WHU-SGCC：A novel approach for blending daily satellite (CHIRP) and precipitation observations over the Jinsha River Basin

Gaoyun Shen[1], Nengcheng Chen[1,2], Wei Wang[1,2], Zeqiang Chen [1,2]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
[2] Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China

Correspondence to: ZeqiangChen@whu.edu.cn; Tel.: +86 13871025965; Fax: +86 27 68778229.

Abstract. Accurate and consistent satellite-based precipitation estimates blended with rain gauge data are important for regional precipitation monitoring and hydrological applications, especially in regions with limited rain gauges. However, existing fusion precipitation estimates often have large uncertainties over mountainous areas with complex topography and sparse rain gauges, and the existing data blending algorithms are very bad at removing the day-by-day random errors. Therefore, the development of effective methods for high-accuracy precipitation estimates over complex terrain and on a daily scale is of vital importance for mountainous hydrological applications. This study aims to offer a novel approach for blending daily precipitation gauge data and the Climate Hazards Group Infrared Precipitation (CHIRP, daily, 0.05°) satellite-derived precipitation developed by the UC Santa Barbara over the Jinsha River Basin for the period of June-July-August from 1994 to 2014. This method is named the Wuhan University Satellite and Gauge precipitation Collaborated Correction (WHU-SGCC). The results show that the WHU-SGCC method is effective in liquid precipitation bias adjustments from point to surface, which is evaluated by multiple error statistics and from different perspectives. Moreover, the accuracy of the spatial distribution of the precipitation estimates derived from the WHU-SGCC method is related to the complexity of the topography. The validation also verifies that the proposed approach is effective in the detection of the major precipitation events inside the Jinsha River Basin with the daily precipitation less than 25 mm. This study indicates that the WHU-SGCC approach is a promising tool to monitor monsoon precipitation over the Jinsha River Basin, the complicated mountainous terrain with sparse rain gauge data, considering the spatial correlation and the historical precipitation characteristics. The daily precipitation estimations at 0.05° resolution over the Jinsha River Basin during the summer seasons from 1990 to 2014, derived from WHU-SGCC are available at the PANGAEA Data Publisher for Earth & Environmental Science portal (https://doi.pangaea.de/10.1594/PANGAEA.896615)

1 Introduction

Accurate and consistent estimates of precipitation are vital for hydrological modelling, flood forecasting and climatological studies in support of better planning and decision making (Agutu et al., 2017;Cattani et al., 2018;Roy et al., 2017). In general, ground-based gauge networks include a substantial number of liquid precipitation observations measured with high accuracy, high temporal resolution, and long historical records. However, sparse distribution and point measurements limit the accurate estimation of spatially gridded rainfall (Martens et al., 2013).

Due to the sparseness of rain gauges and their uneven spatial distribution and high proportion of missing data, satellite-derived precipitation data are an attractive supplement offering the advantage of plentiful information with high spatio-temporal resolution over widespread regions, particularly over oceans, high elevation mountainous regions, and other remote regions where gauge networks are difficult to deploy. However, satellite estimates are susceptible to systematic biases that can influence hydrological modelling and the retrieval algorithms are relatively insensitive to light rainfall events, especially in complex terrain, resulting in underestimation of the magnitude of precipitation events (Behrangi et al., 2014;Thiemig et al., 2013;Yang et al., 2017). Without adjustments, inaccurate satellite-based precipitation estimates will lead to unreliable

assessments of risk and reliability (AghaKouchak et al., 2011).

Accordingly, there are many kinds of precipitation estimates combining multiple sources datasets. Since 1997, the Tropical Rainfall Measurement Mission (TRMM) has improved satellite-based rainfall retrievals over tropical regions (Kummerow et al., 1998;Simpson et al., 1988), and then applies a stepwise method for blending daily TRMM Multisatellite Precipitation Analysis (TMPA) output with rain gauges in South America (Vila et al., 2009). The Global Precipitation Measurement (GPM) satellite was launched after the success of the TRMM satellite by the cooperation of National Aeronautics and Space Administration (NASA) and Japan Aerospace Exploration Agency (JAXA) on February 27, 2014 (Mahmoud et al., 2018;Ning et al., 2016). The main core observatory satellite (GPM) cooperates with the ten other satellites (partners) to offer the high spatiotemporal resolution products (0.1° × 0.1°- half- hourly) of the global real-time precipitation estimates (Mahmoud et al., 2019). The Geostationary Operational Environmental Satellite (GOES)-R Series is the geostationary weather satellites, which significantly improves the detection and observation of environmental phenomena. The Advanced Baseline Imager (ABI) onboard the GOES-R platform will provide images in 16 spectral bands, spatial resolution of 0.5 to 2 km (2 km in the infrared and 1–0.5 km in the visible), and full-disk scanning every 5 minutes over the continental United States. The GOES-R Series will offer the enhanced capabilities for satellite-based rainfall estimation and nowcasting (Behrangi et al., 2009;Schmit et al., 2005). The Global Precipitation Climatology Project (GPCP) is one of the successful projects for blending rain gauge analysis and multiple satellite-based precipitation estimates, and constructed a relatively coarse-resolution (monthly, 2.5° × 2.5°) global precipitation dataset (Adler et al., 2003;Huffman et al., 1997). To improve the resolution of this satellite-based dataset, the GPCC network data was incorporated into remote sensing information with Artificial Neural Networks (PERSIANN) rainfall estimates, which provides finer temporal and spatial resolutions (daily, 0.25° × 0.25°) (Ashouri et al., 2015). The CPC Merged Analysis of Precipitation (CMAP) product is a data blending and fusion analysis of gauge data and satellite-based precipitation estimates (Xie and Arkin, 1996). CMAP has a long-term dataset series from 1979, while the resolution is relatively coarse. Although the aforementioned products are widely used and have performed well, the data resolution cannot achieve high accuracy in precipitation monitoring over the Jinsha River Basin, China.

Currently, the Climate Hazards Group Infrared Precipitation with Station data (CHIRPS) developed by the UC Santa Barbara, which has a higher spatial resolution (0.05°), can solve the scale problem. CHIRPS is a long-term precipitation data series, which merges three types of information: global climatology, satellite estimates and in situ observations. Table 1 shows the temporal and spatial resolution of current major satellite-based precipitation datasets. The CHIRPS precipitation dataset with several temporal and spatial scales has been evaluated in Brazil (Nogueira et al., 2018;Paredes-Trejo et al., 2017), Chile (Yang et al., 2016;Zambrano-Bigiarini et al., 2017), China (Bai et al., 2018), Cyprus (Katsanos et al., 2016b;Katsanos et al., 2016a), India (Ali and Mishra, 2017) and Italy (Duan et al., 2016). Nevertheless, the temporal resolutions of the aforementioned applications were mainly at seasonal and monthly scales, lacking the evaluation of daily precipitation. Additionally, despite the great potential of gauge-satellite fusing products for large-scale environmental monitoring, there are still large discrepancies with ground observations at the sub-regional level where these data are applied. Furthermore, the CHIRPS product reliability has not been analysed in detail for the Jinsha River Basin, China, particularly on a daily scale. The existing research indicates that estimations over mountainous areas with complex topography often have large uncertainties and systematic errors due to the topography, seasonality, climate impact and sparseness of rain gauges (Derin et al., 2016;Maggioni and Massari, 2018;Zambrano-Bigiarini et al., 2017) Moreover, Bai et al., 2018 evaluates CHIRPS over mainland China and indicates that the performance of CHIRPS is poor over the Sichuan Basin and the Northern China Plain, which have complex terrains with substantial variations in elevation. Additionally, Trejo et al., 2016 shows that CHIRPS overestimates low monthly rainfall and underestimates high monthly rainfall using several numerical metrics, and rainfall event frequency is overestimated excluding the rainy season.

Table 1 Coverage and spatiotemporal resolutions of major satellite precipitation datasets

Product	Temporal resolution	Spatial resolution	Period	Coverage
TRMM 3B42	3hours	0.25°	1998-present	50°S-50°N
GPM	30min/Hourly/ 3hours/Daily/3Day/7 Day/Monthly	0.1°/0.25°/0.05°/5°	2014-present	60°S-60°N 70°N-70°S 90°N-90°S
GOES-R	5min/15min	0.5-2 km	2016-present	the continental United States/ western hemisphere
GPCP	Monthly/Pentad	2.5°	1979-(delayed) present	90°S-90°N
PERSIANN-CDR	Daily	0.25°	1983-(delayed) present	60°S-60°N
CMAP	Monthly	2.5°	1979- present	90°S-90°N
CHIRPS	Annual/Monthly/ Dekad/Pentad/Daily	0.05°/0.25°	1981- present	50°S-50°N

To overcome these limitations, many studies have focused on proposing effective methodologies for blending rain gauge observations and satellite-based precipitation estimates, and sometimes radar data to take advantage of each dataset. Many numerical models are established among these datasets for high-accuracy precipitation estimations, such as bias adjustment by a quantile mapping (QM) approach (Yang et al., 2016), Bayesian kriging (BK) (Verdin et al., 2015) and a conditional merging technique (Berndt et al., 2014). Among aforementioned methods, the QM approach is a distribution-based approach, which works with historical data for bias adjustment and is effective in reducing the systematic bias of regional climate model precipitation estimates at monthly or seasonal scales (Chen et al., 2013). However, the QM approach offers very limited improvement in removing day-by-day random errors. The BK approach shows very good model fit with precipitation observations. Unfortunately, the Gaussian assumption of the BK model is invalid for daily scales. Overall, there is a lack of effective methods for high-accuracy precipitation estimates over complex terrain on a daily scale.

As such, due to the poor performance of CHIRPS data at the sub-regional scale and the shortcomings of the existing blending algorithms, the aim of this article is to offer a novel approach for blending daily liquid precipitation gauge data and the Climate Hazards Group Infrared Precipitation (CHIRP) satellite-derived precipitation estimates developed by the UC Santa Barbara, over the Jinsha River Basin. Here, we will use precipitation to name liquid precipitation throughout the text. The CHIRP is the raw data of CHIRPS before blending in rain gauge data. The objective is to build corresponding precipitation models that consider terrain factors and precipitation characteristics to produce high-quality precipitation estimates. This novel method is named the Wuhan University Satellite and Gauge precipitation Collaborated Correction (WHU-SGCC) method. We demonstrate this method by applying it to daily precipitation over the Jinsha River Basin during the summer seasons from 1990 to 2014. The results support the validity of the proposed approach for producing refined satellite-gauge precipitation estimates over mountainous areas.

