Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

WHU-SGCC：A novel approach for blending daily satellite (CHIRP) and precipitation observations over 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, gridded precipitation data and the Climate Hazards Group Infrared Precipitation (CHIRP, daily, 0.05°) satellite-derived precipitation estimates over the Jinsha River Basin for the period of June-July-August in 2016. 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 precipitation bias adjustments from point to surface, which is evaluated by categorical indices. 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 precipitation events that are less than 20 mm. This study indicates that the WHU-SGCC approach is a promising tool to monitor monsoon precipitation over 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 Jinsha River Basin in summer 2016, 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 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 distributed 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, the retrieval algorithms for satellite-based precipitation estimates are susceptible to systematic biases in hydrologic modelling and 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 without adjustment 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

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

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 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.

Currently, the Climate Hazards Group Infrared Precipitation with Station data (CHIRPS), 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 sparseness of rain gauges (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	3h	0.25°	1998-present	50°S-50°N
GPCP	Monthly/Pentad	2.5°	1979-(delayed) present	90°S-90°N
PRESSIANN-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

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

effective methods for high-accuracy precipitation estimates over complex terrain on a daily scale.

As such, the aim of this article is to offer a novel approach for blending daily precipitation gauge data, gridded precipitation data and the Climate Hazards Group Infrared Precipitation (CHIRP) satellite-derived precipitation estimates over Jinsha River Basin. 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 in summer 2016. 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 precipitation gauges, gridded observations 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 ~180 × 10^4 km^2. In 2016, the average precipitation in the Yangtze River Basin was 12053 mm and the total precipitation was 21478.71 billion m^3, which is 10.9% higher than the annual average total precipitation. Yangtze River is divided into nine sub-regions, 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 × 103 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 climate variation within the basin. Average annual precipitation in the Jinsha River Basin is approximately 3433.45 mm, the total annual precipitation north of Shigu is 937.25 mm, while south of Shigu annual precipitation is 2496.20 mm. The climate of the Jinsha River Basin has more precipitation during the warm season (June-July-August, JJA), which is affected by oceanic southwest and southeast monsoons and is drier in cold season (December to February). Therefore, the blending of satellite estimations with gauged observations during the summer (JJA) is the main focus of this research.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

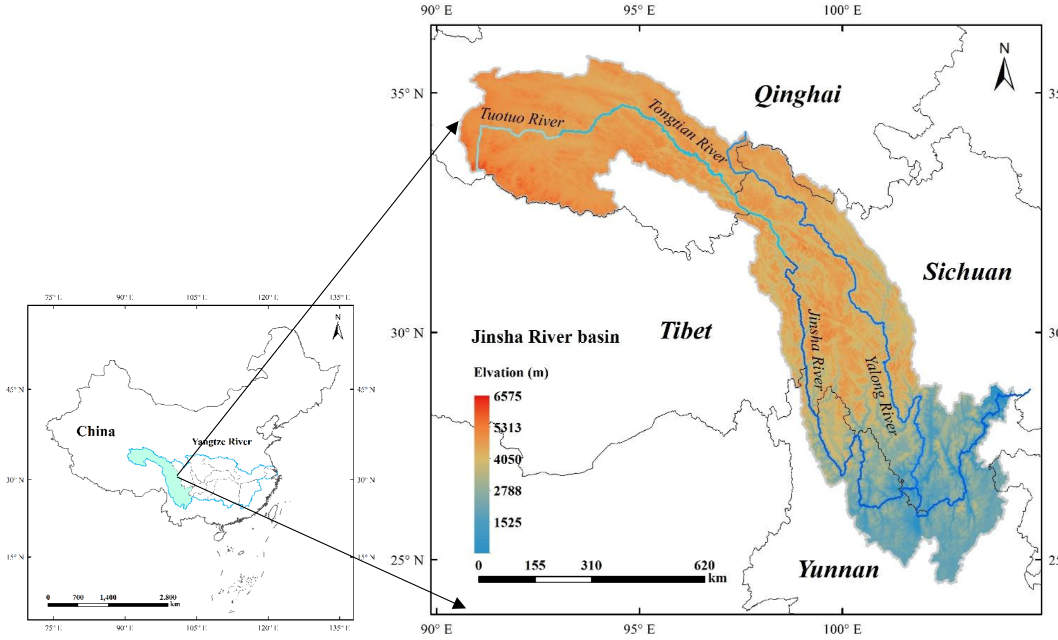


Figure 1 Location of the study area with key topographic features.

2.2 Study Data

2.2.1 Precipitation gauged observations

Daily rain gauge observations at 30 national standard rain stations in the Jinsha River Basin for JJA 2016 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

