J. Geogr. Sci. 2022, 32(11): 2271-2290
DOI: https://doi.org/10.1007/s11442-022-2047-9

Spatio-temporal accuracy evaluation of MSWEP daily precipitation over the Huaihe River Basin, China:

A comparison study with representative satellite- and reanalysis-based products

LI Lingjie[1,2], WANG Yintang[1,2], *WANG Leizhi[1], HU Qingfang[1,2], ZHU Zhenduo[3], LI Liping[4], LI Chengxi[5]

1. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering Science, Nanjing Hydraulic Research Institute, Nanjing 210029, China;
2. Yangtze Institute for Conservation and Development, Nanjing 210098, China;
3. Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14220, USA;
4. Lvliang Meteorology Bureau of Shanxi Province, Lishi 033000, Shanxi, China;
5. La Salle College Preparatory, Pasadena, CA 91107, USA

Abstract: Multi-source weighted-ensemble precipitation (MSWEP) is one of the most popular merged global precipitation products with long-term spanning and high spatial resolution. While various studies have acknowledged its ability to accurately estimate precipitation in terms of temporal dynamics, its performance regarding spatial pattern and extreme rainfall is overlooked. To fill this knowledge gap, the daily precipitation of two versions of MSWEP (MSWEP V2.1 & V2.2) are compared with that of three representative satellite- and reanalysis-based products, namely the Tropical Rainfall Measuring Mission (TRMM 3B42 V7), the climate prediction center morphing technique satellite-gauge merged product (CMORPH BLD), and the fifth-generation reanalysis product of the European Centre for Medium Range Weather Forecasts (ERA5). The comparison is made according to the dense daily rainfall observations from 539 rain gauges over the Huaihe River Basin in China during 2006–2015. The results show that MSWEP V2.1, MSWEP V2.2 and CMORPH BLD have better performance on temporal accuracy of precipitation estimation, followed by ERA5 and TRMM 3B42 V7. MSWEPs yield the most even spatial distribution across the basin since it takes full advantage of the multi datasets. As the weighted-ensemble method is independently carried out on each grid in MSWEPs, the spatial distribution of local precipitation is changed by different source data, which results in that MSWEPs perform worse than CMORPH BLD in terms of the representation of precipitation spatial pattern. In addition, the capability of MSWEPs to de-

Received: 2021-07-31 **Accepted:** 2021-12-27
Foundation: National Key R&D Program of China, No.2021YFC3000104; National Natural Science Foundation of China, No.52009081, No.51479118; Special Funded Project for Basic Scientific Research Operation Expenses of the Central Public Welfare Scientific Research Institutes of China, No.Y519006
Author: Li Lingjie (1992–), Engineer and PhD Candidate, specialized in hydrology and water resources. E-mail: ljli@nhri.cn
***Corresponding author:** Wang Leizhi (1991–), PhD and Senior Engineer, specialized in hydrology. E-mail: wanglz@nhri.cn

scribe the spatial structure in the rainy season is lower than that in the dry season. Strong precipitation (⩾100 mm/d) events are better represented in TRMM 3B42 V7 products than in MSWEPs. Finally, based on the comparison results, we suggest to improve the merging algorithm of MSWEP by considering the precipitation spatial self-correlation and adjusting the merging weights based on the performance of the source datasets under different precipitation intensities.

Keywords: MSWEP; temporal accuracy; spatial pattern; extreme precipitation; Huaihe River Basin

1 Introduction

A better understanding on spatio-temporal distribution of precipitation is important for the researches on climate change, ecological conservation, utilization of water and land resources, and the prevention and mitigation of hydro-meteorological and geological disasters (Tapiador *et al*., 2012; Kirschbaum *et al*., 2017; Monsieurs *et al*., 2018). Precipitation shows complex spatio-temporal variations due to the influences of topography, climate and human activities. Therefore, it is one of the meteorological variables which is difficult to be measured, estimated and forecasted. Though rain gauges can provide reliable and accurate observation, the number of gauges is insufficient, especially in the areas with lakes, mountains and deserts (Dos Reis *et al*., 2017; Sun *et al*., 2018). As a result, it is not possible to get reliable and continuous spatio-temporal distribution of precipitation over a big domain based on rain gauge observations without any other data sources.

In recent decades, a series of satellite- and reanalysis-based gridded precipitation estimates (GPEs) have been developed with advanced sensors and quantitative retrieval algorithms. Major representative products were used in scientific community including the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA), the climate prediction center morphing technique (CMORPH), the precipitation estimation from remotely sensed information using artificial neural networks, the global precipitation measurement integrated multi-satellite retrievals for global precipitation measurement, the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis of global climate (ERA5), and the Japanese 55-year reanalysis (JRA-55) (Joyce *et al*., 2004; Huffman *et al*., 2007, 2010; Ebita *et al*., 2011; Hou *et al*., 2014; Ashouri *et al*., 2015; Hersbach *et al*., 2020). These products have been widely used in meteorological and hydrological studies at various scales (Su *et al*., 2019; Chen *et al*., 2020; Jiang *et al*., 2020; Palash *et al*., 2020; Tarek *et al*., 2020; Beck *et al*., 2021), and the products can be considered as alternatives to gauge observations in some areas. Several studies have revealed that the use of GPEs is limited by its coarse spatio-temporal resolutions, though it has certain advantages in characterizing the spatial distribution of precipitation. At the same time, the data bias is affected by many factors, such as climate conditions, topography, geographic locations, spatio-temporal scales and precipitation intensity (Maggioni *et al*., 2016; Sun *et al*., 2018; Li *et al*., 2020; Ma *et al*., 2020; Tang *et al*., 2020).

It is clear that both gauge observations and GPE products have their own strengths. Hence, integrating these products from various sources can enable a better balance between accurately estimating precipitation and characterizing its spatial distribution. Multi-source weighted-ensemble precipitation (MSWEP), developed by Beck *et al*. (2017a; 2019b), is one of the most popular multi-source products. MSWEP is a global dataset (including ocean ar-

eas) with comparatively high spatio-temporal resolution (0.1°×0.1°, 3 h) and long temporal coverage (spanning from 1979 to 2021) by merging weighted ground observations, satellite-derived and reanalysis-based precipitation estimations (Beck *et al*., 2019b). The comparisons between MSWEP and other GPE products were conducted in a wide range of spatial scales, from regional to global (Beck *et al*., 2017b; Massari *et al*., 2017; Sun *et al*., 2018; Beck *et al*., 2019a; Xu *et al*., 2019). Products of MSWEP-ng V2.0 (a non-gauge version) were assessed at global scale by compared with 12 ensemble-, satellite-, and reanalysis-based non-gauge-corrected products (Beck *et al*., 2017b). The results showed that MSWEP-ng V2.0 has the highest temporal correlation with gauge observations, followed by reanalysis- and satellite-based precipitation products. Moreover, MSWEP-ng V2.0 showed a better performance on long-term averaged precipitation estimations. Evaluation of eleven gauge-corrected precipitation datasets in the conterminous United States (CONUS) demonstrated that MSWEP V2.2 generally characterizes daily precipitation most satisfactorily with the Kling–Gupta efficiency approaching 1 for all grids (Beck *et al*., 2019a). Over China, MSWEP V2.1 has capabilities of capturing the temporal variations of daily precipitation, and effectively identifying and classifying precipitation events (Xu *et al*., 2019). But, MSWEP underestimates the precipitation over the monsoon regions in China and Australia, while there are overestimations in the Qinghai-Tibet Plateau, China (Awange *et al*., 2019; Liu *et al*., 2019; Xu *et al*., 2019). Furthermore, the averaged accuracy of MSWEP V2.1 in India is lower than that of the Climate Hazards Group Infrared Precipitation dataset (Prakash, 2019).

The MSWEP precipitation time series at rain gauge stations or grids were examined by many studies to assess the averaged accuracy and spatial distributions of relevant indicators. However, it is still not clear that how the spatial accuracy in MSWEP is, as well as the spatial structure. To obtain a better knowledge of the performance of MSWEP, it is critical to perform detailed assessments in different regions with a comprehensive comparison between MSWEP and other representative GPE products. Additionally, analyzing the causes of the limitations of MSWEP from the perspectives of the fusion algorithm and source datasets could lead to further improvements in the algorithm.

The Huaihe River, located in the climate transition zone, suffers lots of floods. After enormous efforts were done in more than 70 years, the capabilities of flooding defense and projection over the Huaihe River Basin (HRB) have been effectively strengthened. However, flood modeling and forecasting for the HRB still need to be improved. In particular, more reliable GPEs with higher spatio-temporal resolutions are urgently needed to improve the accuracy of flood prediction. To this end, this study focuses on the comprehensive comparison of the performances of five GPEs over the Bengbu Basin (BB) located in the upper reaches of the HRB, namely MSWEPs (V2.2 and V2.1), two satellite-based precipitation products (TRMM 3B42 V7 and CMORPH BLD), and the most updated reanalysis-based precipitation product (ERA5). Daily precipitation data measured at 539 rain gauge stations in the BB during 2006–2015 are adopted to construct a 0.25°×0.25°-grid benchmark precipitation dataset, by using a state-of-the-art spatial estimation method based on the double geographically weighted regression. These products are then compared from the perspectives of overall performance, temporal accuracy, spatial accuracy, and estimation performance for precipitations of different intensities, so as to evaluate their abilities in identifying and esti-

mating daily precipitation in the BB. In addition, the positive and negative aspects of MSWEPs compared to satellite- and reanalysis-based precipitations are presented, especially in terms of spatial accuracy and its sensitivities to precipitation intensities. Finally, the possible causes for the differences of performances between MSWEPs and other data products are discussed, and suggestions for the improvement of the multi-source weighted-merging method are proposed.