The remainder of this paper is organized as follows: Section 2 describes the study region and rain gauges and CHIRPS dataset used in this study. Section 3 presents the principle of the WHU-SGCC approach for high-accuracy precipitation estimates. The results and discussion are analysed in Section 4, and conclusions and future work are presented in Section 5.

2 Study Region and Data

2.1 Study Region

The Yangtze River, one of the largest and most important rivers in Southeast Asia, originates on the Tibetan Plateau and extends approximately 6300 km eastward to the East China Sea. The river's catchment proximately covers an area of approximately ~180×10^4 km^2 and the average annual precipitation is approximately 1100 mm (Zhang et al., 2019).Yangtze River is divided into nine sub-basins, the upper drainage basin is the Jinsha River Basin, which flows through the provinces of Qinghai, Sichuan, and Yunnan in western China. The total river length is 3486 km, accounting for 77% of the length of the upper Yangtze River, and covering a watershed area of 460×10^3 km^2. The location of the Jinsha River Basin is shown in Fig. 1, and covers the

eastern part of the Tibetan Plateau and the part of the Hengduan Mountains. The southern portion of the river basin is the Northern Yunnan Plateau and the eastern portion includes a wide area of the southwestern margin of the Sichuan basin. Crossing complex and varied terrains, the elevation of the Jinsha River ranges from 263 to 6575 m above sea level, which results in significant temporal and spatial weather variation within the basin. The average annual precipitation of the Jinsha River Basin is approximately 710 mm, the average annual precipitation of the lower reaches is approximately 900-1300 mm, while the average annual precipitation of the middle and upper reaches is approximately 600-800 mm (Yuan et al., 2018). The climate of the Jinsha River Basin has more precipitation during the summer season (June-July-August, JJA), which is affected by oceanic southwest and southeast monsoons. Therefore, the blending of satellite estimations with gauged observations during the JJA is the main focus of this research.

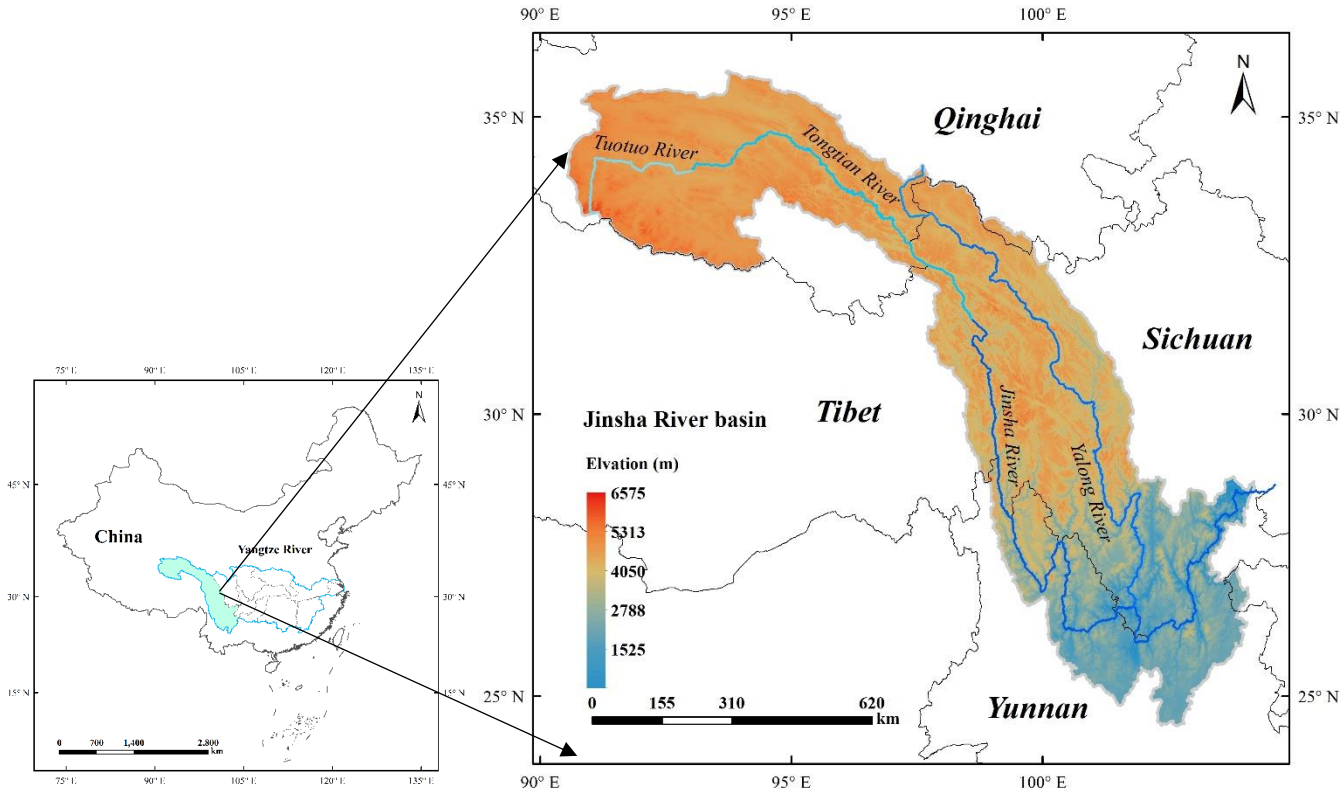

Figure 1 Location of the study area with key topographic features.

2.2 Study Data

2.2.1 Precipitation gauge observations

Daily rain gauge observations at 30 national standard rain stations in the Jinsha River Basin during the JJA from 1990 to 2014 were provided by the National Climate Centre (NCC) of the China Meteorological Administration (CMA) （http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_MUL_DAY_V3.0.html, last access: 10 December, 2018）, which imposes a strict quality control at station-provincial-state levels. Station identification numbers and relevant geographical characteristics are shown in Table 2, and their uneven spatial distribution is shown in Fig. 2. The selected rain gauges are located in Qinghai, Tibet, Sichuan and Yunnan Provinces but are mainly scattered in Sichuan Province, and the number of rain gauges in the northern river basin is less than in the southern river basin. In this study, the gauge observations were used as the reference data in bias adjustment of satellite precipitation estimations.

Table 2 Geographical characteristics of rain stations.

Station number	Province	Lat (°N)	Lon (°E)	Elevation (m)
52908	Qinghai	35.13	93.05	4823
56004	Qinghai	34.13	92.26	4744
56021	Qinghai	34.07	95.48	5049
56029	Qinghai	33.00	96.58	4510
56034	Qinghai	33.48	97.08	4503
56144	Tibet	31.48	98.35	4743
56038	Sichuan	32.59	98.06	4285
56146	Sichuan	31.37	100.00	4703
56152	Sichuan	32.17	100.20	4401
56167	Sichuan	30.59	101.07	3374
56247	Sichuan	30.00	99.06	2948
56251	Sichuan	30.56	100.19	4284
56257	Sichuan	30.00	100.16	3971
56357	Sichuan	29.03	100.18	4280
56374	Sichuan	30.03	101.58	3902
56459	Sichuan	27.56	101.16	3002
56462	Sichuan	29.00	101.30	4019
56475	Sichuan	28.39	102.31	1850
56479	Sichuan	28.00	102.51	2470
56485	Sichuan	28.16	103.35	2060
56565	Sichuan	27.26	101.31	2578
56571	Sichuan	27.54	102.16	1503
56666	Sichuan	26.35	101.43	1567
56671	Sichuan	26.39	102.15	1125
56543	Yunnan	27.50	99.42	3216
56586	Yunnan	27.21	103.43	2349
56651	Yunnan	26.51	100.13	2449
56664	Yunnan	26.38	101.16	1540
56684	Yunnan	26.24	103.15	2184
56778	Yunnan	25.00	102.39	1975

The multi-year (1990-2014) average annual precipitation during the JJA over the Jinsha River Basin increases from north to south (Fig. 2). The spatial distribution of precipitation is uneven, with an average annual precipitation ranging from less than 250 mm to more than 600 mm during the summer seasons.

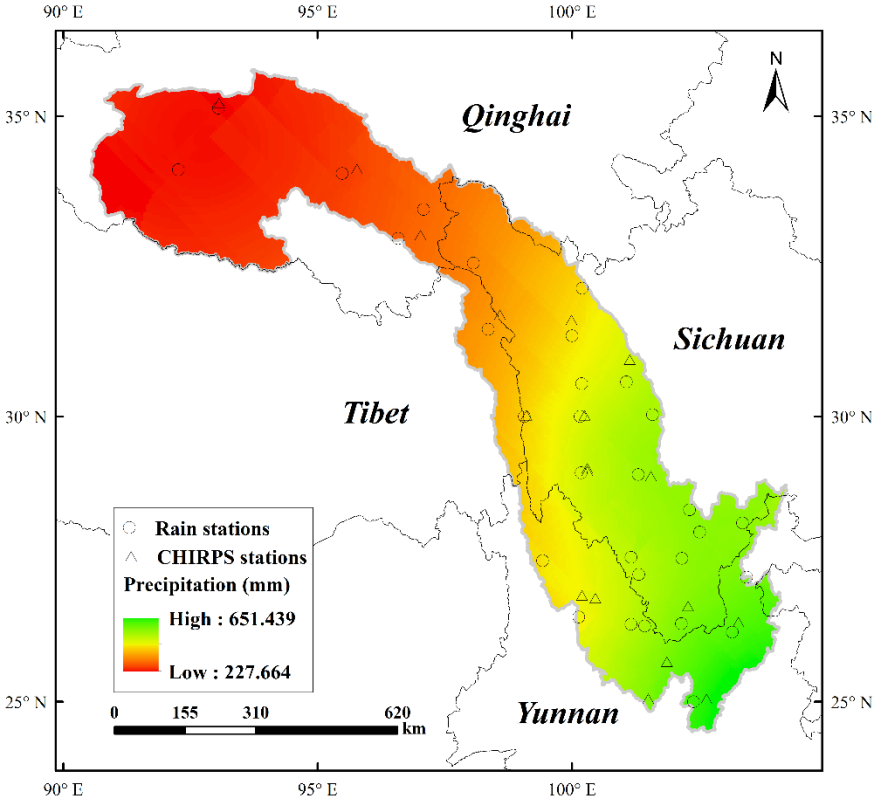

Figure 2 The multi-year (1990-2014) average annual precipitation during JJA over the Jinsha River Basin. 30 rain stations were provided by the China Meteorological Administration stations, the other 18 CHIRPS fusion stations were provided by the Climate Hazards Group UC Santa Barbara online at ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/diagnostics/global_monthly_station_density/tifs/p05/ (last

access: 10 December, 2018).

