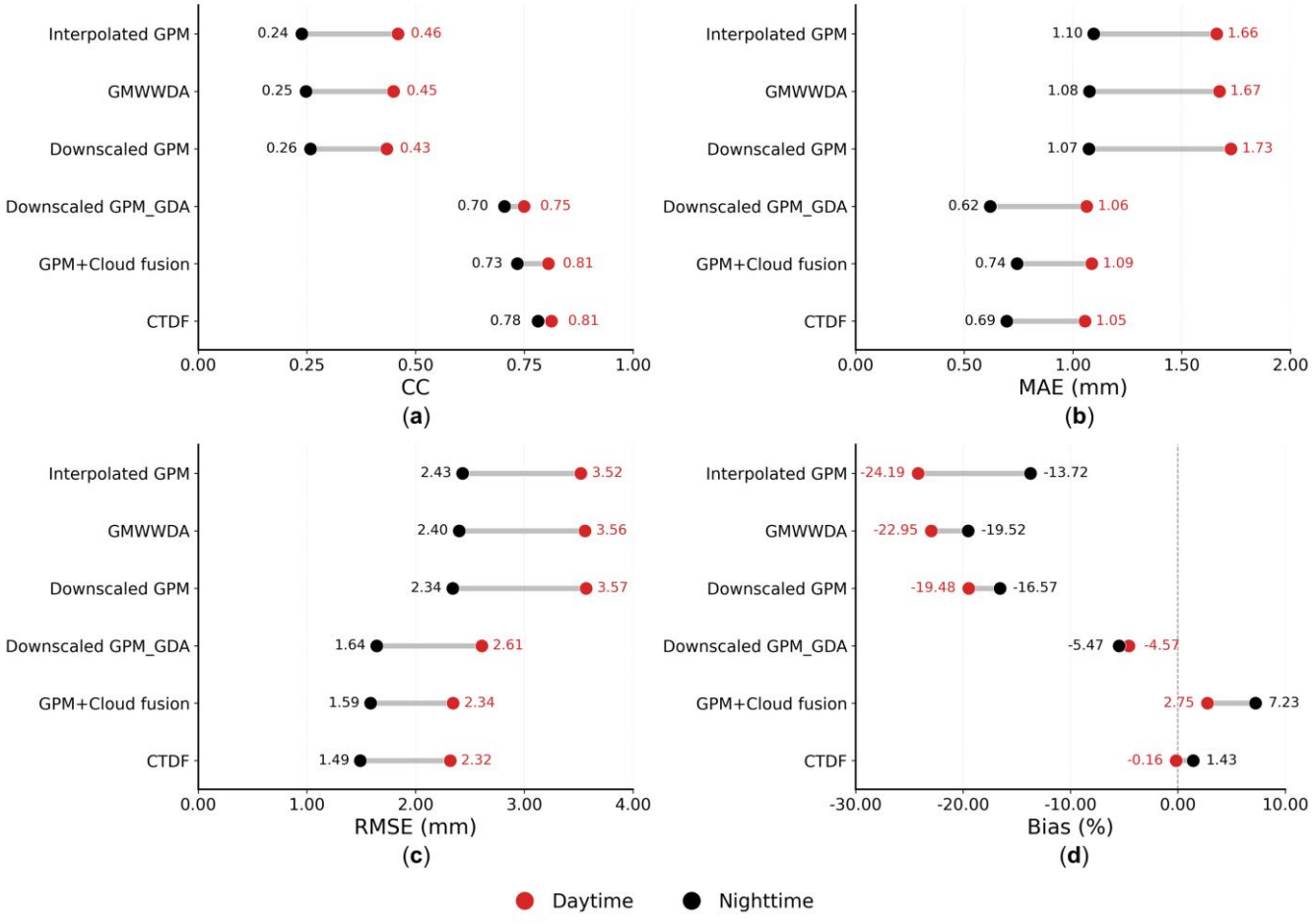


Figure 9. Daytime-nighttime performance comparison of hourly precipitation estimates at 0.01° resolution generated by different schemes: (a) CC, (b) MAE, (c) RMSE, (d) Bias (%).