2 Study area and data

2.1 Study area

The HRB is located between the Yangtze River and the Yellow River, covering an area of approximately 2.7×10^5 km^2. The majority of the HRB is vast plain, with the Tongbai Mountain and the Funiu Mountain in its northwest, the Dabie Mountain in its southwest, and the Yimeng Mountain in its northeast. The Bengbu hydrological station is the main control node in the middle reach of the Huaihe River. The study area covers the upper catchment of Bengbu Station over the HRB (Figure 1a). The area of the catchment (111.9°–117.5°E, 30.9°–34.9°N) is about 1.17×10^5 km^2, of which the mountains occupy only 15.4%. The southeastern part of the study region contains many lakes (Figure 1b). Importantly, rain gauge stations with a relatively long observation period are densely distributed in the area (Figure 1c). Thereby, plenty observations can be used to evaluate the performance of GPEs effectively and accurately. The annual mean precipitation in the BB is approximately 920 mm, and it gradually decreases from south to north. Precipitation in the rainy season (June to September) accounts for 50%–75% of the total annual-precipitation (Zhang *et al*., 2020).

2.2 Data

2.2.1 Rain gauge observations

Daily precipitation observations from 539 gauge stations (Figure 1c) during the period from 2006 to 2015 are collected from the Annual Hydrological Report of the HRB. On average, the effective coverage of a rain gauge is 217 km^2, and the density of gauges over the mountainous region is reasonably high.

2.2.2 GPEs

MSWEP is a global high-resolution (0.1°×0.1°, 3 h) precipitation dataset encompassing the period from 1979 to 2021. MSWEP is produced by integrating three types of datasets using the weighted ensemble algorithm. The sources of datasets include four uncorrected satellite-based precipitation products (TMPA 3B42RT, CMORPH, global satellite mapping of precipitation and gridded satellite (GridSat), Global Satellite Mapping of Precipitation (GSMaP)), two reanalysis-based precipitation products (ECMWF interim reanalysis and JRA-55), and the observations from over 75,000 rain gauges worldwide. MSWEP has been optimized and updated several times through the improvement of algorithm and data sources. The merging process of MSWEP consists of three main parts: the quality control and validation of source data, the weight calculation and integration of satellite- and reanalysis-based precipitation, and the correction of the merged data by comparison with gauge observations. The whole process is performed in 10 specific steps (refer to Beck *et al*. (2019b) for more details). The MSWEP V2.1 and V2.2 are selected to be evaluated in this study. Compared to

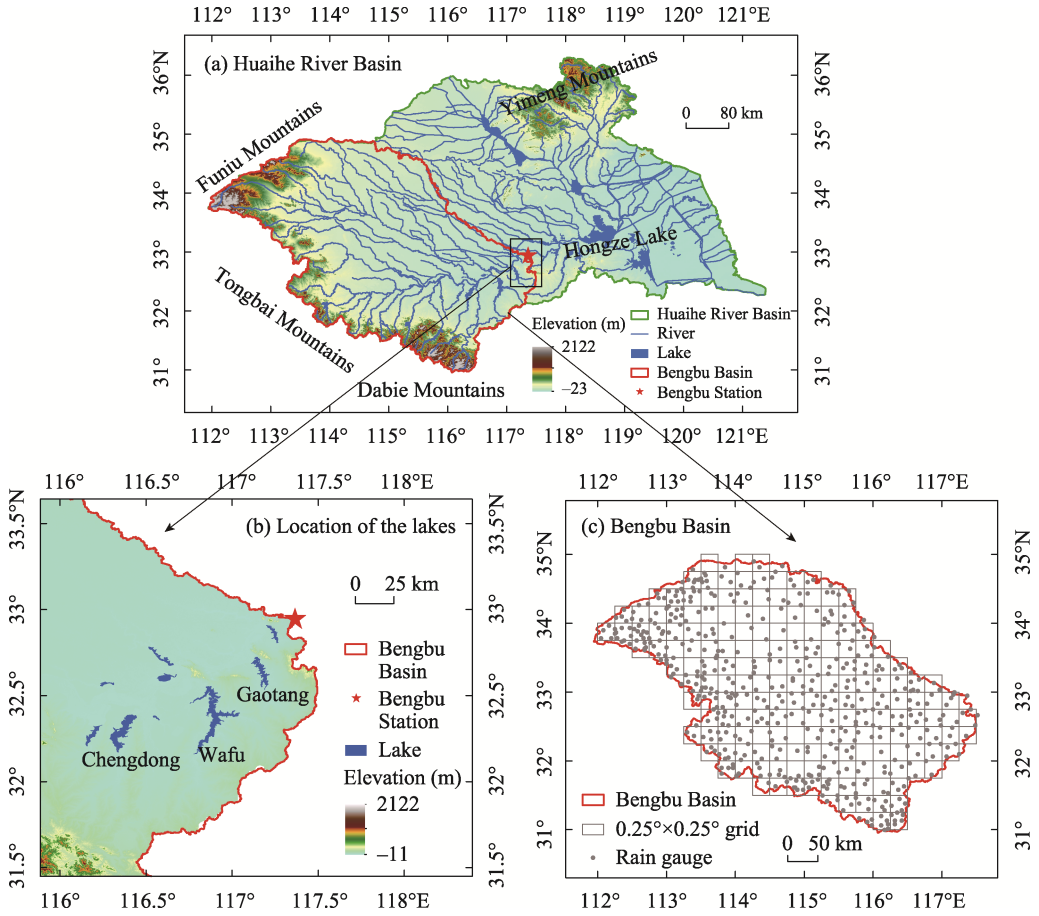

Figure 1 Geographical locations of (a) the Bengbu Basin, (b) lakes and (c) rain gauge stations

MSWEP V2.1, V2.2 has the following updates. (1) The quality control of precipitation data retrieved from GridSat infrared images is included. The incorporated GridSat-based rainfall estimates have been expanded from 1980 rather than 1983. (2) The problem of intensity and trend amplification caused by the cumulative distribution function (CDF) is addressed. The operation of rescaling the merged CDF-corrected estimates has been added to match the merged non-CDF-corrected estimates. The MSWEP V2.1 and V2.2 from 2006 to 2015 are downloaded from *www.gloh2o.org*.

The representative satellite-based precipitation products used in this study are TRMM 3B42 V7 and CMORPH BLD. The former is retrieved by using a multi-satellite estimation algorithm that integrates infrared and microwave data from TRMM and other satellites. The estimates are rescaled by the monthly gauge observations (Huffman *et al.*, 2007, 2010). The retrieval procedure of CMORPH RAW is that the movement vector of cloud systems is calculated first based on the high-resolution infrared brightness temperature data observed by geostationary satellites. Then, the instantaneous precipitation distribution which is based on the passive microwave data from low-earth orbit satellites is extrapolated to the target time along the movement vector (Joyce *et al.*, 2004). The probability density function matching is applied to CMORPH RAW to generate the CMORPH CRT product with less bias. The CMORPH BLD is obtained by using the optimal interpolation to merge CMORPH CRT with

gauge observations. TRMM 3B42 V7 and CMORPH BLD are both scientific products with a spatio-temporal resolution of 0.25°×0.25° and 3h. The reanalysis-based precipitation product selected in this study is the latest ERA5 which is based on the ECMWF integrated forecasting system (IFS) Cy41r2. ECMWF IFS employs a hybrid data assimilation system to get ten-member ensemble 4D-VAR data assimilation (Hersbach *et al*., 2020). The data are presented at 0.25°×0.25° grids and 1-hour intervals. The above three datasets in 2006–2015 are used to compare the accuracy of two MSWEP precipitation estimations.

3 Methodology

The framework of the research method includes three parts (Figure 2). First, data processing is conducted to obtain benchmark data and five GPEs with harmonious spatio-temporal resolutions. Second, detectability and error indicators are employed to compare two MSWEP versions and other GPEs. Then, the strengths and weaknesses of MSWEP are identified and the possible causes are discussed. Finally, advices on the improvement of the multi-source precipitation information fusion algorithm are proposed based on the results.