2.2.2 CHIRPS satellite-gauge fusion precipitation estimates

The CHIRPS v.2 dataset, a satellite-based daily rainfall product, is available online at ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/global_daily/tifs/p05/ (last access: 10 December, 2018). It covers a quasi-global area (land only, 50° S-50° N) with several temporal scales (daily, 3-day, 6-day or monthly time steps) and high spatial resolution (0.05°) (Rivera et al., 2018). This dataset contains a wide variety of satellite-based rainfall products derived from multiple data sources and incorporates four data types: monthly precipitation from CHPClim v.1.0 (Climate Hazards Group's Precipitation Climatology version 1) derived from the combination of the satellite fields, gridded physiographic indicators, and in situ climate normal with the geospatial modelling approach based on moving window regressions and inverse distance weighting interpolation (Funk et al., 2015 b), quasi-global geostationary thermal infrared satellite observations (TRMM 3B42 version 7), atmospheric model rainfall fields CFS (Climate Forecast System) from NOAA, and surface based precipitation observations from various sources including national or regional meteorological services. The differences from other frequently used precipitation products are the higher resolution of 0.05° and the longer-term data series from 1981 to the present (Funk et al., 2015 a).

CHIRPS is the blended product of a two-part process. First, IR precipitation (IRP) pentad rainfall estimates are fused with corresponding CHPClim pentad data to produce an unbiased gridded estimate, called the Climate Hazards Group IR Precipitation (CHIRP), which is available online at ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRP/daily/ (last access: 10 December, 2018). In the second part of the process, CHIRP data is blended with in situ precipitation observations obtained from a variety of sources including national and regional meteorological services by means of a modified inverse-distance weighting algorithm to create the final blended product, CHIRPS (Funk et al., 2014). The daily CHIRP satellite-based data over the Jinsha River Basin during the summer seasons from 1990 to 2014 was selected as the input for WHU-SGCC blending with rain observations, and the corresponding daily CHIRPS blended data was used for comparisons of precipitation accuracy.

The blended in situ daily precipitation observations come from a variety of sources such as: the daily GHCN archive (Durre et al., 2010), the Global Summary of the Day dataset (GSOD) provided by NOAA's National Climatic Data Center, the World Meteorological Organization's Global Telecommunication System (GTS) daily archive provided by NOAA CPC, and over a dozen national and regional meteorological services. The number of daily CHIRP observation stations in the Jinsha River Basin was only 18, compared to the 30 rain gauge stations provided by CMA (Fig. 2).

3 Methods

3.1 The WHU-SGCC approach

In this study, the approach of the WHU-SGCC is to estimate the precipitation for every pixel by blending satellite estimates and rain gauge observations considering the terrain factors and precipitation characteristics. There were four steps to establish the numerical relationship between gauge stations and the corresponding satellite pixels, and interpolation for the remaining pixels. On this basis, the WHU-SGCC method identifies the geographical locations and topographical features of each pixel and applies the four classification and blending rules. A flowchart of the WHU-SGCC method is shown in Fig. 3. The proposed approach was evaluated over the Jinsha River Basin based on 30 gauge stations and CHIRP satellite-based precipitation estimations during the JJA from 1990 to 2014. The leave-one-out cross validation step was applied to computing the out-of-sample adjusted error with gauge stations.

The basic description of the WHU-SGCC method is given below, with details illustrated separately in later sections:

1) Classify all regional pixels into four types: C1 (pixel including one gauge station in its area), C2 (pixel statistically similar to C1), C3 (pixel statistically similar to C2) and C4 (remaining pixels).

2) Analyse the relationships between precipitation observations and the C1, C2, and C3 pixel types, and with the C4 pixels. These relationships are described by four rules, detailed below as Rules 1 through 4.

3) Establish regression models and screen target pixels based on the five aforementioned rules.

4) Correct all precipitation pixels in daily regional precipitation images.

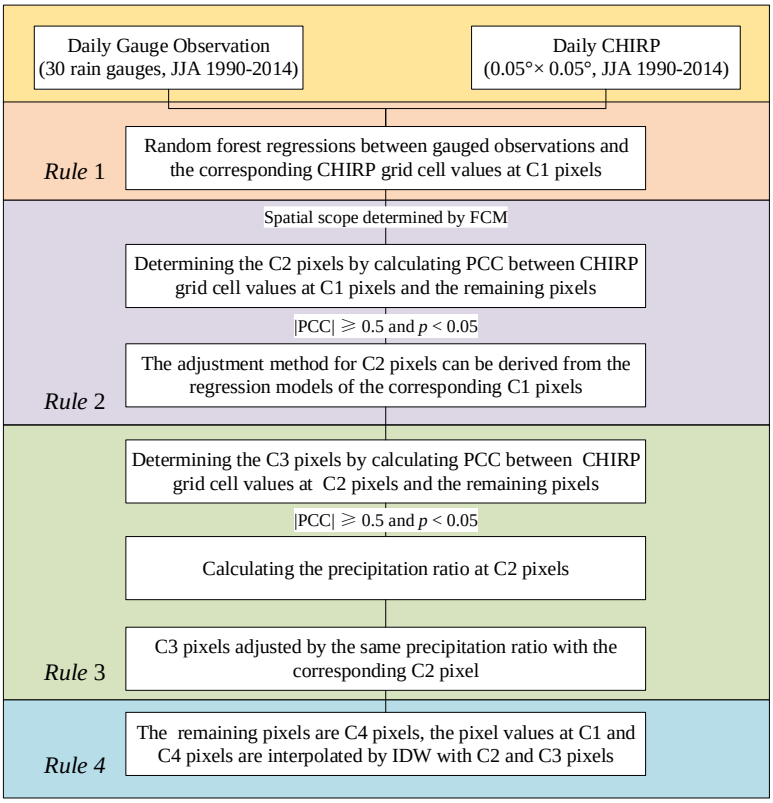

Figure 3 Flowchart of the WHU-SGCC approach with the four rules applied in this study.

3.1.1 Assumptions

1) Gauge observations are the most accurate, or "true", values for reference purposes.

2) No major terrain change occurred during the twenty years.

3) Pearson's Correlation Coefficient (PCC) can indicate the statistically similarity of rainfall characteristics among pixels over a seasonal scale.

3.1.2 Rule 1 of the WHU-SGCC method

In general, satellite precipitation estimations deviated from ground-based measurements, which were assumed to be the true values. Rule 1 aims to establish a regression model between each gauge historical observations and the corresponding CHIRP grid cell value. The regression relationship was derived by random forest regression (RFR) at each gauge station. RFR is a machine-learning algorithm for a predictive model with a large set of regression trees in which each tree in the ensemble is grown from a bootstrap (Johnson, 1998) sample drawn with replacement from the training set. The final prediction is obtained by combining the results of the prediction methods applied to each bootstrap sample (Genuer et al., 2017). The predicted value is calculated by the mean of all trees.

$$Y_o = f_{RFR}(Y_s) \quad (1)$$

where Y_o denotes each gauge historical observations and Y_s denotes the corresponding CHIRP grid cell values at C1 pixels, f_{RFR} is constructed from the time series Y_o (dependent variable) and Y_s (independent variable) by means of RFR. The number of decision trees was set to 500, which was determined by out-of-bag (OOB) error (Figure S1). The OOB error

reached the minimum value when the number of decision trees was less than 500.

The Rule 1 builds the statistical relationships between gauge observations and corresponding CHIRP grid cell values, which is the key idea in correcting the satellite-based precipitation estimations in the whole study area. As there are 30 gauge stations in the study area, 30 regression relationships at C1 pixels were derived from Rule1. The values of C1 pixels are not corrected in Rule 1, but interpolated in Rule 4.

3.1.3 Rule 2 of the WHU-SGCC method

It is reasonable to assume that there are some pixels that are statistically similar to the precipitation characteristics of C1 pixels in a certain spatial scope. Therefore, it is feasible to adjust the satellite estimation bias of C2 pixels by referring to the appropriate regression relationships at corresponding C1 pixels based on Rule 1.

First, the spatial scope in which pixels may have highly similar characteristics is established. Some studies indicate that geographical location, elevation and other terrain information influences the spatial distribution of rainfall, especially in mountainous areas with complex topography (Anders et al., 2006;Long and Singh, 2013). The size of the spatial range is an important parameter to distinguish spatial similarity and heterogeneity. In the WHU-SGCC method, the approach of fuzzy c-means (FCM) clustering was explored to determine the spatial range considered as each pixel's terrain factors including longitude, latitude, elevation, slope, aspect and curvature. FCM method was developed by J.C. Dunn in 1973 (Dunn, 1973), and improved in 1983 (Wang, 1983). It is an unsupervised fuzzy clustering method and the steps are as follows (Pessoa et al., 2018):

1) Choose the number of clusters *c*. The optimum number of clusters was determined by *L*(*c*) which was derived from the inter-distance and inner-distance of samples in Eq. (2). It is ensured that the distance between the same samples is smaller, while the distance between the different samples is larger.

$$L(c)=\frac{\sum_{i=1}^{c}\sum_{j=1}^{n} w_{ij}^{m} \| c_i-\bar{x} \|^2 /(c-1)}{\sum_{i=1}^{c}\sum_{j=1}^{n} w_{ij}^{m} \| x_j-c_i \|^2 /(n-c)} \tag{2}$$

In Eq. (2), the denominator is inner-distance and the molecular is inter-distance. The initial value of *c* is 1 and the maximum value of *c* is the number of gauge stations in this study area. The optimum number of clusters was optimized to maximize the L(*c*). For this reason, *c* value is conducted in the range of 1 to the number of gauge stations with an incremental interval value of 1 in this study.