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

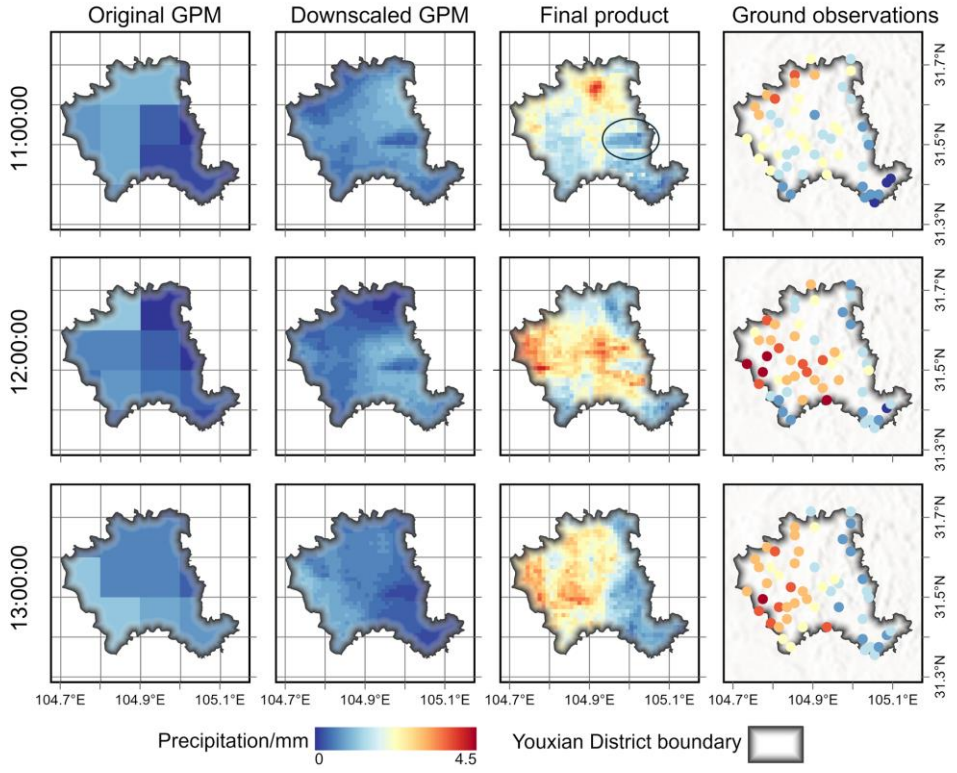
3.4 Spatial distribution of hourly 0.01° precipitation

3.4.1 Spatial patterns of representative events

To visually demonstrate the performance of the two-stage framework, two representative precipitation events were selected for case analysis: a moderate rainfall event on 28 May 2024 (Fig. 10) and a rainstorm event on 8 July 2024 (Fig. 11). For each event, the main precipitation period was extracted, and the hourly evolution of spatial precipitation distributions at different stages of the CTDF method (original GPM, downscaled GPM, and final product) was compared.

The two events exhibit distinct precipitation characteristics. The 28 May event is characterized by relatively weak precipitation (maximum of 4.5 mm h^{-1}) and gentle spatial gradients, whereas the 8 July event features intense precipitation (maximum of 35 mm h^{-1}) with a pronounced precipitation center. The original GPM product captures the general spatial distribution of precipitation but exhibits evident underestimation of precipitation intensity, which becomes more pronounced as precipitation intensity increases. The first-stage spatial downscaling mainly enriches fine-scale (<0.1°) spatial details while largely preserving the large-scale precipitation patterns of the original GPM product. After the second-stage fusion, both precipitation intensity and spatial distribution become much closer to observations. For the July convective event, the final product clearly captures the complete evolution of the heavy precipitation center. For the May stratiform event, the final product clearly depicts the spatial distribution of precipitation intensity with more distinct boundaries between local high- and low-precipitation areas. Consistent with the pattern shown in Fig. 6, the final downscaled grids also exhibit artifacts arising from the discontinuity of the coarse-resolution background field, as marked by the black circles in Fig. 10 and Fig. 11. Similar artifacts can also be visually identified in the spatial distribution results presented by other downscaling studies (He et al., 2025; Zhu et al., 2023). Because CTDF treats each pixel independently, the grid structure of the coarse-resolution input tends to persist in the output when the fine-scale information provided by the auxiliary variables is insufficient (Li and Shao, 2010).

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026



360 **Figure 10.** Spatial distribution of precipitation from 11:00 to 13:00 UTC on 28 May 2024.