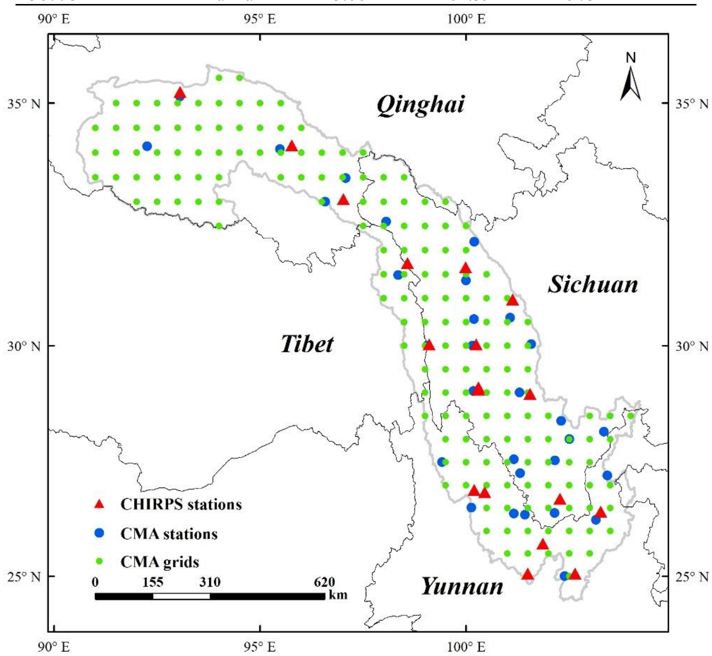


Figure 2 Jinsha River Basin with 18 CHIRPS fusion stations, 30 gauge stations and 170 grid points provided by the China Meteorological Administration stations.

2.2.2 Gridded precipitation observations

The gridded precipitation data developed by CMA with 0.5°× 0.5° resolution on a daily scale, was interpolated from 2472 gauge observations with a thin plate spline algorithm from 1961 to the present. Over the Jinsha River Basin, a total of 170 gridded points were selected as the supplementary data for observations in JJA 2016, due to the 2472 gauged station data that were not shared on CMA (http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_PRE_DAY_GRID_0.5.html, last access: 10 December, 2018). The even distribution of daily gridded precipitation observations is shown in Fig. 2.

2.2.3 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 (Climate Hazards Group Precipitation Climatology), quasi-global geostationary thermal infrared satellite observations (TRMM 3B42 version 7), atmospheric model rainfall fields CFS (Climate Forecast System) from NOAA, and precipitation observations from various

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

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).

CHIRPS is the 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 Jinsha River Basin in JJA 2016 was selected as the input for WHU-SGCC blending with rain observations, and the corresponding daily CHIRPS 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 and 170 grid points provided by CMA; hence, the number of CHIRP stations limited the accuracy of spatial rainfall estimates (Fig. 2).

3 Methods

3.1 The WHU-SGCC approach

In this study, the approach of the WHU-SGCC is to estimate precipitation for every pixel by blending satellite estimates and rain gauge observations considering terrain factors and precipitation characteristics. There were five steps to establish the numerical relationship between gauged stations and corresponding satellite pixels and other pixels. On this basis, the WHU-SGCC method identifies the geographical locations and topographical features of each pixel and applies the classification principles of the SICR approach, including five classification and blending rules. The basic description of the WHU-SGCC method is given below, with details illustrated separately in later sections.

1) Classify all regional pixels into five types: C1 (pixel including one gauged station in its area), C2 (pixel including one gridded point), C3 (pixel physically similar to C1C2), C4 (pixel physically similar to C3) and C5 (remaining pixels). The training samples represented 70% of total gauged stations and gridded points, and the remaining data were used to test model performance.

2) Analyse the relationships between precipitation observations and the C1, C2, and C3 pixel types, and with the C4 and C5 pixels. These relationships are described by five rules, detailed below as Rules 1 through 5.

3) Bias-adjust, establish regression models and screen target pixels based on the five aforementioned rules.

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

5) A flowchart of the WHU-SGCC method is shown in Fig. 3. The proposed approach was evaluated for the Jinsha River Basin for JJA 2016. From that data, the training samples represented 70% of total gauged stations and gridded points, and the remaining data were used to verify the model performance.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

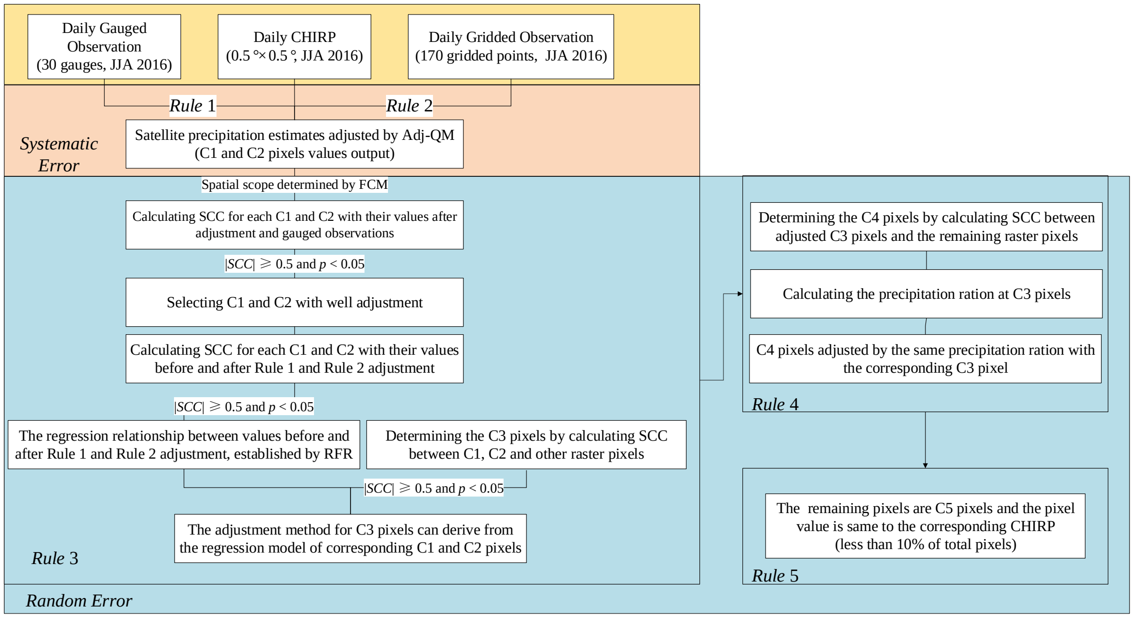


Figure 3 Flowchart of the WHU-SGCC approach with the five rules applied in this study.