2) Assign coefficients randomly to each data point x_i for the degree to which it belongs in the *i* th cluster $w_{ij}(x_i)$:

$$c_i^{(t)}=\frac{\sum_{j=1}^{n} w_{ij}^{m} x_i}{\sum_{j=1}^{n} w_{ij}^{m}} \quad (3), \qquad w_{ij}=\frac{1}{\sum_{k=1}^{c}\left(\frac{\| x_i-c_i \|}{\| x_i-c_k \|}\right)^{\frac{2}{m-1}}} \quad (4), \qquad \bar{x}=\frac{\sum_{i=1}^{c}\sum_{j=1}^{n} w_{ij}^{m} x_j}{n} \quad (5)$$

where *x* is a finite collection of *n* elements that will be partitioned into a collection of *c* fuzzy clusters, c_i is the centre of each cluster, *m* is the hyper-parameter that controls the level of cluster fuzziness and w_{ij} is the degree to which element x_i belongs to c_i and $\bar{x}$ is the center vector of collection. In Eq. (3), $c_j^{(t)}$ represents the cluster centre in iteration *t*. If the minimum improvement in objective function between two consecutive iterations satisfies the following equation, the algorithm

terminates in iteration *t* (Eq. (6)):

$$\|c_i^{(t)} - c_i^{(t+1)}\| < \varepsilon \tag{6}$$

3) Minimize the objective function F_c to achieve data partitioning.

$$F_c = \sum_{j=1}^{n}\sum_{i=1}^{c} w_{ij}^{m} \| x_j - c_i \|^2 \tag{7}$$

The results of FCM are the degree of membership of each pixel to the cluster centre as represented by numerical value. Pixels in each cluster have similar terrain features and precipitation characteristics.

Second, as mentioned above, the aim of Rule 2 is to derive an adjustment method for C2 pixels based on learning from Rule 1. With the establishment of a regression relationship between gauge observations and the corresponding CHIRP grid cell values of the C1 pixels by RFR method, the determination of C2 pixels follows a considerable procedure. With exception of the C1 pixels, the remaining pixels in each cluster represent potential C2 pixels called R pixels. Pearson's correlation coefficient (PCC) and *p*-values between the satellite estimations (CHIRP grid cell values) at R pixels and the C1 pixels are the criteria for final determination of C2 pixels. The PCC is defined as follows:

$$PCC_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{8}$$

where *n* is the number of samples, x_i and y_i are individual samples (CHIRP grid cell values at C1 and C2 pixels), $\bar{x}$ is the arithmetic mean of *x* calculated by $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y}$ is the arithmetic mean of *y* calculated by $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

The value range of the PCC is between -1 and +1. If there are no repeated data values, a perfect PCC of +1 or −1 occurs when each of the variables is a perfect monotone function of the other. However, if the value is close to zero, there is zero correlation. In addition, correlation is not only determined by the value of the correlation coefficient but also from the correlation test's *p*-value. The critical values for PCC and *p*-value are 0.5 and 0.05, thus a PCC value higher than 0.5 and a *p*-value lower than 0.05 indicate the data are significantly correlated (Zhang and Chen, 2016). Therefore, the final determination of C2 pixels must meet the following criteria:

$$|\mathrm{PCC}| \geq 0.5 \quad and \quad p < 0.05 \tag{9}$$

Each R pixel has *m* PCC and *p*-values (the number of C1 pixels in the cluster), and the subset of C2 pixels is identified by excluding the data that failed the correlation test and retaining both the data with a maximum PCC of at least 0.5 and a *p*-value lower than 0.05, and the corresponding index of C1 pixels. The selected C2 pixels are statistically similar to the precipitation characteristics of corresponding C1 pixels in their defined spatial scope.

After identifying the C2 pixels and their corresponding C1 pixels, the adjustment method for C2 pixels is derived from the regression model for the C1 pixels.

$$C2_{as} = f_{RFRc}(Y_s) \tag{10}$$

where $C2_{as}$ is the adjusted satellite precipitation estimate and Y_s is the CHIRP grid cell value at the C2 pixels, and f_{RFRc} is the f_{RFR} of corresponding C1 pixel.

3.1.4 Rule 3 of the WHU-SGCC method

Recognizing that precipitation has a spatial distribution, the assumption that C3 pixels are statistically similar to the precipitation characteristics of C2 pixels is adopted to establish the adjustment method for C3 pixels.

First, the determination of C3 pixels in each spatial cluster is based on the selection of C2 pixels. The satellite-based estimation values at the remaining pixels with exception of the C1and C2 pixels are used to calculate the PCC and *p*-value with the satellite-based estimation values at the C2 pixels in the same cluster. The results of each pixel's *k* PCC and *p*-value (the number of C2 pixels in the cluster) are evaluated based on the correlation test (Eq. (9)) that the pixels with a maximum PCC of is at least 0.5, and then the corresponding index of C2 pixels are retained. The selected pixels called C3 pixels, which are statistically similar to the precipitation characteristics of the corresponding C2 pixels in the defined spatial scope.

After identifying the C3 pixels, a method for merging the CHIRP grid cell values at C3 pixels (Y_s) and the target reference values of $C2_{as}$ at the corresponding C2 pixels is applied to estimate the adjusted precipitation values at C3 pixels. This method combines Y_s and $C2_{as}$ values in one variable, as shown in Eq. (11):

$$w_i = \frac{C2_{as_i} + \lambda}{Y_{s_i} + \lambda} \quad i=1,\ldots,\ n \tag{11}$$

where λ is a positive constant set to 10 mm (Sokol, 2003), $C2_{as}$ is the adjusted precipitation values at the C2 pixels, Y_{s_i} is extracted from the CHIRP grid cell values at the corresponding location of the C2 pixel, and *n* is the number of C2 pixels in each spatial cluster.

Each *w* of the C3 pixels is assigned the same value as the corresponding C2 pixel. Therefore, the values of C3 pixels are derived from Eq. (12):

$$C3_{as} = \max(w \times (Y_s + \lambda) - \lambda, 0) \tag{12}$$

where $C3_{as}$ is the adjusted target precipitation value at one C3 pixel and Y_s is the corresponding CHIRP grid cell value. To avoid precipitation estimates below 0, Eq. (12) sets these negative values to 0.

3.1.5 Rule 4 of the WHU-SGCC method

Excluding the C1C2and C3 pixels, the number of remaining pixels, called C4 pixels which are adjusted by Inverse Distance Weighted (IDW). IDW is based on the concept of the first law of geography from 1970. It was defined as e*verything is related to everything else, but near things are more related than distant things*. Therefore, the attribute value of an unsampled point is the weighted average of known values within the neighbourhood and the distance weighting can be determined by IDW (Lu and Wong, 2008). In Rule 4, IDW is used to interpolate the unknown spatial precipitation data from the C2 and C3 pixels adjusted precipitation values. The IDW formulas are given as Eq. (13) and Eq. (14).

$$R_{as} = \sum_{i=1}^{n} w_i R_i \tag{13}$$

$$w_i = \frac{d_i^{-\alpha}}{\sum_{i=1}^{n} d_i^{-\alpha}} \quad \text{with} \quad \sum_{i=1}^{n} w_i = 1 \tag{14}$$

where R_{as} is the unknown spatial precipitation data, R_i is the adjusted precipitation values at C2 and C3 pixels, *n* is the number of C2 and C3 pixels, d_i is the distance from each C2 or C3 pixel to be unknown grid cell, α is the power which is generally specified as a geometric form for the weight. Several researches (e.g., Simanton and Osborn 1980; Tung 1983) have experimented with variations in a power, the small α tends to estimate values with the averages of sampled grids in the neighbourhood, while large α tends to give larger weights to the nearest points and increasingly down-weights points farther away (Chen and Liu, 2012;Lu and Wong, 2008). The value of α has an influence on the spatial distribution of information from precipitation observations. For this reason, α value is conducted in the range of 0.1 to three (0.1, 0.3, 0.5, 1.0, 1.5, 2.0,

2.5 and 3.0) in this study.

It is noted that the unknown spatial precipitation data including C1 and C4 pixels, because C1 pixels values were not adjusted in Rule 1.

In the end, after applying these four rules, we obtained complete daily adjusted regional precipitation maps for the summer (JJA) over the Jinsha River basin.

3.2 Accuracy assessment

The performance of the WHU-SGCC adjusted precipitation estimates was evaluated by nine statistical indicators: Pearson's correlation coefficient (PCC), root mean square error (RMSE), mean absolute error (MAE), relative bias (BIAS), the Nash-Sutcliffe efficiency coefficient (NSE), probability of detection (POD) and false alarm ratio (FAR) and critical success index (CSI). PCC, RMSE, MAE and BIAS were used to evaluate how well the WHU-SGCC method adjusted satellite estimation bias, while POD, FAR and CSI were used to evaluate the precipitation event predictions (Su et al., 2011). PCC measures strength of the correlation relationship between the satellite estimations and observations. MAE represents the average magnitude of error estimations, considering both systematic and random errors. The NSE (Nash and Sutcliffe, 1970) determines the relative magnitude of the variance of the residuals compared to the variance of the observations, bounded by minus infinity to 1. A negative value indicates a poor precipitation estimate and the value of an optimal estimate is equal to 1. BIAS measures the mean tendency of the estimated precipitation to be larger (positive values) or smaller (negative values) than the observed precipitation, with an optimal value of 0.

POD, also known as the hit rate, represents the probability of rainfall detection. FAR is defined as the ratio of the false detection of rainfall to the total number of rainfall events. All of the accuracy assessment indices are shown in Table 3.

Table 3 Accuracy assessment indices.

Accuracy assessment Index	Unit	Formula	Range	Optimal value
Pearson's Correlation Coefficient (*SCC*)	NA	$\mathrm{SCC}=\frac{\sum_{i=1}^{n}(Y_{oi}-\bar{Y}_o)(C_i-\bar{C})}{\sqrt{\sum_{i=1}^{n}(Y_{oi}-\bar{Y}_o)^2}\cdot\sqrt{\sum_{i=1}^{n}(C_i-\bar{C})^2}}$	[-1,1]	1
Root Mean Square Error (RMSE)	Mm	$\mathrm{RMSE}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(C_i-Y_{oi})^2}$	[0,+∞)	0
Mean Absolute Error (MAE)	Mm	$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\lvert C_i-Y_{oi}\rvert$	[0, +∞)	0
Relative Bias (BIAS)	NA	$\mathrm{BIAS}=\frac{\sum_{i=1}^{n}(C_i-Y_{oi})}{\sum_{i=1}^{n}Y_{oi}}$	(-∞, +∞)	0
Nash-Sutcliffe Efficiency Coefficient (*NSE*)	NA	$\mathrm{NSE}=1-\frac{\sum_{i=1}^{N}(C_i-Y_{oi})^2}{\sum_{i=1}^{N}(C_i-\bar{Y}_o)^2}$	(-∞,1]	1
Probability of Detection (POD)	NA	*POD=H/(H+M)*	[0,1]	1
False Alarm Ratio (FAR)	NA	*FAR=F/(H+F)*	[0,1]	0
Critical Success Index (CSI)	NA	*CSI=H/(H+M+F)*	[0,1]	1

Note: Y_{oi} is the observation data and C_i is the adjusted value using the WHU-SGCC method for test sample pixel; $\bar{Y}_o$ is the arithmetic mean of Y_o and is given by $\bar{Y}_o=\frac{1}{n}\sum_{i=1}^{n}Y_{oi}$; $\bar{C}$ is the arithmetic mean of *C* and is given by $\bar{C}=\frac{1}{n}\sum_{i=1}^{n}C_i$;

H represents the number of both observed and estimated precipitation events (successfully forecasted), and F is the number of false alarms when observed precipitation was below the threshold and estimated precipitation was above threshold (false alarms). M is the number of events in which the estimated precipitation was below the threshold and observed precipitation was above the threshold (missed forecasts). POD and FAR values are dimensionless numbers ranging from 0 to 1. The precipitation threshold (event/no event) was set to 0.1 mm/day.