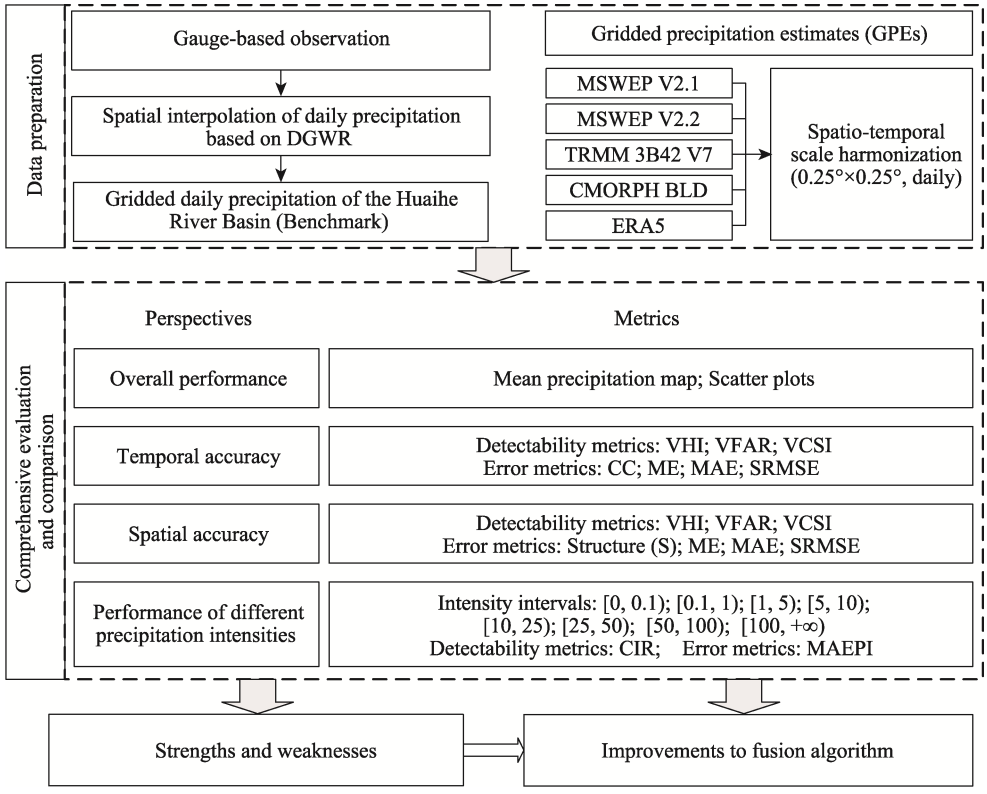

Figure 2 The framework of the research method

3.1 Metrics of accuracy evaluation

This study evaluates the performance of the five GPEs based on the dense rain gauge observations from 2006 to 2015 in the BB. All the 539 gauge observations are not integrated in

the GPEs. The evaluation of GPEs includes the overall performance, temporal accuracy, spatial accuracy, and estimation abilities under different precipitation intensities.

Firstly, overall performance is evaluated by scatter plots and the spatial distribution maps of daily mean precipitation derived from gridded precipitation datasets. Secondly, temporal accuracy is calculated by comparing the time series of the five GPEs with that of benchmark at all grids. The evaluation indexes include categorical and quantitative metrics. The commonly used categorical indexes are the probability of detection (POD), false alarm ratio (FAR) and critical success index (CSI). These indexes can measure the capability of detecting precipitation events, but they cannot well evaluate the accuracy of precipitation estimation. Even if the values of these indexes are the same, the estimated precipitation may vary considerably. As a result, in this study, volumetric indexes proposed by AghaKouchak and Mehran (2013), namely the volumetric hit index (VHI, Equation [1]), volumetric false alarm ratio (VFAR, Equation [2]) and volumetric critical success index (VCSI, Equation [3]), are used to assess the performance of the products. The values of these indexes range between 0 and 1. The larger VHI and VCSI, and the smaller VFAR indicate better performance. The quantitative metrics include mean error (ME, and it equals benchmark minus GPE), mean absolute error (MAE), standardized root-mean-square error (SRMSE) and correlation coefficient (CC). The smaller ME, MAE and SRMSE, as well as the larger CC, indicate higher estimation accuracy of GPEs. Thirdly, spatial accuracy is evaluated by comparing the spatial pattern of GPEs with that of benchmark precipitation in corresponding grids in a specified period. The above-mentioned indexes could be used to evaluate spatial accuracy. While the similarity of local spatial patterns cannot be quantitatively analyzed using CC, because it does not consider the spatial correlation of precipitation. Furthermore, GPEs and gridded benchmark precipitation during a specified period can be treated as images. Then, the structural similarity index (Robertson *et al*., 2014) is used, which considers luminance, contrast and structure (S) of an image. It could provide the local differences of the mean, variance and correlation. In this study, only S is adopted as the evaluation index (Equation [4]) (Plouffe *et al*., 2015). As the S gets closer to 1, the spatial structures between the GPEs and benchmark become more similar. Finally, eight precipitation intensity intervals, separated by 0.1, 1, 5, 10, 25, 50 and 100 mm/d, are considered. The correct identification ratios (CIR, Equation [5]) of the intervals are used to assess the identification capability of the GPEs for precipitation events with different intensities. The MAEs in each interval are calculated, which are written as MAE_{PI}s (Equation [6]) in order to distinguish it from those of the spatio-temporal accuracies.

$$\mathrm{VHI}=\frac{\sum_{i=1}^{N}\left(GPE_i \mid GPE_i>T \,\&\, G_i>T\right)}{\sum_{i=1}^{N}\left(GPE_i \mid GPE_i>T \,\&\, G_i>T\right)+\sum_{i=1}^{N}\left(G_i \mid GPE_i \leqslant T \,\&\, G_i>T\right)} \tag{1}$$

$$\mathrm{VFAR}=\frac{\sum_{i=1}^{N}\left(GPE_i \mid GPE_i>T \,\&\, G_i \leqslant T\right)}{\sum_{i=1}^{N}\left(GPE_i \mid GPE_i>T \,\&\, G_i>T\right)+\sum_{i=1}^{N}\left(GPE_i \mid GPE_i>T \,\&\, G_i \leqslant T\right)} \tag{2}$$

$$\mathrm{VCSI}=\frac{\sum_{i=1}^{N}\left(GPE_i \mid GPE_i>T \,\&\, G_i>T\right)}{\sum_{i=1}^{N}\left[GPE_i \mid \left(GPE_i>T \,\&\, G_i>T\right)+G_i \mid \left(GPE_i \leqslant T \,\&\, G_i>T\right)+GPE_i \mid \left(GPE_i>T \,\&\, G_i \leqslant T\right)\right]} \tag{3}$$

$$S(GPE,G)=\frac{\sigma_{GPE,G}+\mathrm{c}}{\sigma_{GPE}\sigma_{G}+\mathrm{c}} \tag{4}$$

$$\mathrm{CIR}=\frac{\sum_{i=1}^{N}\left(1\,|\,GPE_i\in[T_{\mathrm{down}},T_{\mathrm{up}})\ \&\ G_i\in[T_{\mathrm{down}},T_{\mathrm{up}})\right)}{\sum_{i=1}^{N}\left(1\,|\,GPE_i\in\forall R^{+}\cup\{0\}\ \&\ G_i\in[T_{\mathrm{down}},T_{\mathrm{up}})\right)} \tag{5}$$

$$\mathrm{MAE}_{\mathrm{PI}}=\frac{1}{n}\sum_{j=1}^{n}|\,GPE_i-G_i\,|,\quad \text{where } GPE_i\in\forall R^{+}\cup\{0\}\ \&\ G_i\in[T_{\mathrm{down}},T_{\mathrm{up}}) \tag{6}$$

For temporal accuracy, in Equations (1)–(3), GPE_i and G_i are the precipitation of the GPE and benchmark at a certain grid in the period i, respectively, and N denotes the number of temporal periods. For spatial accuracy, in Equations (1)–(3), GPE_i and G_i represent the precipitation of the GPE and benchmark at the grid i in a certain period, respectively, and N is the number of grids. In Equations (1)–(3), T is the threshold for binary classification of precipitation events, and T = 0.1 mm/d. In Equation (4), σ_{GPE} and σ_G denote the standard deviations of the GPE and benchmark in the 3 × 3 sliding window at a certain period, respectively, and $\sigma_{\mathrm{GPE},\,G}$ is the covariance of the GPE and benchmark. In order to stabilize the equation, c is a constant used when the variability of mean or variance is close to 0 (e.g., when there are large consistent patches) (Plouffe *et al*., 2015), and its calculation formula can be obtained from Robertson *et al*. (2014). In Equations (5) and (6), T_{down} and T_{up} are the lower and upper boundaries of the precipitation intensity intervals, and $\forall R^{+}\cup\{0\}$ denotes an arbitrary non-negative real number.