3.1.1 Assumptions

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

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

3) Spearman's Correlation Coefficient (SCC) can indicate the 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 observed data, which were assumed to be the true values. Rule 1 adjusts the biases in the satellite estimations. For every C1, its precipitation value was derived from a quantile mapping (QM) approach. It has been shown that the QM method is the best for reducing systematic biases of regional satellite precipitation estimates because of its independence from predetermined functions (Themessl et al., 2011;Chen et al., 2013).

QM is a nonparametric empirical approach that considers a time-dependent correction function. This approach is designed to transform the cumulative distribution function (CDF) of satellite data into the CDF of data at each station.

$$Y_o = \mathrm{h}(Y_s) \tag{1}$$

where the variable Y_s is the distribution of the observed variable Y_o. In this study Y_o denotes each gauge or gridded precipitation data point location from CMA and Y_s denotes the corresponding CHIRP grid cell value. The objective of QM is to correct the daily precipitation amount from a climate simulation and the transformation h is defined as Eq. (2):

$$Y_o = H_o^{-1}(H_s(Y_s)) \tag{2}$$

where the H_s is the CDF of Y_s and H_o^{-1} is the inverse CDF (or quantile function) corresponding to Y_o (Gudmundsson et al., 2012).

Notably, we separately calculate CDFs at each gauge and gridded pixel using the historical daily rainfall from the JJA in 2016.

The result of a QM adjustment is $\bar{Y}_{Qm}$, which is approximately the same as the CDF of the gauge observations on a seasonal scale, which is distinct from daily data. The suitable scale of the CDF is seasonal because the QM cannot effectively remove

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

the day-by-day random errors in CHIRP estimates. Therefore, on the basis of $\bar{Y}_{Qm}$, the adjustment result, $\widehat{C1}_{as}$, for each C1 pixel is derived from the minimum absolute value of the difference between the gauge observations and satellite estimations before and after applying the QM adjustment, referred to as the adjusted QM (Adj-QM) method, as shown in Eq. (3) - Eq. (5).

$$D_{QM} = |\bar{Y}_{QM} - Y_o| \tag{3}$$

$$D_s = |Y_s - Y_o| \tag{4}$$

$$\widehat{C1}_{as} = \begin{cases} \bar{Y}_{QM}, & D_{QM} \le D_s \\ Y_s, & D_s \le D_{QM} \end{cases} \tag{5}$$

where D_{QM} is the absolute value of the difference between $\bar{Y}_{Qm}$ and Y_o, and D_s is the absolute value of difference between $\bar{Y}_s$ and Y_o.

3.1.3 Rule 2 of the WHU-SGCC method

Commonly, a few of the national standard stations have free access, and these stations are unevenly distributed and do not satisfied the accuracy needed for regional precipitation estimation. Under these circumstances, the gridded precipitation data developed by CMA are applied as the supplementary data for observations with uniform spatial distribution. Therefore, Rule 2 is same as Rule 1 with different input data. $\widehat{C2}_{as}$ is the adjusted target precipitation of one C2 pixel.

3.1.4 Rule 3 of the WHU-SGCC method

The aforementioned methods improve the accuracy of satellite precipitation estimations based on historical observations data for C1 and C2 pixels. It is reasonable to assume that there are some pixels that are physically similar to the precipitation characteristics of C1 and C2 pixels in a certain spatial scope. Therefore, it is feasible to adjust the satellite estimation bias of C3 pixels by building numerical relationships between C1 and C2 pixels before and after adjustments based on Rule 1 and Rule 2.

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 *t*. The number of clusters was set as the default value of 20 considering the algorithm efficiency and clustering results.

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

$$c_j^{(t)} = \frac{\sum_{i=1}^{n} w_{ij}(x_i)^m x_i}{\sum_{i=1}^{n} w_{ij}(x_i)^m} \quad (6), \qquad w_{ij} = \frac{1}{\sum_{k}^{c} \left(\frac{\| x_i - c_j \|}{\| x_i - c_k \|}\right)^{\frac{2}{m-1}}} \quad (7)$$

where *x* is a finite collection of *n* elements that will be partitioned into a collection of *c* fuzzy clusters, c_j 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

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

belongs to c_j. In Eq. (6), $c_j^{(t)}$ represents the cluster centre in iteration *t*.

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

$$F_c = \sum_{i=1}^{n}\sum_{j=1}^{c} w_{ij}^{m} \| x_i - c_j \|^2 \tag{8}$$

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.