4 Results and Discussion

There were 18482 daily pixels adjusted by blending satellite estimations (CHIRP) and observations (gauge stations and gridded points) using the WHU-SGCC approach during the JJA from 1990 to 2014. The number of pixels adjusted by each rule in the WHU-SGCC method is shown in Table 4. The number of C1 pixels was the number of training gauge stations accounting 0.16% of the total pixels (18482) inside the basin. Due to the leave-one-out cross validation step, the different training samples will have the different number of C2, C3 and C4 pixels respectively inside the Jinsha River Basin. The number of C4 pixels was approximately 10822 with the percentage around 60%, the number of C3 pixels was approximately 4331 with the percentage ranging from 21.72% to 24.40%, and the number of C2 pixels was approximately 3300 with the percentage ranging from 15.59% to 18.36%.

Table 4 The number of each class pixels adjusted by each rule using the WHU-SGCC method inside the Jinsha River Basin.

Validation gauge station	C1 Pixels (%)	C2 Pixels (%)	C3 Pixels (%)	C4 Pixels (%)
52908	29 (0.16%)	3066 (16.59%)	4224 (22.85%)	11163 (60.40%)
56004	29 (0.16%)	2882 (15.59%)	4111 (22.24%)	11460 (62.01%)
56021	29 (0.16%)	3311 (17.91%)	4510 (24.40%)	10632 (57.53%)
56029	29 (0.16%)	3338 (18.06%)	4447 (24.06%)	10668 (57.72%)
56034	29 (0.16%)	3300 (17.86%)	4427 (23.95%)	10726 (58.03%)
56038	29 (0.16%)	3209 (17.36%)	4014 (21.72%)	11230 (60.76%)
56144	29 (0.16%)	3347 (18.11%)	4442 (24.03%)	10664 (57.70%)
56146	29 (0.16%)	3183 (17.22%)	4480 (24.24%)	10790 (58.38%)
56152	29 (0.16%)	3173 (17.17%)	4176 (22.59%)	11104 (60.08%)
56167	29 (0.16%)	3362 (18.19%)	4346 (23.51%)	10745 (58.14%)
56247	29 (0.16%)	3385 (18.32%)	4416 (23.89%)	10652 (57.63%)
56251	29 (0.16%)	3301 (17.86%)	4348 (23.53%)	10804 (58.46%)
56257	29 (0.16%)	3313 (17.93%)	4043 (21.88%)	11097 (60.04%)
56357	29 (0.16%)	3352 (18.14%)	4390 (23.75%)	10711 (57.95%)
56374	29 (0.16%)	3341 (18.08%)	4294 (23.23%)	10818 (58.53%)
56459	29 (0.16%)	3345 (18.10%)	4334 (23.45%)	10774 (58.29%)
56462	29 (0.16%)	3380 (18.29%)	4377 (23.68%)	10696 (57.87%)
56475	29 (0.16%)	3345 (18.10%)	4344 (23.50%)	10764 (58.24%)
56479	29 (0.16%)	3305 (17.88%)	4212 (22.79%)	10936 (59.17%)
56485	29 (0.16%)	3393 (18.36%)	4419 (23.91%)	10641 (57.57%)
56543	29 (0.16%)	3373 (18.25%)	4384 (23.72%)	10696 (57.87%)
56565	29 (0.16%)	3241 (17.54%)	4450 (24.08%)	10762 (58.23%)
56571	29 (0.16%)	3306 (17.89%)	4263 (23.07%)	10884 (58.89%)
56586	29 (0.16%)	3387 (18.33%)	4434 (23.99%)	10632 (57.53%)
56651	29 (0.16%)	3340 (18.07%)	4432 (23.98%)	10681 (57.79%)
56664	29 (0.16%)	3368 (18.22%)	4262 (23.06%)	10823 (58.56%)
56666	29 (0.16%)	3323 (17.98%)	4431 (23.97%)	10699 (57.89%)
56671	29 (0.16%)	3356 (18.16%)	4367 (23.63%)	10730 (58.06%)
56684	29 (0.16%)	3335 (18.04%)	4278 (23.15%)	10840 (58.65%)
56778	29 (0.16%)	3347 (18.11%)	4277 (23.14%)	10829 (58.59%)

4.1 Spatial Clustering from the FCM method

To adjust the pixels other than for the gauged stations, the pixels statistically similar to the C1 were selected. According to Rule 2, C2 pixels were identified in a spatial scope defined by the FCM method. In the following experiments of Rule 2, we set the parameters $m = 2, \varepsilon = 0.00001$ and the maximum number of iterations was set 1000 (an enough large value with the consideration of the algorithm efficiency). In order to determine the optimal numbers of clusters, *c* value was conducted in the

range from 1 to 30 with an incremental interval value of 1 in this study. During the running of FCM approach, the values of $L(c)$ were shown in Fig 4. Figure 4 shows the optimum number of clusters was 22, with the number of iterations was 690 less than the maximum number of iterations.

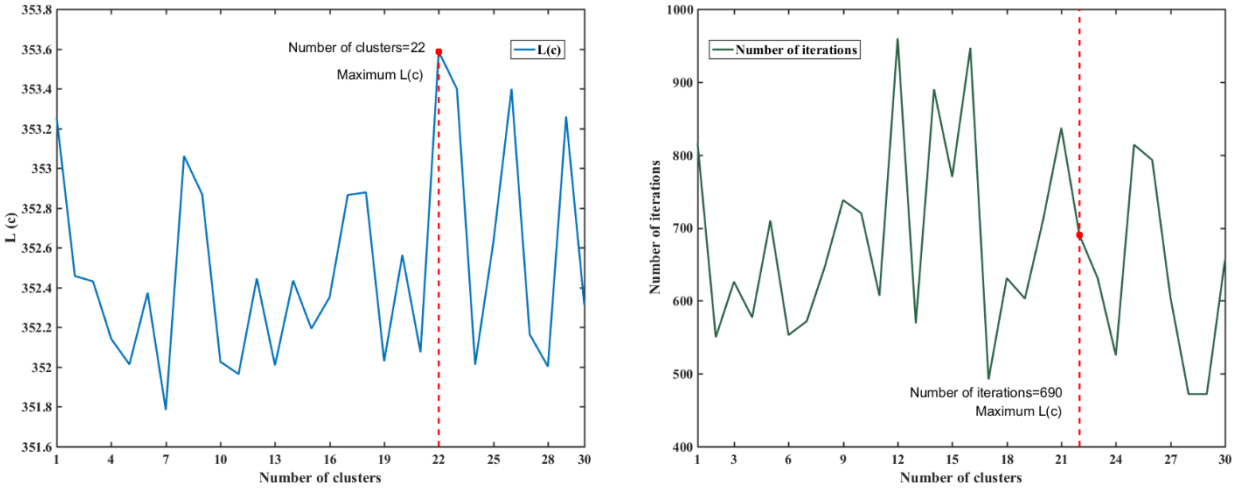

Figure 4 The optimum number of clusters determined by the maximum $L(c)$ with the iterative process.

Therefore, the number of clusters was set to 22 and the number of iterations was still set to 1000 for fully operations by means of FCM. The spatial clusters results with consideration of the terrain factors was shown in Fig. 5. Overall, the spatial results of FCM have many of the same characteristics as spatial areas defined by terrain changes, especially with respect to slope and runoff directions, which may influence regional rainfall to some extent.

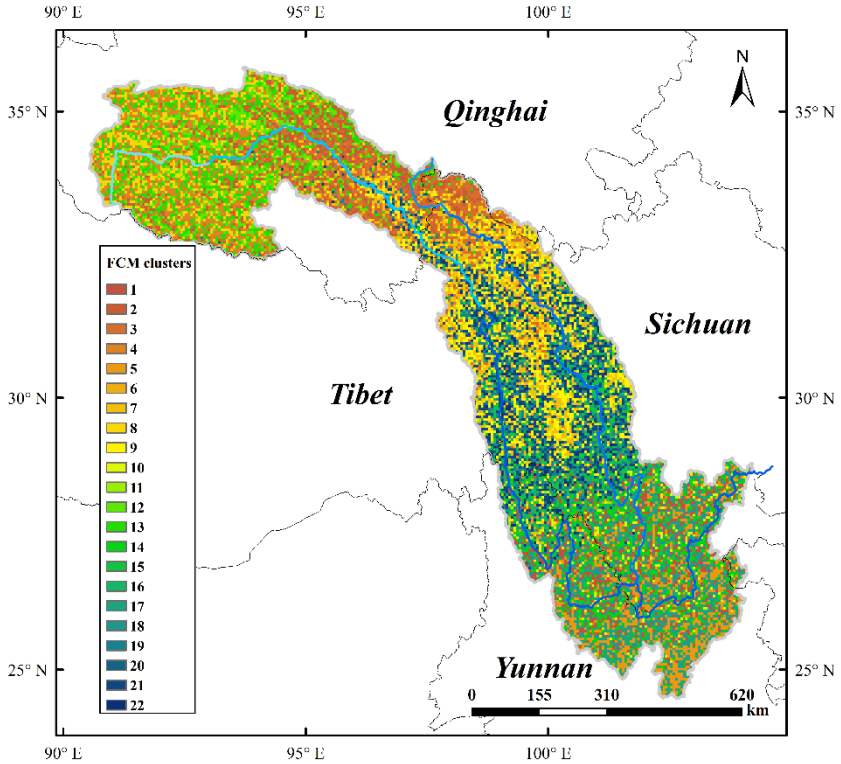

Figure 5 Spatial clustering as defined by FCM for the Jinsha River Basin.