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

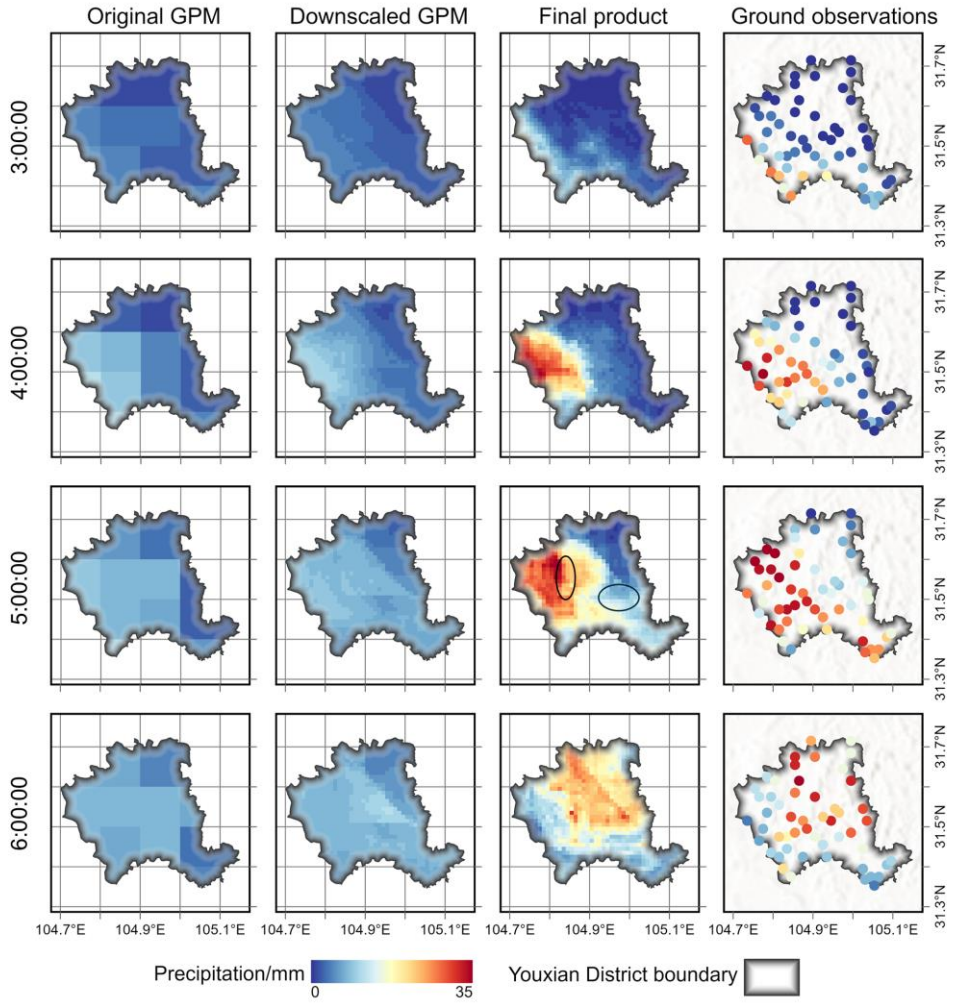


Figure 11. Spatial distribution of precipitation from 03:00 to 06:00 UTC on 8 July 2024.

3.4.2 Spatial heterogeneity analysis under different cloud-type-dominated precipitation

To further quantitatively evaluate the ability of each stage to represent the spatial heterogeneity of precipitation, precipitation events dominated by convective clouds (six events) and stratiform clouds (eight events) were separately selected. Based on the coordinates of observation stations, the spatial coefficient of variation (CV) at each time step was calculated to quantify the degree of spatial heterogeneity.

Figure 12 presents the statistical distributions of precipitation CV for the three stages and observations under the two cloud-type regimes. According to the observed CV, convective precipitation (median CV = 1.06) exhibits substantially higher spatial heterogeneity than stratiform precipitation (median CV = 0.97), reflecting the localized nature of convective systems (Houze, 1997). Both the Original GPM and the Downscaled GPM substantially underestimate the spatial heterogeneity of precipitation. For convective events, the median CV of the Original GPM is only 0.24, representing a 77 % underestimation relative to

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

observations. In contrast, the CTDF method markedly alleviates this deficiency, with median CV reaching 0.65 for convective events (a 170 % increase relative to the Original GPM) and 0.70 for stratiform events (a 255 % increase). The results indicate that the first-stage downscaling can improve the representation of spatial heterogeneity to some extent, with this capability being more pronounced for stratiform events, suggesting a more stable relationship between environmental factors and stratiform precipitation. Convective precipitation is regulated by microphysical processes that cannot be inferred from macroscale environmental factors (Rosenfeld and Lensky, 1998). The second stage, by establishing relationships between cloud properties and observed precipitation, can capture hourly local physical processes to a certain extent.

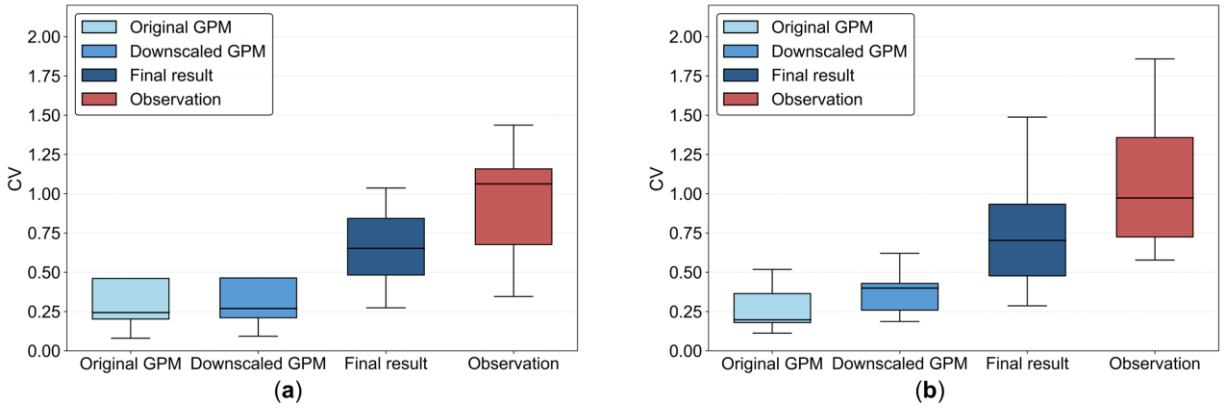


Figure 12. Comparison of the coefficient of variation (CV) for precipitation events dominated by different cloud types: (a) convective cloud-dominated events, (b) stratiform cloud-dominated events.

3.5 Independent validation using 2023 precipitation events

The CTDF was trained and validated exclusively on data samples from April to October 2024; consequently, its adaptability to variable climatic conditions remains uncertain. Owing to constraints in the station network construction schedule, the number of available rainfall stations in the study area in 2023 was substantially lower than in 2024, and anomalies were detected at several stations. To evaluate the model's predictive performance across additional precipitation periods, quality control was applied to the 2023 observational data, and five spatially well-distributed stations (S1–S5; Fig. 1) were selected, together with hourly observation records from four representative precipitation events. In accordance with GB/T 28592–2012 (Standardization Administration of China, 2012), these events span four intensity grades—moderate rain, heavy rain, torrential rain, and heavy torrential rain—thereby providing independent validation samples for assessing the model's cross-year generalization ability across different precipitation patterns.