3.2 Data processing

3.2.1 Preparation of benchmark precipitation

When discrete gauge observations are directly used to evaluate gridded precipitation products, the precipitation in the areas without gauges cannot be validated. Hence, it is of vital importance to prepare a reliable high-resolution gridded benchmark precipitation dataset. GWR (geographically weighted regression), a local multiple linear regression method, can efficiently utilize the spatial autocorrelation of the regression variable and its cross-correlations with explanatory factors (Brunsdon *et al*., 1998; Chao *et al*., 2018). As a result, GWR has an advantage in presenting non-stationary spatial variations. Short-duration precipitations, such as daily precipitation, are highly discontinuous. Based on the work by Barancourt *et al*. (1992), the double GWR (DGWR) is proposed. First, the precipitation probabilities at each grid are estimated by using GWR according to the binary classification results (rain/no-rain) of gauge observations. Then, the precipitation probability threshold is optimized by pursuing the highest CIR to identify the status (rain/no-rain) of each grid. Here, the CIR denotes the percentage of correct results in identifying rain/no-rain status of gauge stations. Next, according to the gauge observations, precipitation is estimated for each grid, and then the estimates and status of precipitation are multiplied to calculate the gridded benchmark precipitation. To validate the DGWR, k-means clustering is employed to randomly divide the 539 gauge stations into four categories. The spatial distributions of the four categories of gauge networks for modeling and validation are illustrated in Figure 3. Cross-validation is used to evaluate the estimation accuracies (Table 1). DGWR accurately identifies the precipitation status of gauge observation, and the CIR of precipitation status at

the gauge stations used for validation is basically equivalent to the modeling gauge network. In addition, if the quantitative errors of the estimates are smaller, the average accuracies of different networks are closer to each other. In conclusion, DGWR can provide a reasonably generalization and a relatively high accuracy. The China Gauge-based Daily Precipitation Analysis (CGDPA), a gauge-based precipitation produced by the China Meteorological Administration, is often employed as benchmark dataset. According to Shen and Xiong (2015), ME of the CGDPA is between 0.6 mm/d and 3.0 mm/d over the study region and the maximum value of CC is 0.557. Compared to the CGDPA, the benchmark gridded precipitation produced by utilizing GWR performs better. So it is a feasible method for the spatial estimation of daily precipitation. To obtain more accurate benchmark precipitation, all gauge information including precipitation observation, latitudes, longitudes, and altitudes are employed to construct the reference dataset with a spatial resolution of 0.25°×0.25°.

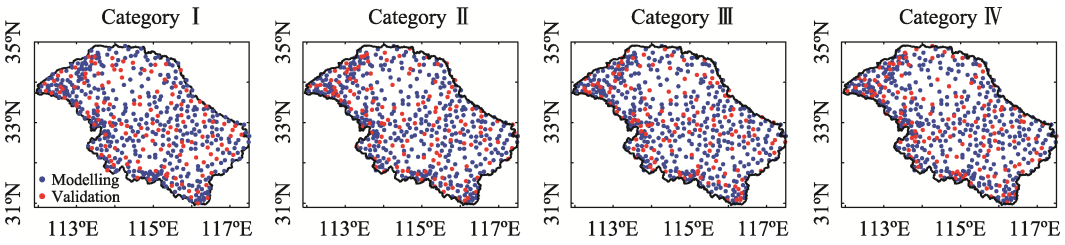

Figure 3 Spatial distributions of gauge station networks in the Bengbu Basin for the cross-validation of double geographically weighted regression (DGWR)

Table 1 Accuracies of the DGWR-based precipitation estimates using four categories of gauge networks

Gauge networks	Modeling gauges				Validation gauges			
	CIR	MAE (mm/d)	SRMSE	CC	CIR	MAE (mm/d)	SRMSE	CC
Category I	0.87	1.1	1.36	0.85	0.84	1.6	1.77	0.76
Category II	0.87	1.1	1.33	0.85	0.85	1.7	1.67	0.77
Category III	0.87	1.1	1.35	0.85	0.85	1.6	1.61	0.78
Category IV	0.87	1.1	1.35	0.86	0.85	1.7	1.65	0.77

Note: The threshold value is set as 0.1 mm/d.

3.2.2 Processing the GPEs

Considering the spatio-temporal resolutions of the benchmark precipitation dataset and five GPEs, evaluation is carried out for 0.25°×0.25° grids at a daily scale. Thus, spatial aggregation and temporal accumulation are implemented for MSWEP to obtain daily precipitation at 0.25°×0.25° grids during 2006–2015. The spatial resolutions of TRMM 3B42 V7, CMORPH BLD and ERA5 meet the requirements, and the daily precipitation datasets are generated by summing in the time dimension.

4 Results and discussion

4.1 Overall performance

Figure 4 is the scatter plots of five GPEs versus the benchmark at 0.25°×0.25° grids. As can

be seen, the data points of MSWEP V2.2, V2.1, and CMORPH BLD are more concentrated. These three products demonstrate remarkably similar and high quantitative accuracies, with MAEs and CCs being 1.2 mm/d and 0.86, respectively. On the contrary, TRMM 3B42 V7 and ERA5 are relatively scattered with larger errors. Figure 5 shows the spatial distributions of daily mean precipitation of the benchmark and five GPEs. All GPEs clearly depict spatial variation of daily precipitation in the BB, and indicate that precipitation gradually decreases from south to north. MSWEP V2.2 and V2.1 successfully present the low- and high-precipit-

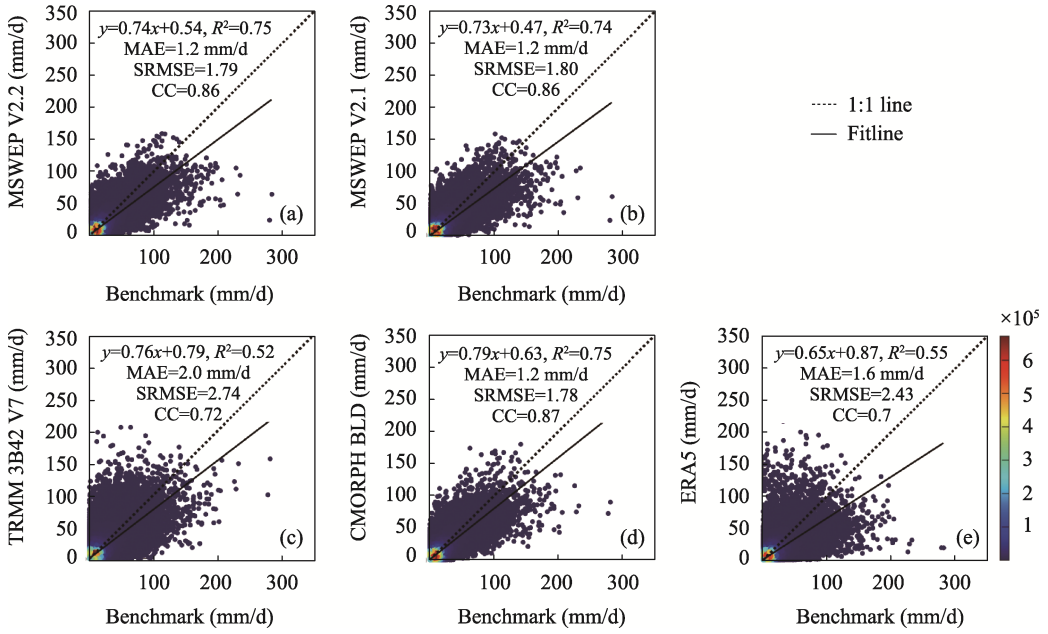

Figure 4 Scatter plots of daily precipitation of the five gridded precipitation estimates in the Bengbu Basin versus the benchmark dataset at 0.25°×0.25° grids

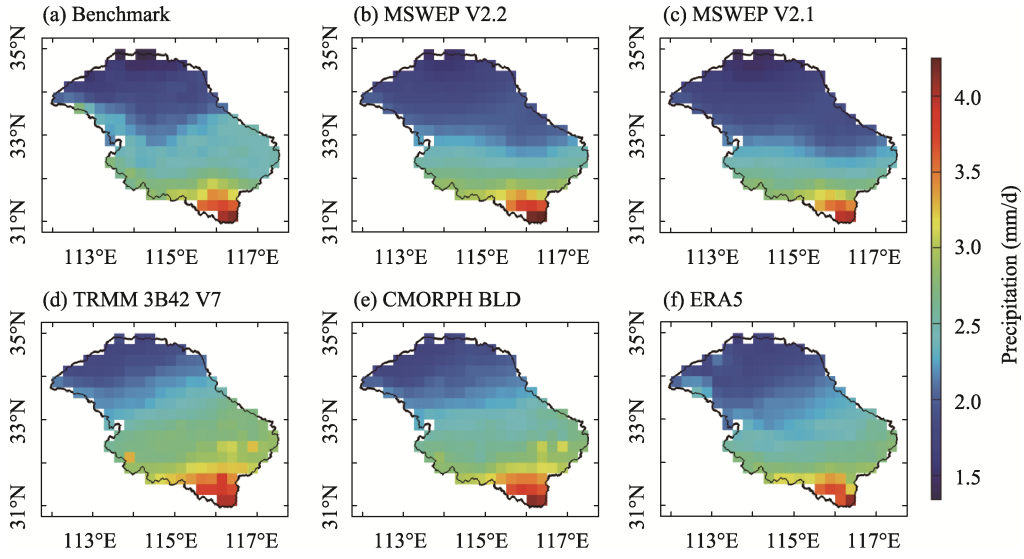

Figure 5 Spatial distributions of daily mean precipitation of the benchmark dataset and the five gridded precipitation estimates in the Bengbu Basin from 2006 to 2015

ation areas in the north and south, respectively. However, the precipitation in the middle of the BB is underestimated in MSWEP V2.2 and V2.1, while other three GPEs overestimate the precipitation in the north. The spatio-temporal accuracies of GPEs are discussed in the following sections 4.2 and 4.3.