4.2 Model performance based on overall accuracy evaluations

To test the performance of the WHU-SGCC method for precipitation estimates, the statistical analyses of PCC, RMSE, BAE, BIAS, NSE, POD, FAR, and CSI were calculated and are presented in Table 5 (The results were derived from the α =0.1 for

IDW in Rule 4 with the comparison with other values in RMSEs). Compared with the satellite images of CHIRP and CHIRPS, the results of the WHU-SGCC provide the greatest improvements for regional daily precipitation estimates over the Jinsha River Basin during JJA from 1990 to 2014. After bias adjustment of the WHU-SGCC, PCC was improved by 3.34% and 31.81% compared to CHIRP and CHIRPS, respectively. Meanwhile, the RMSE and MAE of the WHU-SGCC decreased by 6.91% and 6.59% compared to CHIRP, and by 22.71% and 22.15% compared to CHIRPS. Although, the absolute value of BIAS of WHU-SGCC was no significant improvement than CHIRP and slightly higher than CHIRPS, all of the values were approximately to 0. This results of BIAS indicates that the all three kinds of data were much the same on the performance. Nevertheless, the NSE of the WHU-SGCC reached -0.0137, an increase of 93.33% and 98.32% compared to CHIRP and CHIRPS, respectively. The NSE of WHU-SGCC was still far less than 1, but it was improved to be zero that indicates the adjusted results are close to the average level of the rain gauge observations, while the NSEs of CHIRP and CHIRPS were much worse. It is noted that the POD of WHU-SGCC was approximate to 1, better than CHIRP and CHIRPS, and the FAR of WHU-SGCC was 0.11, lower than CHIRP and CHIRPS, which represents the better ability on precipitation event predictions of the WHU-SGCC.

Table 5 Overall accuracy assessment during the JJA from 1990 to 2014.

Statistic	WHU-SGCC	CHIRP	CHIRPS
PCC	0.2536	0.2454	0.1924
RMSE	8.7608	9.4108	11.3354
MAE	5.4564	5.8415	7.0088
BIAS	-0.0167	-0.0443	-0.0134
NSE	-0.0139	-0.2083	-0.8293
POD	0.9932	0.9578	0.4351
FAR	0.1146	0.2323	0.1601
CSI	0.8799	0.7405	0.4010

4.3 Model performance based on the spatial distributions

The spatial distributions of the statistical comparisons between observations and WHU-SGCC precipitation estimations are shown in Fig. 6. The variation of PCC as seen in Fig. 6 (a) shows that low correlations are observed in areas with lower elevation, particularly in the southern Jinsha River Basin where there is higher precipitation and a greater density of rain gauges. This result is in contrast to the result in (Rivera et al., 2018), because of the few days for heavy rains in this study area. The higher correlations noted over the north central area of the river basin are in a drier region with complex terrain and sparse rain gauges. With respect to the spatial distribution of RMSE, Fig. 6 (b) indicates that smaller errors are scattered in the northwest area of the river basin, with values lower than 5 mm, while the highest errors, which are over 10 mm, are located over the border between the lower reaches of the Jinsha Jiang River and the river basin. All the values of MAE are below 10 mm and the spatial behaviour is similar to that of the RMSE. Fig. 6 (c) shows that the lower MAE values are located over the mountainous region southwest of Qinghai and west of Sichuan, with values below 6 mm. The spatial distribution of the BIAS indicates that the WHU-SGCC has good agreement with the observations, with the most values ranging from -10%~10%. All the spatial distribution statistics indicate that the statistical relationships established during the process of the WHU-SGCC method is susceptible to the mode values of the rain gauge stations data. Although the average annual precipitation in the southern Jinsha River Basin was more than 600 mm (Fig.2), the days of light rain were still in the great percentage that limited the performance over the south area, while there were sufficient data with similar precipitation features for WHU-SGCC over the north data. Nevertheless, WHU-SGCC approach is still effective in adjusting the satellite biases by blending with the observations, particularly in the complicated mountainous region where there are higher PCC corresponding to lower values of RMSE, MAE and BIAS.

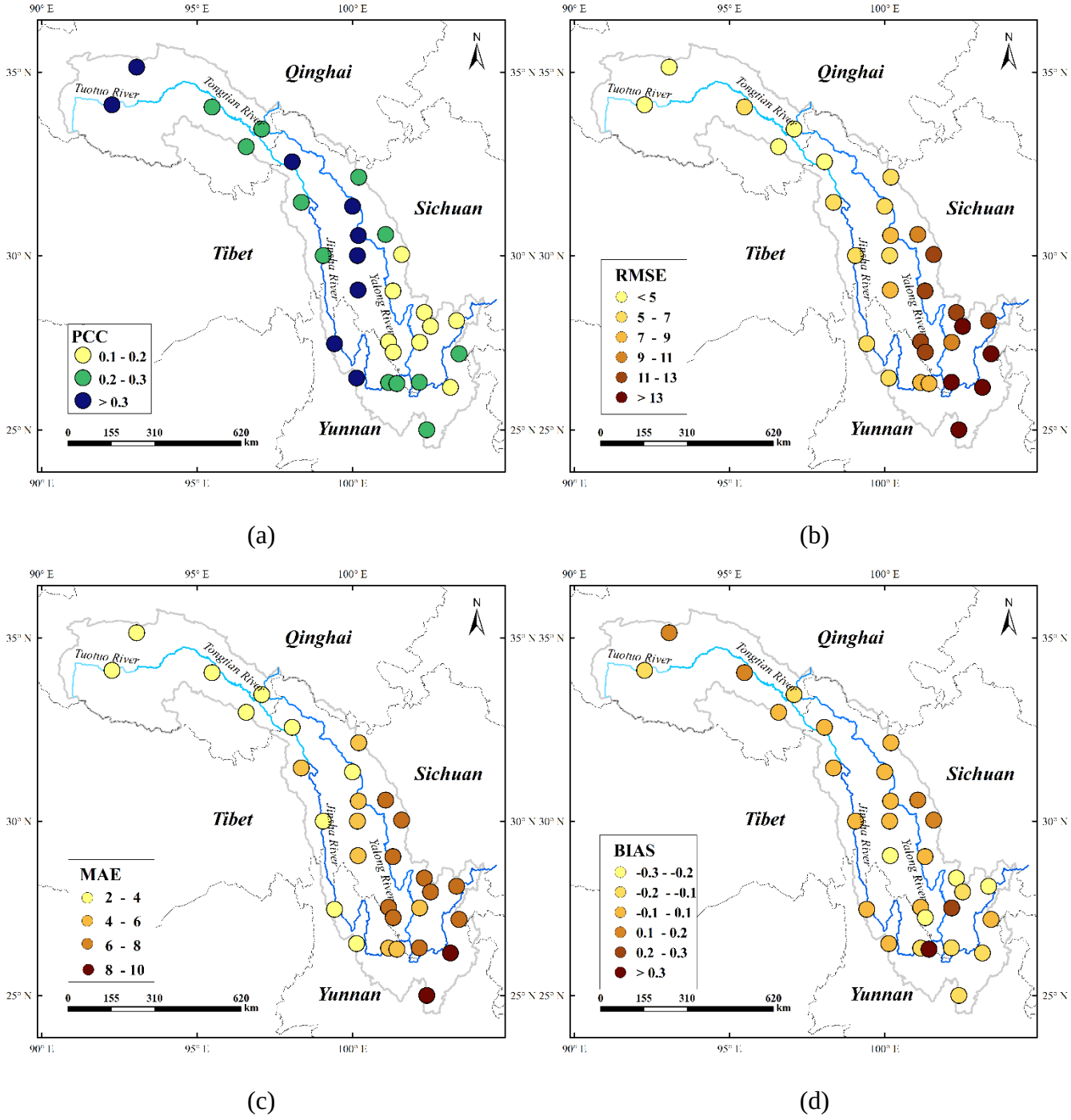

Figure 6 Spatial distribution of the statistical analyses of the overall agreement between observations and the WHU-SGCC estimations on leave-one-out cross validation for JJA from 1990 to 2014: a) Pearson's correlation coefficient, b) root mean square error c) mean absolute error, and d) relative bias.

4.4 Model performance for rain events

To measure the WHU-SGCC performance for different rain events, the daily precipitation thresholds of 0.1, 10, 25, and50mm were considered, and the result is shown in Table 6 and Table 7. The days of each class of rain events at the validation gauge station during JJA from 1990 to 2014 were shown in Table 5. The major rain events inside the Jinsha River Basin were light rain (0.1-10 mm), accounting for 54.76% of the total days (the average percentage of rain event days in its total days at each gauge station), while the days with daily precipitation over the 50 mm was least, only accounting for 0.72%. And the percentage of the daily precipitation of <0.1, 10-25, and 25-50 mm were 26.89%, 14.01% and 3.62% respectively. The result indicated that the average daily precipitation was less than 10 mm, though in the summer seasons during the multi-year. As well as, the spatial distribution of precipitation was also uneven, with an increase from north to south.