Table 5 presents a comparison of the overall accuracy between CTDF predictions and GPM estimates across four precipitation events at five independent stations in 2023. Relative to GPM, the correlation coefficient (CC) of CTDF increased from 0.66 to 0.82, while the MAE and RMSE decreased by approximately 20 % and 26 %, respectively. The absolute Bias was reduced from 49.6 % to 7.3 %. This overall improvement is consistent with the findings from the 2024 validation set, indicating that

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

the accuracy gains of the CTDF method over GPM are preserved under independent cross-year scenarios. Figure 13 further compares the hourly time series of CTDF predictions, GPM estimates, and gauge observations for each precipitation event. Across all events, CTDF reasonably captures the onset and cessation of precipitation as well as intensity fluctuations, exhibiting a markedly better fit to the observations than GPM. For the heavy rain (20–21 April 2023) and torrential rain (10–11 September 2023) events, CTDF predictions closely match the observations in both magnitude and temporal evolution, with the best overall agreement obtained for the torrential rain event. Owing to the limited representation of extreme precipitation in the 2024 training samples, the model substantially underestimates the precipitation intensity during the heavy torrential rain event (26–27 July 2023); the magnitude of underestimation increases with rainfall intensity and reaches its maximum at the peak rainfall rate. For the moderate rain event (10–11 October 2023), the model exhibits substantial overestimation over multiple time periods, consistent with the finding in Section 3.3.2 that the model shows considerable bias under light precipitation conditions. In contrast, GPM fails to adequately capture variations in precipitation intensity and is unable to reproduce the magnitude and peak timing during heavy rainfall periods.

Table 5. Daily/hourly derived cloud-property features.

Scheme	CC	MAE (mm)	RMSE (mm)	Bias (%)
GPM	0.66	2.33	4.17	−49.6
CTDF	0.82	1.86	3.09	−7.3

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

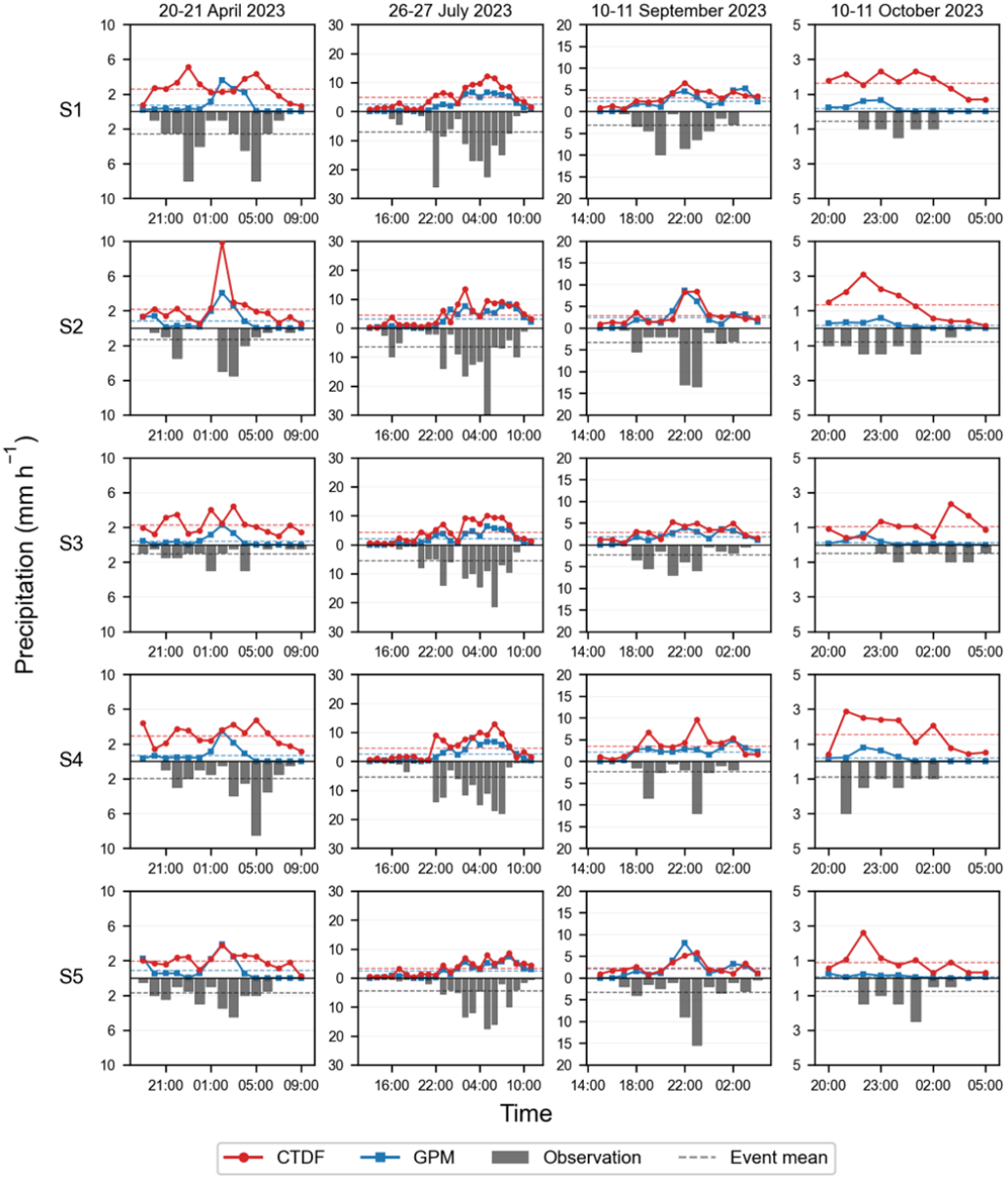


Figure 13. Hourly time series comparison of CTDF predictions, GPM estimates, and gauge observations at five validation stations during four representative precipitation events in 2023.