Second, the adjusted C1 and C2 are employed. SCC was used as the evaluation index for each C1 and C2 with their values after adjustment and gauge observations in JJA:

$$SCC = \frac{\sum_{i=1}^{n}(rgx_i - rg\overline{x})(rgy_i - rg\overline{y})}{\sqrt{\sum_{i=1}^{n}(rgx_i - rg\overline{x})^2}\sqrt{\sum_{i=1}^{n}(rgy_i - rg\overline{y})^2}} \tag{9}$$

Spearman's correlation coefficient is defined as Pearson's correlation coefficient between the ranked variables, and it assesses monotonic relationships (whether linear or not) where *n* is the number of data points in each set, which was the number of each C1 or C2 in the historical JJA dataset. x_i is the *i*th data value in the first data set (satellite estimations after Rule 1 and Rule 2 adjustment, $\widehat{C1}_{as}$ and $\widehat{C2}_{as}$), x_i is converted to its rank, rgx_i, and $rg\overline{x}$ is its average value. Similar definitions exist for rgy_i and $rg\overline{y}$ (gauge and gridded observations at C1 and C2 pixels, Y_o). The value range of the *SCC* is between -1 and +1. If there are no repeated data values, a perfect *SCC* 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, confidence is not only determined by the value of the correlation coefficient but also from the correlation test's *p* value. The critical value is 0.05, thus a *p* lower than 0.05 indicates the data are significantly correlated. Therefore, the C1 and C2 pixels selected for Rule 3 must meet the following criteria:

$$|SCC| \geq 0.5 \quad and \quad p < 0.05 \tag{10}$$

Third, the filtered C1 and C2 pixels after adjustment is used to establish a regression model between the historical $\widehat{C1}_{as}$, $\widehat{C2}_{as}$ and Y_s. To ensure high accuracy, it is necessary to calculate the *SCC* and *p* values between $\widehat{C1}_{as}$, $\widehat{C2}_{as}$ and Y_s, and complete the filtering criteria described above in Eq. (7) before building the regression model. The regression relationship was derived by random forest regression (RFR). 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.

$$\widehat{C1}_{as} \text{ or } \widehat{C2}_{as} = f_{RFR}(Y_s) \tag{11}$$

where f_{RFR} is constructed from the time series $\widehat{C1}_{as}$ or $\widehat{C2}_{as}$ (dependent variable) and the corresponding Y_s data (independent variable) at filtered C1 and C2 pixels in JJA by means of RFR. The number of decision trees was set at the default value of 500.

Fourth, as mentioned above, the aim of Rule 3 is to derive an adjustment method for C3 pixels based on learning from Rule 1 and Rule 2. With the establishment of a regression relationship between values before and after adjustment of the C1 and C2 pixels by RFR method, the determination of C3 pixels follows a considerable procedure. Pixels in each cluster represent

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

potential C3 pixels, with exception of the C1 and C2 pixels and are called R pixels. Spearman's *r* and *p* values between the satellite estimations (CHIRP grid cell values) at R pixels and the C1 and C2 pixels are the criteria for final determination of C3 pixels. Each R pixel has *m SCC* and *p* values (the number of C1 and C2 pixels in the cluster), and the subset of C3 pixels is identified by excluding the data that failed the correlation test and retaining both the data with a maximum *SCC* of at least 0.5 and the corresponding index of C1 and C2 pixels. The selected C3 pixels are physically similar to the precipitation characteristics of corresponding C1 and C2 pixels in their defined spatial scope.

After identifying the C3 pixels and their corresponding C1 and C2 pixels, the adjustment method for C3 pixels is derived from the regression model for the C1 and C2 pixels.

$$C\widehat{3}_{as} = f_{RFRc}(Y_s) \tag{12}$$

where $C\widehat{3}_{as}$ is the adjusted satellite precipitation estimate and Y_s is the CHIRP grid cell value for the C3 pixels, and f_{RFRc} is the f_{RFR} of corresponding C1 and C2 pixels.

3.1.5 Rule 4 of the WHU-SGCC method

Recognizing that precipitation has a spatial distribution, the assumption that C4 pixels are physically similar to the precipitation characteristics of C3 pixels was adopted to establish the adjustment method for C4 pixels.

First, the determination of C4 pixels in each spatial cluster is based on the selection of C3 pixels. The satellite estimation values for regional pixels with exception of the C1, C2 and C3 pixels are used to calculate the *SCC* and *p* values with Y_s for the C3 pixels in the same cluster of the JJA dataset. The results of each pixel's *k SCC* and *p* value (the number of C3 pixels in the cluster) are evaluated based on the correlation test, and the pixels with a maximum *SCC* of at least 0.5, as well as the corresponding index of C3 pixels, are retained. The selected pixels called C4 pixels, which are physically similar to the precipitation characteristics of the corresponding C3 pixels in the defined spatial scope.

After identifying the C4 pixels, a method for merging method the CHIRP grid cell values at C4 pixels (Y_s) and the target reference values of $C\widehat{3}_{as}$ at the corresponding C3 pixels was applied to estimate the adjusted precipitation values for C4 pixels. This method combines Y_s and $C\widehat{3}_{as}$ values in one variable, as shown in Eq. (13):

$$w_i = \frac{C\widehat{3}_{as_i} + \lambda}{Y_{s_i} + \lambda} \quad i=1, \ldots, \ n \tag{13}$$

where λ is a positive constant set to 10 mm (Sokol, 2003), $C\widehat{3}_{as}$ is the adjusted precipitation values for the C3 pixels, Y_{s_i} is extracted from the CHIRP values for the pixel corresponding with the C3 pixel, and *n* is the number of C3 pixels in each spatial cluster.