Table 6 The days of each class of rain events at the validation gauge station during JJA from 1990 to 2014 inside the Jinsha River

Validation gauge station \ Rain event (mm)	<0.1	[0.1,10)	[10,25)	[25,50)	>=50	Total days
52908	637	1186	134	9	0	1966
56004	628	1243	128	3	0	2002
56021	535	1305	166	9	0	2015
56029	556	1328	190	5	0	2079
56034	558	1351	185	17	0	2111
56038	459	1329	222	16	0	2026
56144	562	1153	321	25	0	2061
56146	467	1278	267	19	0	2031
56152	466	1255	307	35	1	2064
56167	565	1234	278	20	0	2097
56247	591	1089	246	34	0	1960
56251	466	1247	320	30	0	2063
56257	336	1212	429	59	0	2036
56357	313	1247	373	63	1	1997
56374	393	1191	351	47	0	1982
56459	487	1080	377	102	13	2059
56462	185	1315	430	86	2	2018
56475	544	983	352	148	20	2047
56479	667	931	298	156	28	2080
56485	588	905	232	100	37	1862
56543	332	1200	289	41	1	1863
56565	526	1020	349	120	13	2028
56571	674	819	301	159	49	2002
56586	730	950	223	79	9	1991
56651	402	1056	391	137	31	2017
56664	727	797	306	166	56	2052
56666	858	791	226	128	44	2047
56671	616	886	289	148	70	2009
56684	768	899	246	114	19	2046
56778	682	930	274	119	43	2048

In terms of performance with respect to different daily rain events, the WHU-SGCC approach had the lowest error, as indicated by RMSE, MAE and BIAS for events with total rainfall lower than 25 mm, but WHU-SGCC performance for total rainfall higher than 25 mm did not improve compared to CHIRP and CHIRPS (Table 6). This negative performance on the total rainfall higher than 25 mm was probably caused by the precipitation conditions inside the Jinsha River Basin (Table 5). The average daily precipitation was less than 10 mm inside the basin, during the multi-year summer seasons, which provided a large amount of rain gauge stations data with the values lower than 10 mm, that caused a significantly impact on the statistical relationships establishment for WHU-SGCC. In hence, the approach of WHU-SGCC is applicable for the detection of rainfall events over the Jinsha River Basin, with the precipitation less than 10 mm, or even than 25mm. Due to the 4.34% of summer days with the daily precipitation over the 25 mm, the performance of WHU-SGCC on these rain events was poorer than the results of CHIRP and CHRPS.

Table 7 Accuracy assessment on liquid precipitation events during the JJA from 1990 to 2014

	RMSE			MAE			BIAS		
Rain Event	WHU-SGCC	CHIRP	CHIRPS	WHU-SGCC	CHIRP	CHIRPS	WHU-SGCC	CHIRP	CHIRPS
<0.1	4.7253	5.0802	7.1643	2.5927	2.9562	2.9145	/	/	/
[0.1,10)	4.1661	6.8684	9.6022	3.9885	4.5534	6.2462	0.8021	1.4435	1.9842
[10,25)	10.4281	11.0848	13.4427	9.2722	9.6866	11.5909	-0.5762	0.6342	0.7559
[25,50)	25.7494	24.5600	25.4975	24.8386	23.0967	23.4927	-0.7784	0.7250	0.7388
≥50	56.6072	54.5037	52.7875	54.4168	52.1557	49.4318	-0.8861	0.8297	0.7852

5 Conclusions

This study provided a novel approach in the WHU-SGCC method for merging daily satellite-based precipitation estimates with observations. A case study of Jinsha River Basin was conducted to verify the effectiveness of the WHU-SGCC approach during the JJA from 1990 to 2014, and the adjusted precipitation estimates were compared to CHIRP and CHIRPS. WHU-SGCC aims to reduce systematic and random errors in CHIRP over a region that has complicated mountainous terrain and sparse rain gauges. To the best of the authors' knowledge, this study is the first to use daily CHIRP and CHIRPS data in this area.

According to our findings, the following conclusions can be drawn: (1) The WHU-SGCC method is effective for the adjustment of precipitation biases from point to surface. The precipitation estimated by the WHU-SGCC method can achieve greater accuracy, which was evaluated with PCC, RMSE, MAE, BIAS, NSE, POD, FAR and CSI. Particularly, the NSE statistic was improved by 93.33% and 98.32% compared to CHIRP and CHIRPS, respectively, and all measured errors were reduced except the BIAS with no significant improvement, but approximately to 0. The results show that compared to CHIRPS, the WHU-SGCC approach can achieve substantial improvements in precipitation estimate accuracy. (2) Moreover, the spatial distribution of precipitation estimate accuracy derived from the WHU-SGCC method is related to the complexity of the topography. These random errors over the lower evaluations and the large size of light precipitation events with short duration rainstorms in the region resulted in a limited improvement in accuracy, with PCC values less than 0.3, However, higher PCC and lower errors were observed over the north central area of the river basin, which is a drier region with complex terrain and sparse rain gauges. All the spatial distribution statistics indicate that the WHU-SGCC method is superior for adjustment of satellite biases by blending with the observations over the complex terrain region. (3) The WHU-SGCC validations for daily rain events confirmed that the model was effective in the detection of precipitation events less than 25 mm due to the less average annual precipitation inside the Jinsha River Basin. According to the comparison, the WHU-SGCC approach achieves error reductions for the RMSE, MAE and BIAS statistics for rain events within the range of 1-25 mm. Specifically, compared with CHIRP, the RMSE value was reduced by approximately by 5.92%-39.44%, the MAE value by 4.28%-12.41%, and the absolute BIAS value by 9.15%-44.43%; compared with CHIRPS, the RMSE and MAE values were reduced by 11.04%-56.61%, and the absolute BIAS value by 23.77%-59.58%.

Therefore, the WHU-SGCC approach can help adjust the biases of daily satellite-based precipitation estimates over the Jinsha River Basin, the complicated mountainous terrains with sparse rain gauges, particularly for precipitation events with less than 25 mm in summer. This approach is a promising tool to monitor monsoon precipitation over the Jinsha River Basin, considering the spatial correlation and historical precipitation characteristics between raster pixels located in regions with similar topographic features. Future development of the WHU-SGCC approach will focus on the following three aspects: 1) improvement of the adjusted precipitation quality by blending in different rain events and applying in all seasons; 2) introduction of more topographic factors and climatic factors to achieve a more accurate spatial distribution of precipitation;

and 3) investigation of the performance over other areas and on the particular hydrological case to validate the WHU-SGCC.

6 Data availability

All the resulting dataset derived from the WHU-SGCC approach is available on PANGAEA, with the following DOI: https://doi.pangaea.de/10.1594/PANGAEA.896615 (Shen et al., 2018). The high-resolution (0.05°) daily precipitation estimation data over the Jinsha River Basin in the summer from 1990 to 2014 can be downloaded in TIFF format.

7 Acknowledgments

This work was supported by the National Natural Science Foundation of China program (no. 41771422), the Nature Science Foundation of Hubei Province (no. 2017CFB616), the fundamental research funds for the central universities (no. 2042017kf0211), and the LIESMARS Special Research Funding.

The authors would like to thank data support: the Climate Hazards Group at the University of California, Santa Barbara, for providing CHIRP and CHIRPS datasets (http://chg.ucsb.edu/data/), and the National Climate Center (NCC) of the China Meteorological Administration (CMA) for providing the daily rain gauged observations and gridded precipitation observations (http://data.cma.cn/). The authors also thank the PANGAEA Data Publisher for Earth & Environmental Science platform for providing the storage to disseminate the data generated in this experiment.

The authors are grateful for the editor and anonymous reviewers for their useful suggestions that clearly improved this paper.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present), J. Hydrometeorol., 4, 1147-1167, doi:10.1175/1525-7541(2003)004<1147:tvgpcp>2.0.co;2, 2003.

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., and Amitai, E.: Evaluation of satellite-retrieved extreme precipitation rates across the central United States, J. Geophys. Res.-Atmos., 116, 11, doi:10.1029/2010jd014741, 2011.

Agutu, N. O., Awange, J. L., Zerihun, A., Ndehedehe, C. E., Kuhn, M., and Fukuda, Y.: Assessing multi-satellite remote sensing, reanalysis, and land surface models' products in characterizing agricultural drought in East Africa, Remote Sens. Environ., 194, 287-302, doi:10.1016/j.rse.2017.03.041, 2017.

Ali, H., and Mishra, V.: Contrasting response of rainfall extremes to increase in surface air and dewpoint temperatures at urban locations in India, Sci Rep, 7, 1228, doi:10.1038/s41598-017-01306-1, 2017.

Anders, A. M., Roe, G. H., Hallet, B., Montgomery, D. R., Finnegan, N. J., and Putkonen, J.: Spatial patterns of precipitation and topography in the Himalaya, Tectonics, Climate, and Landscape Evolution, 398, 39-53, doi:10.1130/2006.2398(03), 2006.

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

Bai, L., Shi, C. X., Li, L. H., Yang, Y. F., and Wu, J.: Accuracy of CHIRPS Satellite-Rainfall Products over Mainland China, Remote Sens., 10, 28, doi:10.3390/rs10030362, 2018.

Behrangi, A., Andreadis, K., Fisher, J. B., Turk, F. J., Granger, S., Painter, T., and Das, N.: Satellite-Based Precipitation Estimation and Its Application for Streamflow Prediction over Mountainous Western US Basins, J. Appl. Meteorol.

Climatol., 53, 2823-2842, doi:10.1175/jamc-d-14-0056.1, 2014.
Behrangi, A., Hsu, K. L., Imam, B., Sorooshian, S., Huffman, G. J., and Kuligowski, R. J.: PERSIANN-MSA: A Precipitation Estimation Method from Satellite-Based Multispectral Analysis, J. Hydrometeorol., 10, 1414-1429, 10.1175/2009jhm1139.1, 2009.
Berndt, C., Rabiei, E., and Haberlandt, U.: Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios, Journal of Hydrology, 508, 88-101, doi:10.1016/j.jhydrol.2013.10.028, 2014.
Cattani, E., Merino, A., Guijarro, J. A., and Levizzani, V.: East Africa Rainfall Trends and Variability 1983-2015 Using Three Long-Term Satellite Products, Remote Sens., 10, 26, doi:10.3390/rs10060931, 2018.
Chen, F. W., and Liu, C. W.: Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan, Paddy Water Environ., 10, 209-222, 10.1007/s10333-012-0319-1, 2012.
Chen, J., Brissette, F. P., Chaumont, D., and Braun, M.: Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America, Water Resources Research, 49, 4187-4205, doi:10.1002/wrcr.20331, 2013.
Derin, Y., Anagnostou, E., Berne, A., Borga, M., Boudevillain, B., Buytaert, W., Chang, C.-H., Delrieu, G., Hong, Y., Hsu, Y. C., Lavado-Casimiro, W., Manz, B., Moges, S., Nikolopoulos, E. I., Sahlu, D., Salerno, F., Rodriguez-Sanchez, J.-P., Vergara, H. J., and Yilmaz, K. K.: Multiregional Satellite Precipitation Products Evaluation over Complex Terrain, J. Hydrometeorol., 17, 1817-1836, 10.1175/jhm-d-15-0197.1, 2016.
Duan, Z., Liu, J. Z., Tuo, Y., Chiogna, G., and Disse, M.: Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales, Sci. Total Environ., 573, 1536-1553, doi:10.1016/j.scitotenv.2016.08.213, 2016.
Dunn, J. C.: A fuzzy relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, Journal of Cybernetics, 3, 32-57, 1973.
Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., and Vose, R. S.: Comprehensive Automated Quality Assurance of Daily Surface Observations, J. Appl. Meteorol. Climatol., 49, 1615-1633, doi:10.1175/2010jamc2375.1, 2010.
Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes, Sci. Data, 2, 21, doi:10.1038/sdata.2015.66, 2015 a.
Funk, C., Verdin, A., Michaelsen, J., Peterson, P., Pedreros, D., and Husak, G.: A global satellite-assisted precipitation climatology, Earth Syst. Sci. Data, 7, 275-287, 10.5194/essd-7-275-2015, 2015 b.
Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Rowland, J., Bo, E., Husak, G. J., Michaelsen, J. C., and Verdin, A. P.: A Quasi-Global Precipitation Time Series for Drought Monitoring Data Series 832, Usgs Professional Paper, Data Series, 2014.
Genuer, R., Poggi, J. M., Tuleau-Malot, C., and Villa-Vialaneix, N.: Random Forests for Big Data, Big Data Res., 9, 28-46, doi:10.1016/j.bdr.2017.07.003, 2017.
Guo, P.-T., Li, M.-F., Luo, W., Tang, Q.-F., Liu, Z.-W., and Lin, Z.-M.: Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach, Geoderma, 237-238, 49-59, 10.1016/j.geoderma.2014.08.009, 2015.
Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B., and Schneider, U.: The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset, Bulletin of the American Meteorological Society, 78, 5-20, doi:10.1175/1520-0477(1997)078<0005:tgpcpg>2.0.co;2, 1997.
Johnson, R. W.: An Introduction to the Bootstrap, Chapman & Hall/CRC Press, 49-54 pp., 1998.
Katsanos, D., Retalis, A., and Michaelides, S.: Validation of a high-resolution precipitation database (CHIRPS) over Cyprus for a 30-year period, Atmospheric Research, 169, 459-464, doi:10.1016/j.atmosres.2015.05.015, 2016a.