4 Conclusions and future work

The CTDF method proposed in this study is a two-stage framework that performs "downscaling first, fusion second" across temporal scales to generate hourly precipitation products at 0.01° resolution. First, to validate the effectiveness and robustness of the proposed method in the study area, its performance was compared with five other schemes, with this comparison

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

designed to demonstrate the advantages and rationality of the method design. Then, to demonstrate the capability of CTDF in representing spatial information of hourly precipitation, the results were visualized in terms of spatial distribution for representative precipitation events, and differences in spatial heterogeneity representation capability across stages were analyzed for different dominant cloud types. Finally, the CTDF model trained on 2024 data was applied to precipitation estimation in 2023 to examine its generalization ability on independent-year data. The main conclusions of this study are as follows:

1. The proposed method effectively produces high-accuracy hourly precipitation products. Compared with five benchmark schemes, it achieves the best performance (CC = 0.81, MAE = 0.88 mm, RMSE = 1.95 mm, Bias = 0.4 %), substantially mitigating the systematic underestimation inherent in the original GPM products. Each stage of the CTDF framework demonstrates distinct advantages, and the capability of XGBoost to capture nonlinear relationships makes it superior to traditional calibration methods.
2. The proposed method exhibits strong robustness under varying conditions. Across three precipitation intensity classes as well as during daytime and nighttime periods, CTDF maintains the best performance with the smallest variability, indicating good adaptability to changes in precipitation intensity and diurnal data differences.
3. From the spatial visualization results, the first-stage downscaling of the proposed method primarily provides refined spatial priors, whereas the second-stage data fusion directly improves intensity estimation. For precipitation events dominated by convective clouds and stratiform clouds, CTDF substantially reduces the gap in spatial variability compared to ground observations, demonstrating stronger capability in representing spatial heterogeneity relative to GPM.
4. Under the independent cross-year scenario, CTDF retains a reasonable capability for precipitation estimation, with its accuracy still markedly superior to that of GPM (CC = 0.82, MAE = 1.86 mm, RMSE = 3.09 mm, Bias = −7.3 %). Nevertheless, the performance of the model under light and extreme precipitation conditions remains constrained by the limited temporal coverage and uneven distribution of the training samples.

Although the proposed method demonstrates strong performance and distinct advantages in the study area, further improvements and extensions are still required in the following aspects:

1. Due to the 0.5 mm sensitivity threshold of ground-based rain gauges, the measurement accuracy for light precipitation is limited, introducing uncertainty into the validation reference itself. In future studies, higher-precision observations (e.g., rain gauges with 0.1 mm sensitivity) could improve the quality of training labels and validation benchmarks for light rain events.
2. The representativeness of the samples used in this study remains limited: spatially, only Youxian District is covered, which is characterized by a subtropical monsoon climate and shallow-hill terrain; temporally, the data are mainly drawn from April to October 2024. Although the temporal generalization ability of the model was preliminarily evaluated using independent samples from 2023, further efforts are still needed to extend the study area to more diverse climatic zones and terrain types, and to progressively expand the training samples as multi-year observational datasets accumulate. This

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

would not only help to further assess the general applicability of the proposed framework, but also potentially reduce the model's dependence on station density.

3. To address the grid residual artifacts identified in Section 3.4.1, future research could first quantify their error contribution to both the accuracy of the precipitation product and downstream applications such as hydrological modeling. A smoothing mechanism could then be incorporated either within the model architecture or as a post-processing step to mitigate these artifacts, while carefully controlling for any additional errors introduced by the smoothing process.

In summary, the two-stage framework proposed in this study follows a progressive division-of-labor strategy, and its generation of high–spatiotemporal-resolution precipitation products can provide data support for fine-scale applications such as urban flood early warning and hydrological modeling in small and medium-sized catchments.

Code and data availability

GPM IMERG final run precipitation data are available at https://gpm.nasa.gov/data/ (Huffman et al., 2023). Himawari-8 cloud property data are available at https://www.eorc.jaxa.jp/ptree/index.html. The all-weather cloud property data were generated using the CldNetV2 model; the model code and parameters are available at https://github.com/metdai/PMD-CldNetV2 (Nie et al., 2025). The TRIMS LST dataset is available at https://data.tpdc.ac.cn (Zhou et al., 2021). MODIS MOD13Q1 NDVI data are available at https://earthengine.google.com/ (Didan, 2021). The SRTM DEM is available at https://srtm.csi.cgiar.org/srtmdata/ (Jarvis et al., 2008). The hourly rain gauge observations are not publicly available; please direct any inquiries to the corresponding author. The code implementing the core functions of the CTDF framework, as well as the GMWWDA benchmark implementation, is available at https://doi.org/10.5281/zenodo.20118493 (He et al., 2026).