Each *w* of the C4 pixels is assigned the same value as the corresponding C3 pixel. Therefore, the value of C4 pixels is derived from Eq. (14):

$$C\widehat{4}_{as} = \max(w \times (Y_s + \lambda) - \lambda, 0) \tag{14}$$

where $C\widehat{4}_{as}$ is the adjusted target precipitation value at one C4 pixel and Y_s is the corresponding CHIRP grid cell value. To avoid precipitation estimates below 0, Eq. (14) sets these negative values to 0.

If there is no C3 pixels in a spatial cluster, the C4 pixels are assumed to be physically similar to the precipitation characteristics of the C1 and C2 pixels and adjusted by the above method in Rule 4.

3.1.6 Rule 5 of the WHU-SGCC method

Excluding the C1, C2, C3 and C4 pixels, the number of remaining pixels, called C5 pixels, is less than 10% of the total number of pixels, and each C5 pixel value ($\hat{C5}_{as}$) is set to be the same as the CHIRP grid cell value at the corresponding position.

In the end, after applying these five rules, we obtained complete daily adjusted regional precipitation maps for summer (JJA) 2016.

3.2 Accuracy assessment

The performance of the WHU-SGCC adjusted precipitation estimates was evaluated by nine statistical indicators: Spearman's correlation coefficient (SCC), 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). SCC, RMSE, MAE and BIAS were used to evaluate how well the SGCC method adjusted satellite estimation bias, while POD, FAR and CSI were used to evaluate precipitation event predictions (Su et al., 2011). SCC measures strength of the nonlinear 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 as Table 3.

Table 3 Accuracy assessment indices.

Accuracy assessment Index	Unit	Formula	Range	Optimal value
Spearman'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 to be adjusted by blending satellite estimations (CHIRP) and observations (gauge stations and gridded points) using the WHU-SGCC approach for the 92 days of JJA 2016. The number of pixels adjusted by each rule in the WHU-SGCC method is shown in Fig. 4. The number of C1 and C2 was nearly 140, as well as 11493 C3 pixels, approximately 6344 C4 pixels, and the number of remaining C5 pixels was no more than 5%.

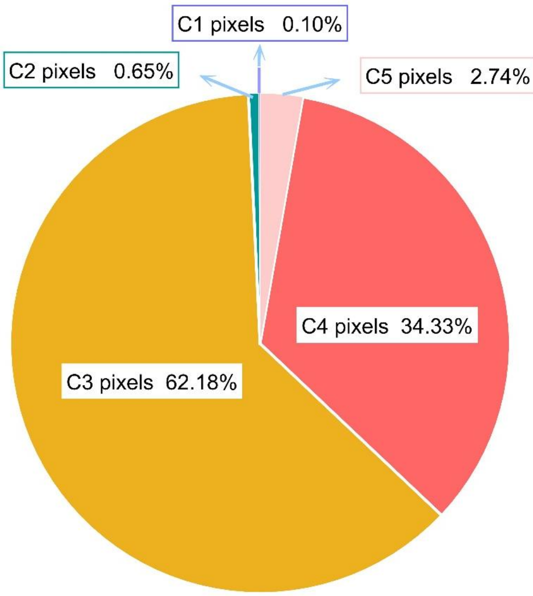
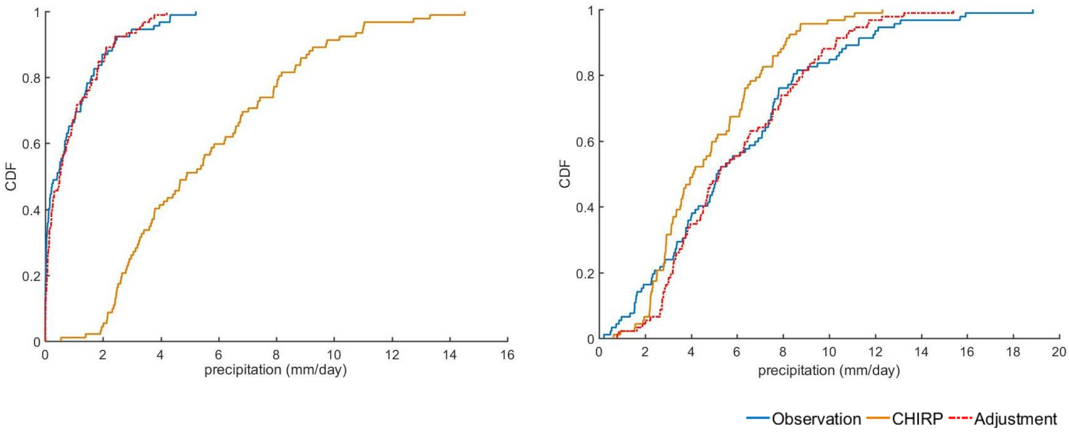


Figure 4 The number of pixels adjusted by each rule using the WHU-SGCC method.