4.2 Temporal accuracy

Table 2 summarizes the mean values of the temporal accuracy indexes of the five GPEs over all 0.25°×0.25° grids. The performances of MSWEP V2.2 and V2.1 are basically the same, as both their VHIs are close to 1, VFARs are smaller than 0.1, and VCSIs are larger than 0.9. The results indicate that both V2.2 and V2.1 have great classification abilities. Furthermore, the CCs of the two products reach up to 0.87, which indicates that the temporal variations of precipitation revealed by both MSWEPs are reasonably comparable to those of the benchmark. The SRMSEs of MSWEPs are 1.79. Two satellite-based precipitation products show a little bit overestimation. It is interesting to note that the identification ability and quantitative accuracy of CMORPH BLD are comparable to those of MSWEPs. Although VHI, VFAR, and VCSI of the ERA5 are similar to those of the CMORPH BLD, the errors of ERA5 products are significantly larger. The performance of TRMM 3B42 V7 is the worst. Overall, the results show that the five GPEs identify the precipitation events reasonably well. However, their quantitative errors cannot be neglected. Similar to the findings of the scatter plots, MSWEP V2.2 and V2.1 and CMORPH BLD have the highest temporal accuracies, followed by ERA5 and TRMM 3B42 V7.

Table 2 Mean values of the temporal accuracy indexes of the five gridded precipitation estimates compared to the benchmark over the Bengbu Basin during 2006–2015

Indexes	MSWEP V2.2	MSWEP V2.1	TRMM 3B42 V7	CMORPH BLD	ERA5
VHI	0.99	0.98	0.91	1.00	0.99
VFAR	0.07	0.06	0.12	0.07	0.09
VCSI	0.92	0.92	0.81	0.93	0.91
CC	0.87	0.87	0.72	0.87	0.75
ME (mm/d)	0.0	–0.1	0.3	0.2	0.1
MAE (mm/d)	1.2	1.2	2.0	1.2	1.6
SRMSE	1.79	1.80	2.77	1.80	2.45

Figure 6 shows the spatial variations of the temporal accuracy indexes of the five GPEs. Except for TRMM 3B42 V7, the other four GPEs yield highly spatially homogeneous VHIs, and slightly higher VFARs and VCSIs in the northwestern hilly areas than in the south. The spatial patterns of the classification indexes are mostly identical for MSWEP V2.2 and V2.1. Additionally, the VCSIs of CMORPH BLD in the south are higher than those of the MSWEPs. Owing to local topographical effects, the TRMM 3B42 V7 and CMORPH BLD produce higher VFARs, but considerably lower VCSIs at some grids in the Funiu Mountains and Tongbai Mountains. Furthermore, abnormal VFARs and VCSIs are produced at some isolated grids in the southeast of the BB (Figure 1b) where there are many large water bodies, such as Gaotang Lake, Wafu Lake and Chengdong Lake. The presence of these water bodies generate similar observation signals to trace precipitation when high-frequency microwaves

detect the inland water bodies, thereby causing significant false alarms (Tian and Peters-Lidard, 2007). The indexes illustrated in Figure 6 also reflect the quantitative estimation of the considered products. For all GPEs, MAE values increase from north to south, while SRMSE values increase from south to north because the precipitation gradually reduces northwards. The errors of TRMM 3B42 V7 and ERA5 are greater than other products. The CCs of CMORPH BLD in several grids in the southern region are higher than those of the MSWEPs.

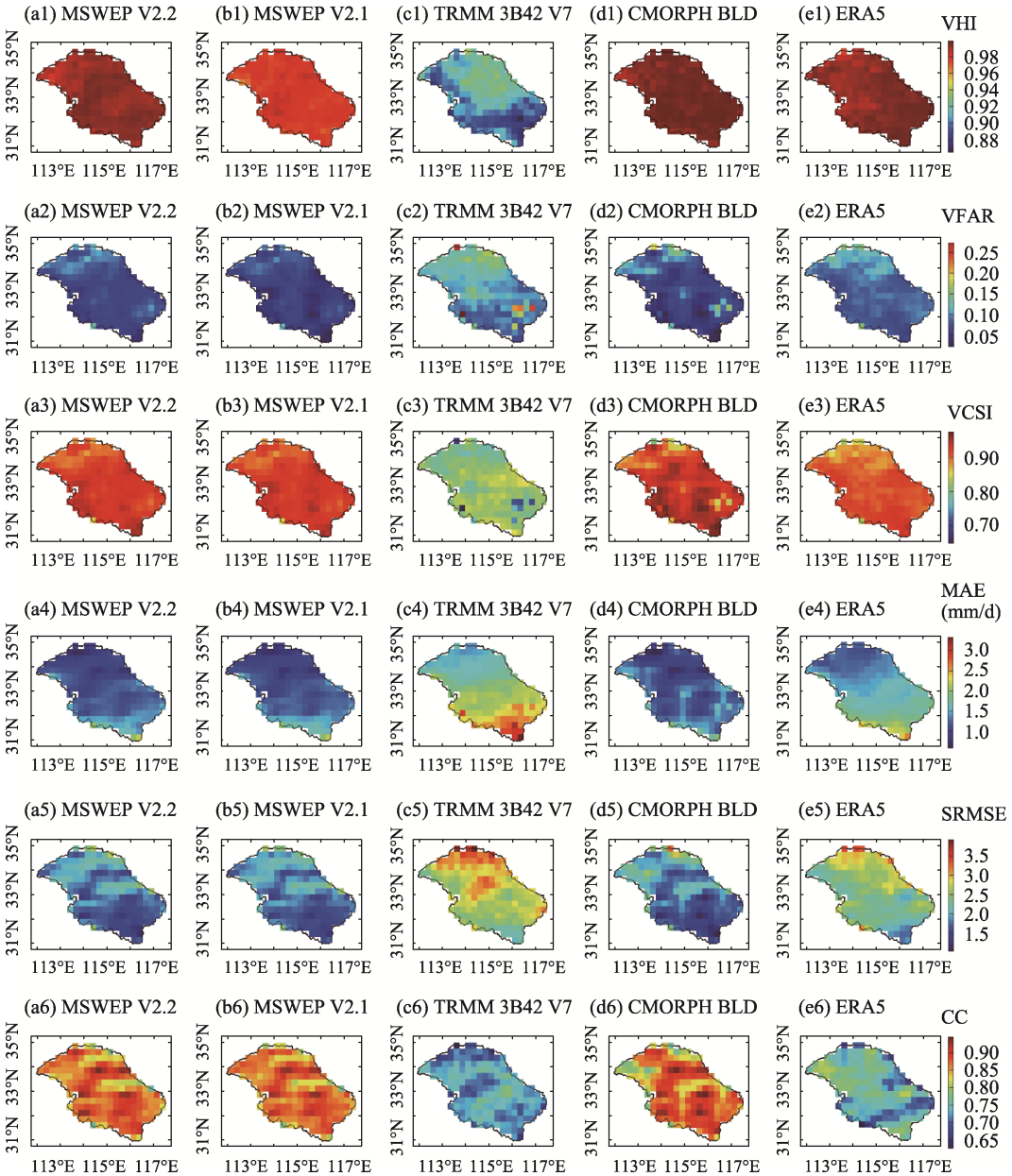

Figure 6 Spatial patterns of temporal accuracy indexes for the five gridded precipitation estimates in the Bengbu Basin compared to the benchmark at 0.25°×0.25° grids

Overall, the MSWEPs demonstrate a higher temporal accuracy and smaller spatial variability compared to the other GPEs. In other words, the performances of the MSWEPs in characterizing temporal variations of daily precipitation over the whole basin are reasonably stable because of the ensemble algorithm adopted. In the MSWEP algorithm, the weights of the four uncorrected satellite-based precipitation products and two reanalysis-based precipitation products at each grid are calculated against the three-day average gauge-based precipitation. The precipitation of different products is then merged by different weights. Thus, the spatial differences of the temporal accuracies of these input datasets are considered. By maintaining the estimation performance of superior datasets in high-precision areas and absorbing the advantages of other data in low-precision areas, it is possible to improve the accuracy of precipitation estimation over a bigger area. The weight assigned to the CMORPH dataset is significantly heavier than the other datasets (Beck *et al*., 2019). As a result, the MSWEPs and CMORPH have relatively consistent spatial patterns of temporal accuracy indexes. Abnormal indicators in the southeast where lakes are concentrated, disappear significantly by merging the reanalysis-based precipitation. Moreover, the CMORPH BLD outperforms MSWEPs at some locations. The gauge network used by the MSWEPs and CMORPH BLD are different in terms of densities and locations, and this difference can induce variations of temporal accuracies at some grids. VHIs of MSWEP V2.2 are slightly higher than those of V2.1, and this is probably due to the quality control of the GridSat infrared archive and the rescaling of the CDF-corrected merged precipitation.