Author contributions

XW was responsible for conceptualization, supervision, and funding acquisition. WH developed the methodology and prepared the original draft. WH and JS developed the methodology and carried out formal analysis. WH and JL were responsible for data curation. WH and DW were responsible for validation and visualization. XW and JL reviewed and edited the article. ZZ provided guidance on methodology.

Competing interests

The contact author has declared that none of the authors has any competing interests.

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

Acknowledgements

We are grateful to the Youxian District Water Resources Bureau for providing the hourly rain gauge observations used in this study.

Financial support

This work was supported by the National Science Foundation for Young Scientists of China (52409100) and Development of an Intelligent Platform for River and Lake Health Identification, Diagnosis, and Early Warning Integrating DeepSeek Large Model (Y125008)

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 25 July 2019, Anchorage, Alaska, USA, https://doi.org/10.1145/3292500.3330701, 2019.

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, Bull. Am. Meteorol. Soc., 96, 69–83, https://doi.org/10.1175/BAMS-D-13-00068.1, 2015.

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T., Okuyama, A., Oyama, R., Sasaki, Y., Shimazu, Y., Shimoji, K., Sumida, Y., Suzuki, M., Taniguchi, H., Tsuchiyama, H., Uesawa, D., Yokota, H., and Yoshida, R.: An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites, J. Meteorol. Soc. JPN Ser. II, 94, 151–183, https://doi.org/10.2151/jmsj.2016-009, 2016.

Camps-Valls, G., Fernández-Torres, M.-Á., Cohrs, K.-H., Höhl, A., Castelletti, A., Pacal, A., Robin, C., Martinuzzi, F., Papoutsis, I., Prapas, I., Pérez-Aracil, J., Weigel, K., Gonzalez-Calabuig, M., Reichstein, M., Rabel, M., Giuliani, M., Mahecha, M. D., Popescu, O.-I., Pellicer-Valero, O. J., Ouala, S., Salcedo-Sanz, S., Sippel, S., Kondylatos, S., Happé, T., and Williams, T.: Artificial intelligence for modeling and understanding extreme weather and climate events, Nat. Commun., 16, 1919, https://doi.org/10.1038/s41467-025-56573-8, 2025.

Chen, C., Hu, B., and Li, Y.: Easy-to-use spatial random-forest-based downscaling-calibration method for producing precipitation data with high resolution and high accuracy, Hydrol. Earth Syst. Sci., 25, 5667–5682, https://doi.org/10.5194/hess-25-5667-2021, 2021.

Chen, H., Chandrasekar, V., Cifelli, R., and Xie, P.: A machine learning system for precipitation estimation using satellite and ground radar network observations, IEEE Trans. Geosci. Remote Sens., 58, 982–994, https://doi.org/10.1109/TGRS.2019.2942280, 2020.

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

Chen, H., Yong, B., Kirstetter, P.-E., Wang, L., and Hong, Y.: Global component analysis of errors in three satellite-only global precipitation estimates, Hydrol. Earth Syst. Sci., 25, 3087–3104, https://doi.org/10.5194/hess-25-3087-2021, 2021.

Chen, H., Wen, D., Du, Y., Xiong, L., and Wang, L.: Errors of five satellite precipitation products for different rainfall intensities, Atmos. Res., 285, 106622, https://doi.org/10.1016/j.atmosres.2023.106622, 2023.

Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, 13 August 2016, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Cristiano, E., ten Veldhuis, M.-C., and van de Giesen, N.: Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review, Hydrol. Earth Syst. Sci., 21, 3859–3878, https://doi.org/10.5194/hess-21-3859-2017, 2017.

Didan, K., Munoz, A. B., Solano, R., and Huete, A.: MODIS vegetation index user's guide (MOD13 series), Tech. rep., Vegetation Index and PhenologyLab, https://doi.org/10.5067/MODIS/MYD13C2.006, 2015.

Didan, K.: MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061, NASA Land Processes Distributed Active Archive Center [data set], https://doi.org/10.5067/MODIS/MOD13Q1.061, 2021.

Duan, Z. and Bastiaanssen, W. G. M.: First results from version 7 TRMM $3B_{43}$ precipitation product in combination with a new downscaling–calibration procedure, Remote Sens. Environ., 131, 1–13, https://doi.org/10.1016/j.rse.2012.12.002, 2013.

Feldman, A. F., Konings, A. G., Gentine, P., Cattry, M., Wang, L., Smith, W. K., Biederman, J. A., Chatterjee, A., Joiner, J., and Poulter, B.: Large global-scale vegetation sensitivity to daily rainfall variability, Nature, 636, 380–384, https://doi.org/10.1038/s41586-024-08232-z, 2024.

Gong, Y., Chen, Y., Jiang, Q., and Zhou, Y.: Enhanced error correction and spatial downscaling of precipitation and air temperature in the middle and low reaches of the Yangtze River using a random forest model with the sokol method, Earth Sci. Inf., 18, 433–446, https://doi.org/10.1007/s12145-025-01933-7, 2025.

He, W., Li, J., Wen, D., Su, J., Zhang, Z., and Wang, X.: Cross-temporal downscaling and fusion for hourly 0.01° precipitation estimation: A case study in Youxian District, China (v2.0.0), Zenodo [code], https://doi.org/10.5281/zenodo.20118493, 2026.