4.1 CDFs of Rule 1 and Rule 2 results

Figure 5 shows the daily average precipitation for observations, CHIRP, C1 (Fig. 5 (a)) and C2 (Fig. 5 (b)) in JJA 2016. Compared to the gauge or grid observations, CHIRP estimations deviated from the observations in Jinsha River Basin. However, the adjusted values for the C1 and C2 pixels improved the estimates and approximated the observations with application of Rule 1 and Rule 2 of the WHU-SGCC method. This result demonstrates that Rule 1 and Rule 2 of WHU-SGCC method are effective in correcting consistent biases and considerably reduce the systematic biases of CHIRP. These improvements not only adjust the bias of satellite estimations but also preserve the original CHIRP pixel values which are close to the corresponding observed data. These adjustments provide reliable precipitation estimates for the C1 and C2 pixels, which supports further study using the WHU-SGCC method, especially for areas in which rain gauges are limited.



(a) (b)

Figure 5 CDFs of seasonal mean daily observations, CHIRP, C1 and C2 estimations for the Jinsha River Basin in JJA 2016

4.2 Spatial Clustering of Rule 3 results

To adjust the pixels other than for the gauged and gridded points, the pixels physically similar to the C1 and C2 pixels were selected. According to Rule 3, C3 pixels were identified in a spatial scope defined by the FCM method. Figure 6 shows the twenty spatial clusters with consideration of the terrain factors. 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.

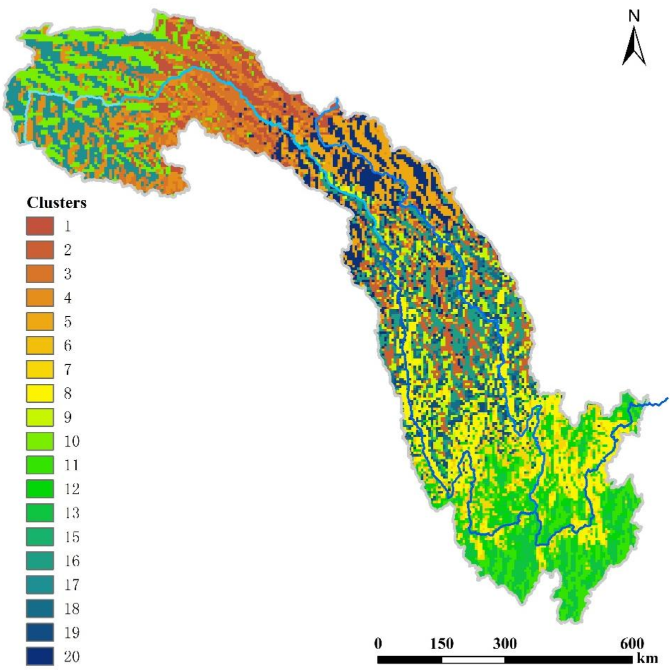


Figure 6 Spatial clustering as defined by FCM for the Jinsha River Basin.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

After Rule 3, each C3 pixel has a good SCC with a C1 or C2 pixel in its cluster; the statistical analysis is shown in Fig. 7. It was found that the average SCC value was 0.6. Therefore, the regression model established in Rule 3 for C1 and C2 before and after adjustment is applicable for each corresponding C3 pixel.

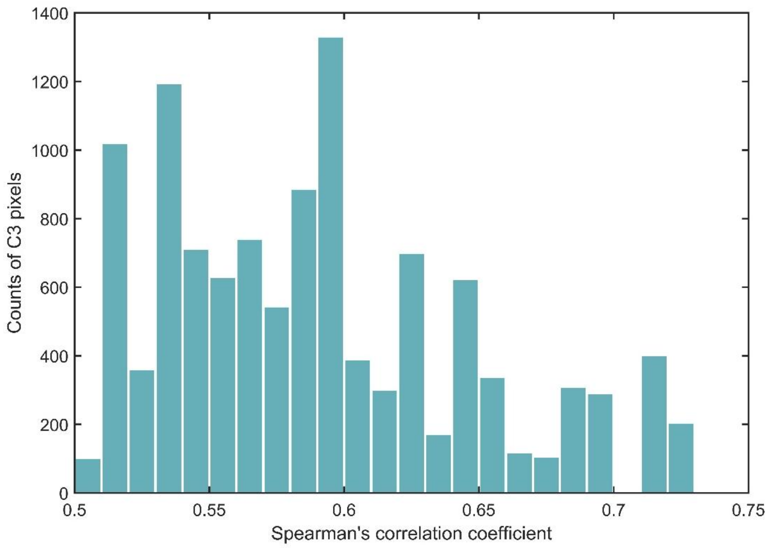


Figure 7 Frequency distribution histogram for Spearman's correlation coefficient (SCC) for C3 pixels and their corresponding C1 and C2 pixels using Rule 3.

It is important to note that 62.18% of the pixels satellite precipitation estimates were adjusted by Rule 3 of the WHU-SGCC method. The accuracy assessment of C3 pixels is shown in Table 4. Validation statistics indicate that compared with the CHIRP and CHIRPS satellite estimations, the WHU-SGCC approach provides best adjustments based on all the statistical indicators at C3 pixels. With the improvement of precipitation accuracy by WHU-SGCC of C3 pixels, the adjustments of C4 pixels, which mainly rely on C3 pixel corrections, are reasonable.