4.3 Spatial accuracy

Table 3 summarizes the mean values of the spatial accuracy indexes for the five GPEs. MSWEP V2.2 demonstrates a greater improvement on spatial classification accuracy than on temporal accuracy compared to MSWEP V2.1. Based on the results of the VHI, VFAR and VCSI in Table 3, the CMORPH BLD outperforms MSWEP V2.2. The spatial VFAR of TRMM 3B42 V7 is smaller than that of ERA5, but the VHI and VCSI indicate a better performance of ERA5 than TRMM 3B42 V7. The CMORPH BLD is the best product in terms of characterizing the spatial structure of daily precipitation in the BB, although its mean S is close to 0.5. The MSWEP V2.2 is weaker than V2.1 in depicting the spatial distribution of precipitation, and the ERA5 is inferior to TRMM 3B42 V7. Quantitative errors suggest that the satellite- and reanalysis-based products generally underestimate the daily precipitation, whereas MSWEPs slightly overestimate it. The MAEs of MSWEP V2.2 and CMORPH BLD

Table 3 Mean values of the spatial accuracy metrics of the five gridded precipitation estimates compared to the benchmark dataset over the Bengbu Basin from 2006 to 2015

Metrics	MSWEP V2.2	MSWEP V2.1	TRMM 3B42 V7	CMORPH BLD	ERA5
VHI	0.93	0.87	0.77	0.96	0.94
VFAR	0.18	0.15	0.18	0.16	0.21
VCSI	0.79	0.75	0.66	0.82	0.76
S	0.41	0.44	0.37	0.49	0.29
ME (mm/d)	–0.2	–0.3	0.3	0.3	0.1
MAE (mm/d)	2.5	2.6	4.3	2.5	3.4
SRMSE	2.62	2.50	6.54	2.84	2.90

are both 2.5 mm/d, which are slightly smaller than that of the MSWEP V2.1. In contrast, the mean SRMSE of MSWEP V2.1 is the smallest, followed by MSWEP V2.2 and CMORPH BLD. Notably, the quantitative error of TRMM 3B42 V7 is significantly greater than that of ERA5.

The temporal variations of the spatial accuracy indexes are further analyzed. The temporal continuity of some spatial indexes, such as POD, is affected by the spatio-temporal distribution of the daily precipitation. Since precipitation is highly random, spatial accuracy indexes show considerable temporal fluctuations. As a result, moving averages of the accuracy indicators are calculated by using a 30-day sliding window to extract their major characteristics. Particularly, VCSI, S and MAE are selected (Figure 7). The abilities of GPEs to identify precipitation events present conspicuous seasonal variations. The GPE products are more capable during flood seasons with small differences. In dry seasons, the performance of MSWEP V2.2 is better than MSWEP V2.1 and CMORPH BLD. It is remarkable that ERA5 outperforms other products during specific periods. However, the VCSIs of TRMM 3B42 V7 are exceptionally low (even close to 0) compared to that of others. It can be summarized that the S values are low during rainy seasons, whereas they are high during dry seasons, and a contrasting pattern is observed in MAE values, which indicates that the

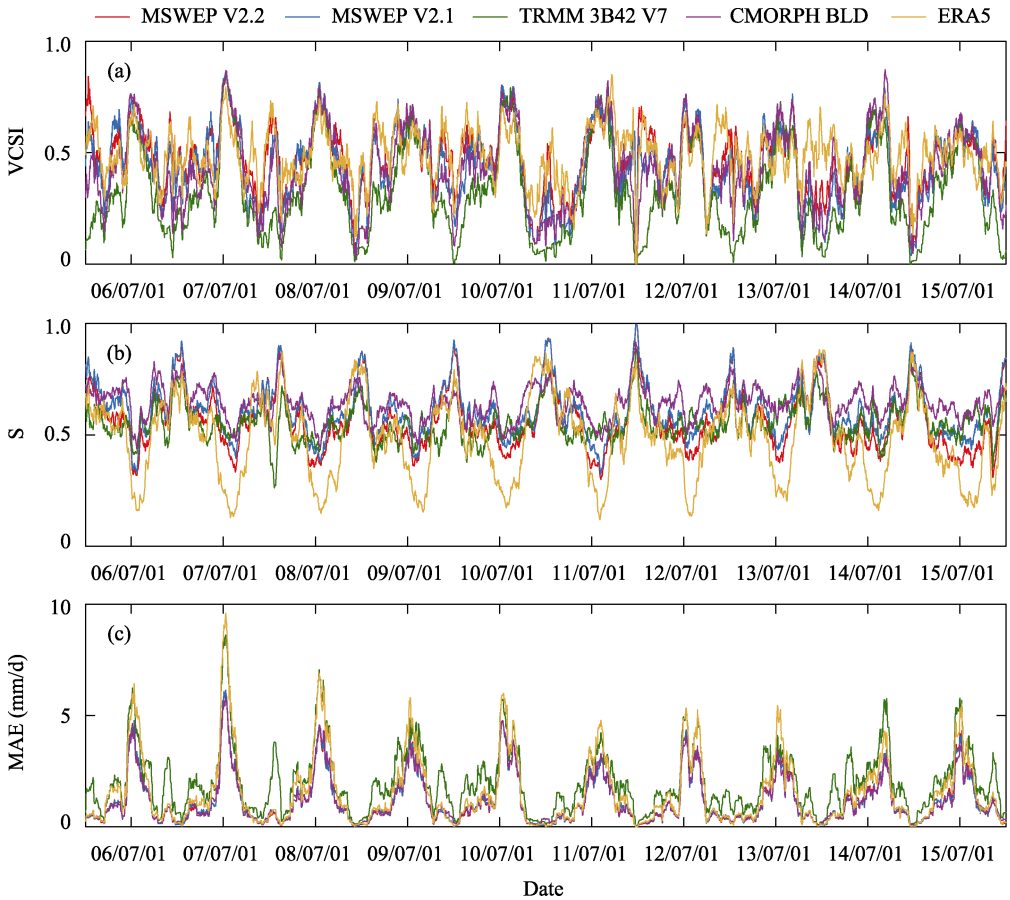

Figure 7 Temporal variations of spatial accuracy indexes of the five gridded precipitation estimates in the Bengbu Basin versus the benchmark dataset

capabilities of characterizing precipitation structure and estimating precipitation are weaker during rainy seasons. The S values of MSWEP V2.2 are lower than those of MSWEP V2.1 in most periods, and are only higher than those of ERA5 during the rainy season. Compared to other GPEs, the CMORPH BLD has the best capability of capturing spatial patterns of precipitation during the rainy season. The MAEs of MSWEP V2.2, MSWEP V2.1 and CMORPH BLD vary synchronously, and they are substantially smaller than those of the TRMM 3B42 V7 and ERA5.

Table 3 shows the S values averaged over all periods, and their temporal variations are given in Figure 7. In fact, the averaged S values over all grids and temporal periods vary with grid points, as shown in Figure 8. The results indicate that MSWEP V2.2 and MSWEP V2.1 reasonably characterize the spatial distribution in the northwestern and southern BB. The S values of CMORPH BLD are generally larger than those of MSWEPs, and they are close to 0.6 in some grids in the southern region with large precipitation. The S values of TRMM 3B42 V7 are mainly between 0.3 and 0.4, except some greater values in the northwestern hilly areas. However, the ERA5 performs the worst in presenting the spatial structure of precipitation over the BB.

Although MSWEPs have a low spatial quantitative error, they are still worse than the CMORPH BLD in terms of spatially identifying daily precipitation and describing the spatial structure of precipitation over the BB. The reason is that the movement vectors of cloud systems and the spatial structure of precipitation are considered in the CMORPH BLD. In contrast, MSWEPs merge multi-source precipitation information at each grid independently. In this way, the spatial correlations of precipitation between neighboring grids are not included in MSWEPs, which could naturally disturb local spatial patterns. Moreover, the extreme events of the CDF-corrected merged precipitation are rescaled, so that the variation patterns are consistent with uncorrected merged data in the fusion scheme of MSWEP V2.2.

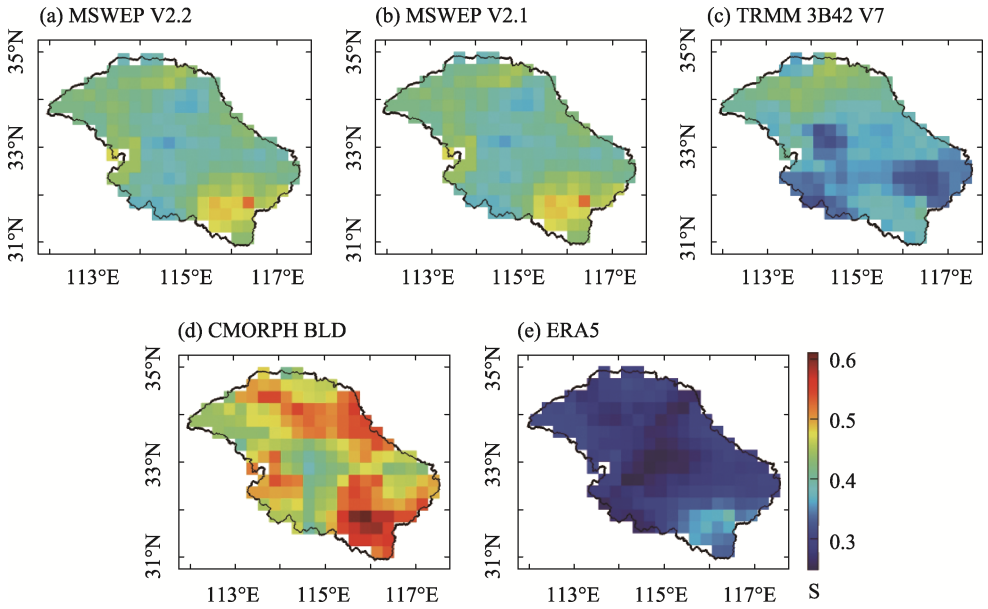

Figure 8 Spatial variations of the structure indexes of the five gridded precipitation estimates in the Bengbu Basin versus the benchmark dataset at 0.25°×0.25° grids

Notably, the trend of precipitation in the rescaling process is obtained through simple linear regression, which may result in the decrease of the mean S of MSWEP V2.2 compared to that of V2.1 in the rainy season (Beck *et al*., 2019b).