He, K., Zhao, W., Brocca, L., Quintana-Seguí, P., and Chen, X.: SMPD-MERG: A hybrid downscaling model for high-resolution daily precipitation estimation via merging surface soil moisture and multisource precipitation data, IEEE Trans. Geosci. Remote Sens., 63, 1–16, https://doi.org/10.1109/TGRS.2025.3561253, 2025.

Hinge, G., Hamouda, M. A., Long, D., and Mohamed, M. M.: Hydrologic utility of satellite precipitation products in flood prediction: a meta-data analysis and lessons learnt, J. Hydrol., 612, 128103, https://doi.org/10.1016/j.jhydrol.2022.128103, 2022.

Houze, R. A.: Stratiform precipitation in regions of convection: a meteorological paradox?, Bull. Am. Meteorol. Soc., 78, 2179–2196, https://doi.org/10.1175/1520-0477(1997)078%3C2179:SPIROC%3E2.0.CO;2, 1997.

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., and Yoo, S. H.: NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG), in: Algorithm Theoretical Basis Document (ATBD), Number 4.4, NASA/GSFC, Greenbelt, MD, USA, 30 pp., 2014.

Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., and Tan, J.: GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V07, Goddard Earth Sciences Data and Information Services Center (GES DISC) [data set], https://doi.org/10.5067/GPM/IMERG/3B-HH/07, 2023.

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled SRTM for the globe Version 4, CGIAR-CSI [data set], http://srtm.csi.cgiar.org, 2008.

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., and Kirschbaum, D. B.: So, how much of the Earth's surface is covered by rain gauges?, Bull. Am. Meteorol. Soc., 98, 69–78, https://doi.org/10.1175/BAMS-D-14-00283.1, 2017.

Kofidou, M., Stathopoulos, S., and Gemitzi, A.: Review on spatial downscaling of satellite derived precipitation estimates, Environ. Earth Sci., 82, 424–456, https://doi.org/10.1007/s12665-023-11115-7, 2023.

Lasser, M., O, S., and Foelsche, U.: Evaluation of GPM-DPR precipitation estimates with WegenerNet gauge data, Atmos. Meas. Tech., 12, 5055–5070, https://doi.org/10.5194/amt-12-5055-2019, 2019.

Li, M. and Shao, Q.: An improved statistical approach to merge satellite rainfall estimates and raingauge data, J. Hydrol., 385, 51–64, https://doi.org/10.1016/j.jhydrol.2010.01.023, 2010.

Ma, Z., Xu, J., He, K., Han, X., Ji, Q., Wang, T., Xiong, W., and Hong, Y.: An updated moving window algorithm for hourly-scale satellite precipitation downscaling: a case study in the southeast coast of China, J. Hydrol., 581, 124378, https://doi.org/10.1016/j.jhydrol.2019.124378, 2020.

Ma, Z., Xu, J., Ma, Y., Zhu, S., He, K., Zhang, S., Ma, W., and Xu, X.: AERA5-Asia: a long-term asian precipitation dataset (0.1°, 1-hourly, 1951–2015, Asia) anchoring the ERA5-land under the total volume control by APHRODITE, Bull. Am. Meteorol. Soc., 103, E1146–E1171, https://doi.org/10.1175/BAMS-D-20-0328.1, 2022.

Nie, L., Chen, Y., and Zhang, D.: All-day cloud property and occurrence probability dataset based on satellite remote sensing data, Sci. Data, 12, 387–408, https://doi.org/10.1038/s41597-025-04659-9, 2025.

Rosenfeld, D. and Lensky, I. M.: Satellite-based insights into precipitation formation processes in continental and maritime convective clouds, Bull. Am. Meteorol. Soc., 79, 2457–2476, https://doi.org/10.1175/1520-0477(1998)079%3C2457:SBIIPF%3E2.0.CO;2, 1998.

Rossow, W. B. and Schiffer, R. A.: ISCCP cloud data products, Bull. Am. Meteorol. Soc., 72, 2–20, https://doi.org/10.1175/1520-0477(1991)072%3C0002:ICDP%3E2.0.CO;2, 1991.

Saharia, M., Kirstetter, P.-E., Vergara, H., Gourley, J. J., Emmanuel, I., and Andrieu, H.: On the impact of rainfall spatial variability, geomorphology, and climatology on flash floods, Water Resour. Res., 57, e2020WR029124, https://doi.org/10.1029/2020WR029124, 2021.

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

Sokol, Z., Szturc, J., Orellana-Alvear, J., Popová, J., Jurczyk, A., and Célleri, R.: The role of weather radar in rainfall estimation and its application in meteorological and hydrological modelling—a review, Remote Sens., 13, 351–388, https://doi.org/10.3390/rs13030351, 2021.

Standardization Administration of China: Grade of precipitation, GB/T 28592–2012, https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=B4A00E4ABCF80F8C6A048C1D0121A97D&refer=outter (last access: 5 May 2026), 2012.

Tang, W., Zhou, J., Ma, J., Wang, Z., Ding, L., Zhang, X., and Zhang, X.: TRIMS LST: a daily 1 km all-weather land surface temperature dataset for China's landmass and surrounding areas (2000–2022), Earth Syst. Sci. Data, 16, 387–419, https://doi.org/10.5194/essd-16-387-2024, 2024.