Katsanos, D., Retalis, A., Tymvios, F., and Michaelides, S.: Analysis of precipitation extremes based on satellite (CHIRPS) and in situ dataset over Cyprus, Natural Hazards, 83, 53-63, doi:10.1007/s11069-016-2335-8, 2016b.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J.: The Tropical Rainfall Measuring Mission (TRMM) sensor package, J. Atmos. Ocean. Technol., 15, 809-817, doi:10.1175/1520-0426(1998)015<0809:ttrmmt>2.0.co;2, 1998.

Long, D., and Singh, V. P.: Assessing the impact of end- member selection on the accuracy of satellite- based spatial variability models for actual evapotranspiration estimation, Water Resources Research, 49, 2601-2618, doi:10.1002/wrcr.20208, 2013.

Lu, G. Y., and Wong, D. W.: An adaptive inverse-distance weighting spatial interpolation technique, Comput. Geosci., 34, 1044-1055, 10.1016/j.cageo.2007.07.010, 2008.Maggioni, V., and Massari, C.: on the performance of satellite precipitation products in riverine flood modeling: A review, Journal of Hydrology, 558, 214-224, 10.1016/j.jhydrol.2018.01.039, 2018.

Mahmoud, M. T., Al-Zahrani, M. A., and Sharif, H. O.: Assessment of global precipitation measurement satellite products over Saudi Arabia, Journal of Hydrology, 559, 1-12, 10.1016/j.jhydrol.2018.02.015, 2018.

Mahmoud, M. T., Hamouda, M. A., and Mohamed, M. M.: Spatiotemporal evaluation of the GPM satellite precipitation products over the United Arab Emirates, Atmospheric Research, 219, 200-212, 10.1016/j.atmosres.2018.12.029, 2019.

Martens, B., Cabus, P., De Jongh, I., and Verhoest, N. E. C.: Merging weather radar observations with ground-based measurements of rainfall using an adaptive multiquadric surface fitting algorithm, Journal of Hydrology, 500, 84-96, doi:10.1016/j.jhydrol.2013.07.011, 2013.

Nash, J. E., Sutcliffe, J. V.: River flow forecasting through conceptual models, Part I - A discussion of principles, Journal of Hydrology, 10, 282–290, doi.org/10.1016/0022-1694(70)90255-6, 1970.

Nogueira, S. M. C., Moreira, M. A., and Volpato, M. M. L.: Evaluating Precipitation Estimates from Eta, TRMM and CHRIPS Data in the South-Southeast Region of Minas Gerais State-Brazil, Remote Sens., 10, 16, doi:10.3390/rs10020313, 2018.

Ning, S., Wang, J., Jin, J., and Ishidaira, H.: Assessment of the Latest GPM-Era High-Resolution Satellite Precipitation Products by Comparison with Observation Gauge Data over the Chinese Mainland, Water, 8, 481-497, doi:10.3390/w8110481,2016.

Paredes-Trejo, F. J., Barbosa, H. A., and Kumar, T. V. L.: Validating CHIRPS-based satellite precipitation estimates in Northeast Brazil, J. Arid. Environ., 139, 26-40, doi:10.1016/j.jaridenv.2016.12.009, 2017.

Pessoa, F. C. L., Blanco, C. J. C., and Gomes, E. P.: Delineation of homogeneous regions for streamflow via fuzzy c-means in the Amazon, Water Pract. Technol., 13, 210-218, doi:10.2166/wpt.2018.035, 2018.

Rivera, J. A., Marianetti, G., and Hinrichs, S.: Validation of CHIRPS precipitation dataset along the Central Andes of Argentina, Atmospheric Research, 213, 437-449, doi:10.1016/j.atmosres.2018.06.023, 2018.

Roy, T., Gupta, H. V., Serrat-Capdevila, A., and Valdes, J. B.: Using satellite-based evapotranspiration estimates to improve the structure of a simple conceptual rainfall–runoff model, Hydrology and Earth System Sciences, 21, 879-896, doi:10.5194/hess-21-879-2017, 2017.

Schmit, T. J., Gunshor, M. M., Menzel, W. P., Gurka, J. J., Li, J., and Bachmeier, A. S.: Introducing the next-generation Advanced Baseline Imager on goes-R, Bulletin of the American Meteorological Society, 86, 1079-+, 10.1175/bams-86-8-1079, 2005.

Shen, G. Y., Chen, N. C., Wang, W., Chen, Z. Q.: High-resolution daily precipitation estimation data derived from Wuhan University Satellite and Gauge precipitation Collaborated Correction method (WHU-SGCC) in TIFF format. PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.896615, 2018.

Simpson, J., Adler, R. F., and North, G. R.: A PROPOSED TROPICAL RAINFALL MEASURING MISSION (TRMM) SATELLITE, Bulletin of the American Meteorological Society, 69, 278-295, doi:10.1175/1520-0477(1988)069<0278:aptrmm>2.0.co;2, 1988.

Sokol, Z.: The use of radar and gauge measurements to estimate areal precipitation for several Czech River basins, Stud. Geophys. Geod., 47, 587-604, doi:10.1023/a:1024715702575, 2003.

Su, F. G., Gao, H. L., Huffman, G. J., and Lettenmaier, D. P.: Potential Utility of the Real-Time TMPA-RT Precipitation Estimates in Streamflow Prediction, J. Hydrometeorol., 12, 444-455, doi:10.1175/2010jhm1353.1, 2011.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., and De Roo, A.: Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin, Journal of Hydrology, 499, 324-338, doi:10.1016/j.jhydrol.2013.07.012, 2013.

Trejo, F. J. P., Barbosa, H. A., Penaloza-Murillo, M. A., Moreno, M. A., and Farias, A.: Intercomparison of improved satellite rainfall estimation with CHIRPS gridded product and rain gauge data over Venezuela, Atmosfera, 29, 323-342, doi:10.20937/atm.2016.29.04.04, 2016.

Verdin, A., Rajagopalan, B., Kleiber, W., and Funk, C.: A Bayesian kriging approach for blending satellite and ground precipitation observations, Water Resources Research, 51, 908-921, 2015.

Vila, D. A., de Goncalves, L. G. G., Toll, D. L., and Rozante, J. R.: Statistical Evaluation of Combined Daily Gauge Observations and Rainfall Satellite Estimates over Continental South America, J. Hydrometeorol., 10, 533-543, doi:10.1175/2008jhm1048.1, 2009.

Wang, P. H.: PATTERN-RECOGNITION WITH FUZZY OBJECTIVE FUNCTION ALGORITHMS - BEZDEK,JC, SIAM Rev., 25, 442-442, 1983.

Xie, P. P., and Arkin, P. A.: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions, Journal of Climate, 9, 840-858, doi:10.1175/1520-0442(1996)009<0840:aogmpu>2.0.co;2, 1996.

Yang, T. T., Asanjan, A. A., Welles, E., Gao, X. G., Sorooshian, S., and Liu, X. M.: Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information, Water Resources Research, 53, 2786-2812, doi:10.1002/2017wr020482, 2017.

Yang, Z., Hsu, K., Sorooshian, S., Xu, X., Braithwaite, D., and Verbist, K. M. J.: Bias adjustment of satellite-based precipitation estimation using gauge observations: A case study in Chile, Journal of Geophysical Research: Atmospheres, 121, 3790-3806, doi:10.1002/2015jd024540, 2016.

Yuan, Z., Xu, J. J., and Wang, Y. Q.: Projection of Future Extreme Precipitation and Flood Changes of the Jinsha River Basin in China Based on CMIP5 Climate Models, Int. J. Environ. Res. Public Health, 15, 17, 10.3390/ijerph15112491, 2018.

Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., and Ribbe, L.: Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile, Hydrology and Earth System Sciences, 21, 1295-1320, doi:10.5194/hess-21-1295-2017, 2017.

Zhang, X., and Chen, N. C.: Reconstruction of GF-1 Soil Moisture Observation Based on Satellite and In Situ Sensor Collaboration Under Full Cloud Contamination, Ieee Transactions on Geoscience and Remote Sensing, 54, 5185-5202, 10.1109/tgrs.2016.2558109, 2016.

Zhang, Y. R., Sun, A., Sun, H. W., Gui, D. W., Xue, J., Liao, W. H., Yan, D., Zhao, N., and Zeng, X. F.: Error adjustment of TMPA satellite precipitation estimates and assessment of their hydrological utility in the middle and upper Yangtze River Basin, China, Atmospheric Research, 216, 52-64, 10.1016/j.atmosres.2018.09.021, 2019.