4.4 Estimation performances under different precipitation intensities

Figure 9 provides the CIRs and MAE_{PI}s of the five GPEs for daily precipitation events under different intensity conditions. All GPEs can identify the no-rain events considerably better than the rain events. When precipitation intensity is lower than 25 mm/d, the identification abilities of all GPEs are gradually strengthened as the precipitation intensity increases. However, when the precipitation intensity exceeds 25 mm/d, the CIR tends to decline. The MSWEP V2.2 is able to identify rain events more effectively than the MSWEP V2.1, but its performance is not as good as that of the CMORPH BLD. The CIR values of all GPEs, except the ERA5, are approximately 0.2 for the precipitation events whose intensities are larger than 100 mm/d. This indicates substantially low capacity of identifying extreme precipitation. The CIR value of the TRMM 3B42 V7 is the largest among all products, which indicates it has the greatest ability to identify extreme precipitation events with intensity larger than 100 mm/d. In terms of quantitative error, the MAE_{PI}s of the five GPEs show a significantly increasing trend with the precipitation intensity. When the precipitation intensity is lower than 50 mm/d, MSWEP V2.2, MSWEP V2.1 and CMORPH BLD show close MAE_{PI}s. Beyond this threshold, the MAE_{PI}s of MSWEP V2.2 and MSWEP V2.1 are greater than that of the CMORPH BLD. When precipitation intensity exceeds 100 mm/d, the MAE_{PI}s of MSWEPs are larger than that of the TRMM 3B42 V7, while the MAE_{PI} of the ERA5 is even larger than 70 mm/d.

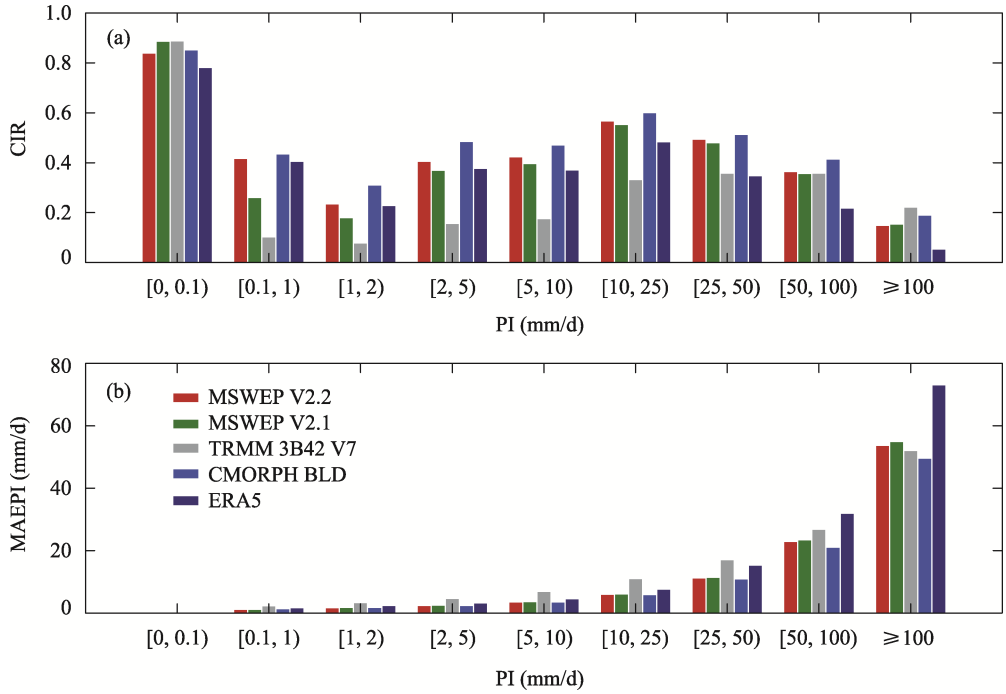

Figure 9 Correct identification ratios (CIRs) and mean absolute errors (MAE_{PI}s) of the five gridded precipitation estimates in the Bengbu Basin under different precipitation intensities

Considering that the precipitation in the BB gradually reduces northwards, the latitudinal variations of the CIRs of all GPEs for extreme daily precipitation events (≥100 mm/d) in each month are further investigated (Figure 10). Figure 10a shows that extreme precipitation events are concentrated in the south of the BB, and mainly occur from June to August. However, some of the extreme precipitation events in the north of the BB are missed in the MSWEPs. The distribution maps of the two versions are very similar. It is possibly because the ERA-interim data underestimates the extreme precipitation due to the influence of the model structure and its parameters. Then, when the ERA-interim data is merged into the MSWEP, extreme precipitation in the north of the BB is overlooked (Beck *et al*., 2019b). Although missing and false phenomena also occur in the TRMM 3B42 V7, the spatio-temporal distribution of the CIR for extreme events of TRMM 3B42 V7 is reasonably consistent with that of the benchmark. Numerous missed precipitation events are also observed in the south of BB for the CMORPH BLD, but it is still better than the ERA5 on capturing extreme precipitation.

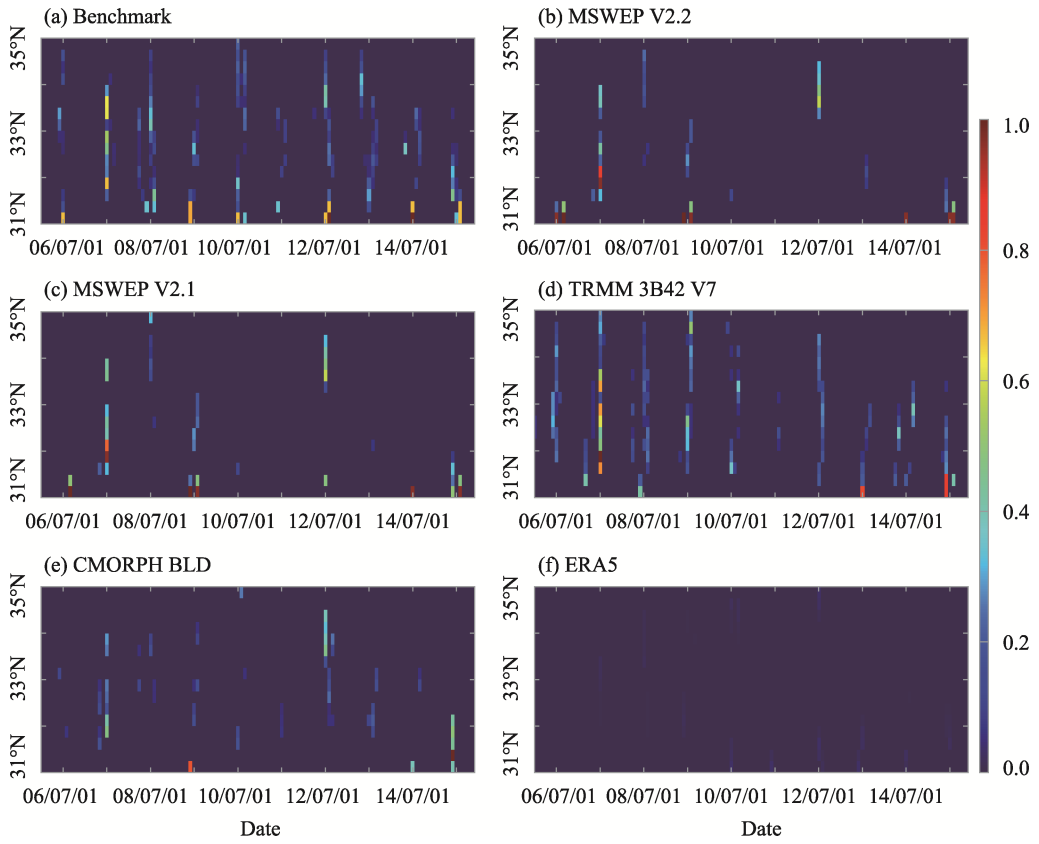

Figure 10 Latitudinal variations of the CIRs of all the gridded precipitation estimates for extreme precipitation events (≥100 mm/d) in all months over the Bengbu Basin from 2016 to 2015

in estimating precipitation with different intensities.