Tesfa, T. K., Leung, L. R., and Ghan, S. J.: Exploring topography-based methods for downscaling subgrid precipitation for use in Earth system models, J. Geophys. Res.: Atmos., 125, e2019JD031456, https://doi.org/10.1029/2019JD031456, 2020.

Trenberth, K. E. and Shea, D. J.: Relationships between precipitation and surface temperature, Geophys. Res. Lett., 32, L14703, https://doi.org/10.1029/2005GL022760, 2005.

Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J.: Estimates of the global water budget and its annual cycle using observational and model data, J. Hydrometeorol., 8, 758–769, https://doi.org/10.1175/JHM600.1, 2007.

Wang, Y., Miao, C., Zhao, X., Zhang, Q., and Su, J.: Evaluation of the GPM IMERG product at the hourly timescale over China, Atmos. Res., 285, 106656, https://doi.org/10.1016/j.atmosres.2023.106656, 2023.

Woods, D., Kirstetter, P.-E., Vergara, H., Duarte, J. A., and Basara, J.: Hydrologic evaluation of the global precipitation measurement mission over the U.S.: flood peak discharge and duration, J. Hydrol., 617, 129124, https://doi.org/10.1016/j.jhydrol.2023.129124, 2023.

Wu, H., Yong, B., and Shen, Z.: Hourly calibration algorithm for Fengyun-2G quantitative precipitation estimates using spatial random forest and improved temporal disaggregation scheme, Atmos. Res., 321, 108061, https://doi.org/10.1016/j.atmosres.2025.108061, 2025.

Yang, L., Meng, X., and Zhang, X.: SRTM DEM and its application advances, Int. J. Remote Sens., 32, 3875–3896, https://doi.org/10.1080/01431161003786016, 2011.

Yin, Y., He, J., Guo, J., Song, W., Zheng, H., and Dan, J.: Enhancing precipitation estimation accuracy: an evaluation of traditional and machine learning approaches in rainfall predictions, J. Atmos. Sol. Terr. Phys., 255, 106175, https://doi.org/10.1016/j.jastp.2024.106175, 2024.

Zhan, C., Han, J., Hu, S., Liu, L., and Dong, Y.: Spatial downscaling of GPM annual and monthly precipitation using regression-based algorithms in a mountainous area, Adv. Meteorol., 2018, 1506017, https://doi.org/10.1155/2018/1506017, 2018.

Zhang, J., Xu, J., Dai, X., Ruan, H., Liu, X., and Jing, W.: Multi-source precipitation data merging for heavy rainfall events based on cokriging and machine learning methods, Remote Sens., 14, 1750, https://doi.org/10.3390/rs14071750, 2022.

https://doi.org/10.5194/egusphere-2026-1468
Preprint. Discussion started: 18 May 2026

Zhang, X., Song, Y., Nam, W.-H., Huang, T., Gu, X., Zeng, J., Huang, S., Chen, N., Yan, Z., and Niyogi, D.: Data fusion of satellite imagery and downscaling for generating highly fine-scale precipitation, J. Hydrol., 631, 130665, https://doi.org/10.1016/j.jhydrol.2024.130665, 2024.

Zhao, C., Yang, Y., Chi, Y., Sun, Y., Zhao, X., Letu, H., and Xia, Y.: Recent progress in cloud physics and associated radiative effects in China from 2016 to 2022, Atmos. Res., 293, 106899, https://doi.org/10.1016/j.atmosres.2023.106899, 2023.

Zhao, N.: A method for merging multi-source daily satellite precipitation datasets and gauge observations over Poyang Lake Basin, China, Remote Sens., 15, 2407, https://doi.org/10.3390/rs15092407, 2023.

Zhou, J., Zhang, X., Tang, W., Ding, L., Ma, J., and Zhang, X.: Daily 1-km all-weather land surface temperature dataset for the China's landmass and its surrounding areas (TRIMS LST; 2000–2024), National Tibetan Plateau / Third Pole Environment Data Center [data set], https://doi.org/10.11888/Meteoro.tpdc.271252, 2021.

Zhu, H., Liu, H., Zhou, Q., and Cui, A.: A XGBoost-based downscaling-calibration scheme for extreme precipitation events, IEEE Trans. Geosci. Remote Sens., 61, 1–12, https://doi.org/10.1109/TGRS.2023.3294266, 2023.

Zhu, J., Yang, Y., Meng, N., Li, R., Ma, J., and Zheng, H.: Constructing a high-precision precipitation dataset on the data-limited qinghai-tibet plateau, Agric. For. Meteorol., 344, 109794, https://doi.org/10.1016/j.agrformet.2023.109794, 2024.

Zhu, S., Ma, Z., Xu, J., He, K., Liu, H., Ji, Q., Tang, G., Hu, H., and Gao, H.: A morphology-based adaptively spatio-temporal merging algorithm for optimally combining multisource gridded precipitation products with various resolutions, IEEE Trans. Geosci. Remote Sens., 60, 1–21, https://doi.org/10.1109/TGRS.2021.3097336, 2022.

Zhuang, Q., Zhou, Z., Liu, S., Wright, D. B., and Gao, L.: The evaluation and downscaling-calibration of IMERG precipitation products at sub-daily scales over a metropolitan region, J. Flood Risk Manage., 16, e12902, https://doi.org/10.1111/jfr3.12902, 2023.