Overall, MSWEPs and the CMORPH BLD outperform the other GPEs in identifying precipitation events with intensities <100 mm/d. While for the precipitation intensity ≥100 mm/d, the comprehensive performance of TRMM 3B42 V7 is the best. At present, in the

process of calculating the weights of satellite- and reanalysis-based precipitation data, the fusion algorithm of MSWEP does not consider the different abilities of various data sources

5 Conclusions

This study comprehensively compared the performances of MSWEPs (V2.1 and V2.2), representative satellite-based (TRMM 3B42 V7 and CMORPH BLD) and reanalysis-based (ERA5) precipitation products with respect to the daily precipitation in the BB. The advantages and disadvantages of MSWEPs were examined, and the probable causes were discussed. The main conclusions are as follows:

(1) Except the TRMM 3B42 V7, the mean VHI, VFAR and VCSI of other GPEs are all close to perfect values. The MAE of MSWEP V2.2, MSWEP V2.1 and CMORPH BLD are similar, and their accuracies are higher than those of the ERA5 and TRMM 3B42 V7. For the MSWEPs, the weighted ensemble algorithm is employed to merge the merits of multiple satellite and reanalysis datasets in various locations. As a result, the spatial heterogeneity of the temporal indicators decreases significantly. Hence, the MSWEPs have better temporal performances than other GPEs over the BB.

(2) For spatial accuracy, the CMORPH BLD has better overall classification results than the MSWEPs. Furthermore, it reproduces a better spatial pattern of daily precipitation, particularly in the south of the BB. However, it has a larger value of mean SRMSE than that of MSWEPs. The spatial accuracy indexes exhibit remarkable seasonal variations. The precipitation identification is more effective in rainy season than that in dry season in all GPE products. However, the capture of spatial patterns of precipitation and quantitative accuracies are better in dry seasons than in rainy seasons. The spatial structure of daily precipitation in rainy seasons estimated by MSWEP V2.2 is quite different to that of the benchmark dataset. The S value of MSWEP V2.2 is only higher than that of the ERA5, because it merges precipitation data from various sources at different weights grid-by-grid independently which results in disturbance to the spatial structures of the source data. Hence, considering local spatial self-correlation of precipitation is beneficial for optimizing the ensemble algorithm.

(3) The optimal GPE vary with precipitation intensity. Neither the MSWEPs nor the CMORPH BLD is always preferable. When the precipitation intensity exceeds 100 mm/d, the TRMM 3B42 V7 provides the best estimation. Based on the accuracies of source datasets under different precipitation intensity conditions, weights should be adjusted accordingly to further improve the precipitation estimation of the MSWEPs.

Overall, compared to satellite- and reanalysis-based precipitation products, the MSWEPs provide the best temporally-accurate estimates with excellent spatial homogeneity in the BB. However, there are still space of improvement for MSWEP in characterizing the precipitation spatial structure and estimating extreme precipitation. With regarding to these two aspects, this research provides meaningful suggestions to optimize the fusion algorithm. These suggestions are also important references for the innovation of multi-source precipitation integration methods.

References

AghaKouchak A, Mehran A, 2013. Extended contingency table: Performance metrics for satellite observations

and climate model simulations. *Water Resources Research*, 49(10): 7144–7149.

Ashouri H, Hsu K L, Sorooshian S *et al*., 2015. PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, 96(1): 69–83.

Awange J L, Hu K X, Khaki M, 2019. The newly merged satellite remotely sensed, gauge and reanalysis-based Multi-Source Weighted-Ensemble Precipitation: Evaluation over Australia and Africa (1981–2016). *Science of the Total Environment*, 670: 448–465.

Barancourt C, Creutin J D, Rivoirard J, 1992. A method for delineating and estimating rainfall fields. *Water Resources Research*, 28(4): 1133–1144.

Beck H E, Pan M, Miralles D G *et al*., 2021. Evaluation of 18 satellite-and model-based soil moisture products using in situ measurements from 826 sensors. *Hydrology and Earth System Sciences*, 25(1): 17–40.

Beck H E, Pan M, Roy T *et al*., 2019a. Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, 23(1): 207–224.

Beck H E, Van Dijk A I J M, Levizzani V *et al*., 2017a. MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Science*, 21(1): 589–615.

Beck H E, Vergopolan N, Pan M *et al*., 2017b. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, 21(12): 6201–6217.

Beck H E, Wood E F, Pan M *et al*., 2019b. MSWEP v2 global 3-hourly 0.1 precipitation: Methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, 100(3): 473–500.

Brunsdon C, Fotheringham S, Charlton M, 1998. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D* (*The Statistician*), 47(3): 431–443.

Chao L, Zhang K, Li Z *et al*., 2018. Geographically weighted regression based methods for merging satellite and gauge precipitation. *Journal of Hydrology*, 558: 275–289.

Chen S, Zhang L, Zhang Y *et al*., 2020. Evaluation of Tropical Rainfall Measuring Mission (TRMM) satellite precipitation products for drought monitoring over the middle and lower reaches of the Yangtze River Basin, China. *Journal of Geographical Sciences*, 30(1): 53–67.

Dos Reis J B C, Rennó C D, Lopes E S S, 2017. Validation of satellite rainfall products over a mountainous watershed in a humid subtropical climate region of Brazil. *Remote Sensing*, 9(12): 1240.

Ebita A, Kobayashi S, Ota Y *et al*., 2011. The Japanese 55-year reanalysis "JRA-55": An interim report. *Sola*, 7: 149–152.

Hersbach H, Bell B, Berrisford P *et al*., 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730): 1999–2049.

Hou A Y, Kakar R K, Neeck S *et al*., 2014. The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 95(5): 701–722.

Huffman G J, Adler R F, Bolvin D T *et al*., 2010. The TRMM Multi-satellite Precipitation Analysis (TMPA). In: Satellite Rainfall Applications for Surface Hydrology. Dordrecht: Springer, 3–22.

Huffman G J, Bolvin D T, Nelkin E J *et al*., 2007. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1): 38–55.

Jiang S, Wei L, Ren L *et al*., 2020. Utility of integrated IMERG precipitation and GLEAM potential evapotranspiration products for drought monitoring over mainland of China. *Atmospheric Research*, 247: 105141.

Joyce R J, Janowiak J E, Arkin P A *et al*., 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, 5(3): 487–503.

Kirschbaum D B, Huffman G J, Adler R F *et al*., 2017. NASA's remotely sensed precipitation: A reservoir for applications users. *Bulletin of the American Meteorological Society*, 98(6): 1169–1184.

Li X, Chen Y, Wang H *et al*., 2020. Assessment of GPM IMERG and radar quantitative precipitation estimation

(QPE) products using dense rain gauge observations in the Guangdong-Hong Kong-Macao Greater Bay Area, China. *Atmospheric Research*, 236: 104834.

Liu J, Shangguan D, Liu S *et al*., 2019. 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.

Ma Q, Li Y, Feng H *et al*., 2020. Performance evaluation and correction of precipitation data using the 20-year IMERG and TMPA precipitation products in diverse subregions of China. *Atmospheric Research*, 249: 105304.

Maggioni V, Meyers P C, Robinson M D, 2016. A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) era. *Journal of Hydrometeorology*, 17(4): 1101–1117.

Massari C, Crow W, Brocca L, 2017. An assessment of the performance of global rainfall estimates without ground-based observations. *Hydrology and Earth System Sciences*, 21(9): 4347.

Monsieurs E, Kirschbaum D B, Tan J *et al*., 2018. Evaluating TMPA rainfall over the sparsely gauged East African Rift. *Journal of Hydrometeorology*, 19(9): 1507–1528.

Palash W, Akanda A S, Islam S, 2020. The record 2017 flood in South Asia: State of prediction and performance of a data-driven requisitely simple forecast model. *Journal of Hydrology*, 589: 125190.

Plouffe C C F, Robertson C, Chandrapala L, 2015. Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: A case study of Sri Lanka. *Environmental Modelling & Software*, 67: 57–71.

Prakash S, 2019. Performance assessment of CHIRPS, MSWEP, SM2RAIN-CCI, and TMPA precipitation products across India. *Journal of Hydrology*, 571: 50–59.

Robertson C, Long J A, Nathoo F S *et al*., 2014. Assessing quality of spatial models using the structural similarity index and posterior predictive checks. *Geographical Analysis*, 46(1): 53–74.

Shen Y, Xiong A, 2016. Validation and comparison of a new gauge-based precipitation analysis over mainland of China. *International Journal of Climatology*, 36(1): 252–265.

Su J, Lü H, Zhu Y *et al*., 2019. Evaluating the hydrological utility of latest IMERG products over the Upper Huaihe River Basin, China. *Atmospheric Research*, 225: 17–29.

Sun Q, Miao C, Duan Q *et al*., 2018. A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, 56(1): 79–107.

Tang G, Clark M P, Papalexiou S M *et al*., 2020. Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. *Remote Sensing of Environment*, 240: 111697.

Tapiador F J, Turk F J, Petersen W *et al*., 2012. Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, 104: 70–97.

Tarek M, Brissette F P, Arsenault R, 2020. Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, 24(5): 2527–2544.

Tian Y, Peters-Lidard C D, 2007. Systematic anomalies over inland water bodies in satellite-based precipitation estimates. *Geophysical Research Letters*, 34(14): L14403.

Xu Z, Wu Z, He H *et al*., 2019. Evaluating the accuracy of MSWEP V2.1 and its performance for drought monitoring over mainland of China. *Atmospheric Research*, 226: 17–31.

Zhang Y, Chen Q, Xia J, 2020. Investigation on flood event variations at space and time scales in the Huaihe River Basin of China using flood behavior classification. *Journal of Geographical Sciences*, 30(12): 2053–2075.