Table 4 Accuracy assessment of C3 pixels for JJA 2016.

Statistic	WHU-SGCC	CHIRP	CHIRPS
SCC	0.3518	0.3176	0.2476
RMSE	5.1776	5.6686	7.0311
MAE	3.5226	3.7353	4.6909
BIAS	-0.0831	-0.2366	-0.2404
NSE	-0.0590	-0.2693	-0.9528
POD	1.0000	0.8900	0.3396
FAR	0.0687	0.0749	0.0763
CSI	0.9313	0.8302	0.3304

4.3 Model performance based on overall accuracy evaluations

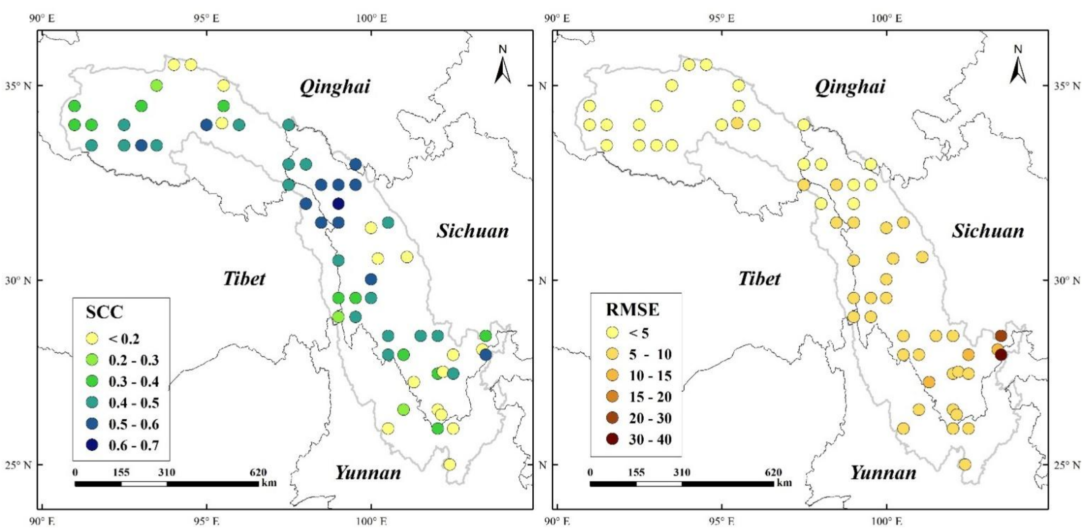
To test the performance of the WHU-SGCC method for precipitation estimates, the statistical analyses of SCC, RMSE, BAE, BIAS, NSE, POD, FAR, and CSI were calculated and are presented in Table 5. Compared with the satellite images of CHIRP and CHIRPS, the results of the WHU-SGCC provide the greatest improvements for regional daily precipitation estimates over the Jinsha River Basin in JJA 2016. After bias adjustment of the WHU-SGCC, SCC was improved by 17.38% and 39.62% compared to CHIRP and CHIRPS, respectively. Meanwhile, the RMSE, MAE and BIAS of the WHU-SGCC decreased by 4.20%, 6.23% and 11.83% compared to CHIRP, and by 19.10%, 24.47% and 41.93% compared to CHIRPS. The NSE of the

WHU-SGCC reached -0.0864, an increase of 0.10 and 0.60 compared to CHIRP and CHIRPS, respectively. It is noted that not only was the POD improved to over 0.95, but the CSI was also improved to over 0.85, and all the FARs were approximately 0.11.

Table 5 Overall accuracy assessment in JJA 2016.

Statistic	WHU-SGCC	CHIRP	CHIRPS
SCC	0.3006	0.2561	0.2153
RMSE	8.3349	8.7003	10.3026
MAE	4.4671	4.7641	5.9146
BIAS	-0.0529	-0.0600	-0.0911
NSE	-0.0864	-0.1838	-0.6599
POD	0.9822	0.9230	0.3686
FAR	0.1023	0.1122	0.1125
CSI	0.8833	0.8266	0.3522

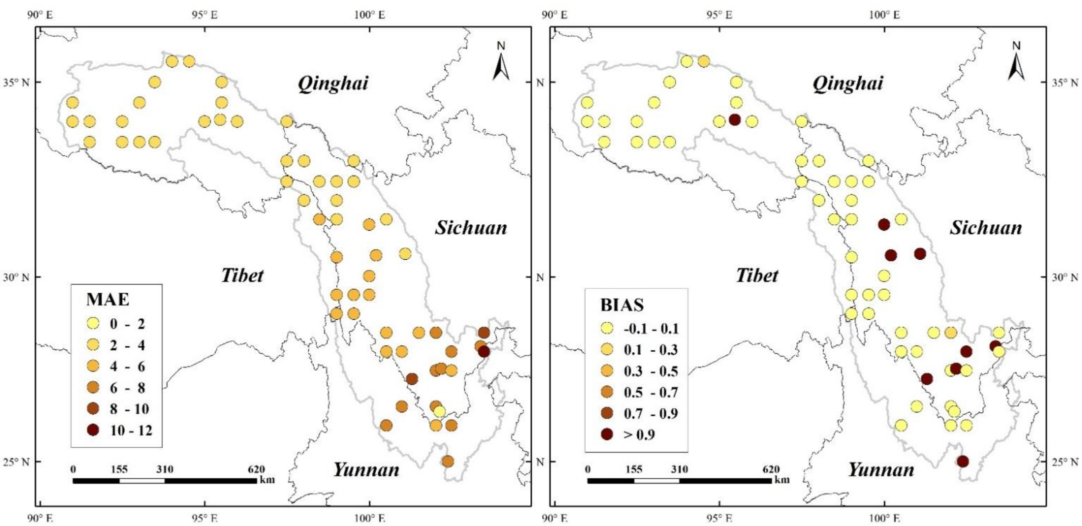
The spatial distributions of the statistical comparisons between observations and WHU-SGCC precipitation estimations are shown in Fig. 8. The variation of SCC as seen in Fig. 8 (a) shows that low correlations are observed in areas with lower elevation, particularly in the southern Jinsha River Basin 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). However, 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. 8 (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 20 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 12 mm and the spatial behaviour is similar to that of the RMSE. Fig. 8 (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 WHU-SGCC is effective in adjusting the satellite biases by blending with the observations, particularly in the complicated mountainous region where there are higher SCC corresponding to lower values of RMSE, MAE and BIAS. The lower SCC and higher errors located over the area southeast of the river basin showed very limited improvement in precipitation estimates.



Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

Open Access Earth System Science Data Discussions

(a) (b)

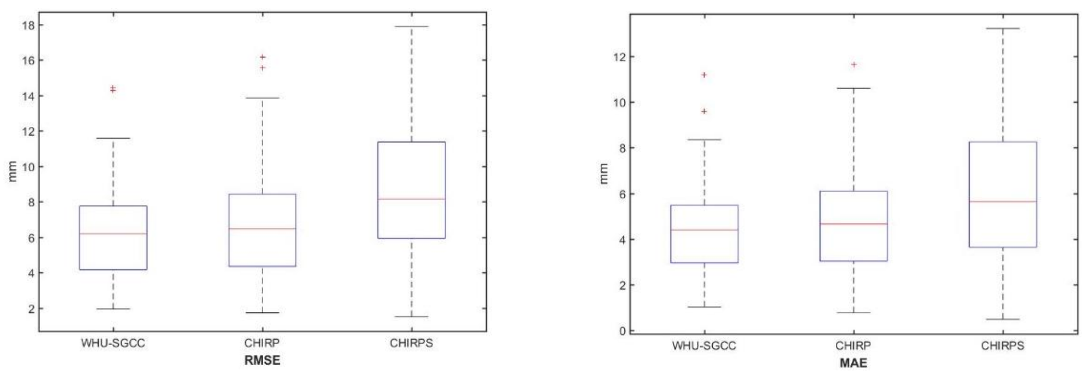


(c) (d)

Figure 8 Spatial distribution of the statistical analyses of the overall agreement between observations and the WHU-SGCC estimations on 30% validation for JJA 2016: a) Spearman's correlation coefficient, b) root mean square error c) mean absolute error, and d) relative bias.

4.4 Model performance based on daily accuracy evaluations

After overall accuracy evaluations for JJA were conducted, further evaluations of daily accuracy were undertaken and the results are shown in Fig. 9. The evaluation of daily accuracy indicates that the WHU-SGCC reduces errors and biases compared to CHIRP and CHIRPS, with especially greatly decreases compared to CHIRPS. The RMSE and MAE derived from the WHU-SFCC were reduced by approximately 5% and 30% compared to CHIRP and CHIRPS, respectively. However, the greatest reduction was reflected in the BIAS, with at least an 18% and 30% reduction compared to CHIRP and CHIRPS, respectively. Therefore, the WHU-SGCC approach is effective for adjustments of daily precipitation estimates, and improves estimate performance.



Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

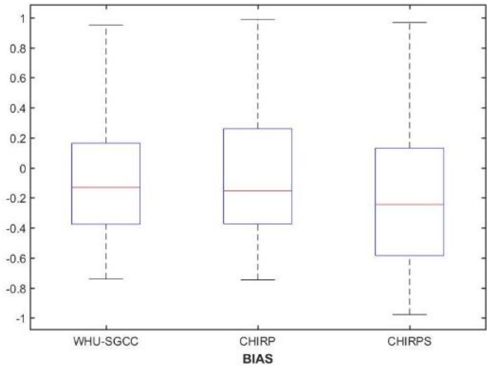


Figure 9 The statistical analysis of the agreement between daily observations and WHU-SGCC, CHIRP and CHIRPS estimates on 30% validation: a) root mean square error b) mean absolute error, and c) relative bias.

The series of daily precipitation differences between WHU-SGCC, CHIRP, CHIRPS and observations is presented in Fig. 10. In this comparison, the WHU-SGCC has the best agreement with the observations, and provides a certain improvement compared to CHIRP, while CHIRPS shows the greatest inconsistencies with the observations. Furthermore, it is noted that differences in precipitation estimates and observations are reduced gradually as the season progresses, especially in August when rainfall is decreased. But at days 36 and 56, short heavy rain events occurred in conjunction with the largest differences in observed WHU-SGCC values. This indicates that short heavy rainstorms may affect the accuracy of precipitation estimates, which deserves further study (Katsanos et al., 2016b;Herold et al., 2017). However, in general, the precipitation estimated using the WHU-SGCC method are superior to other products.

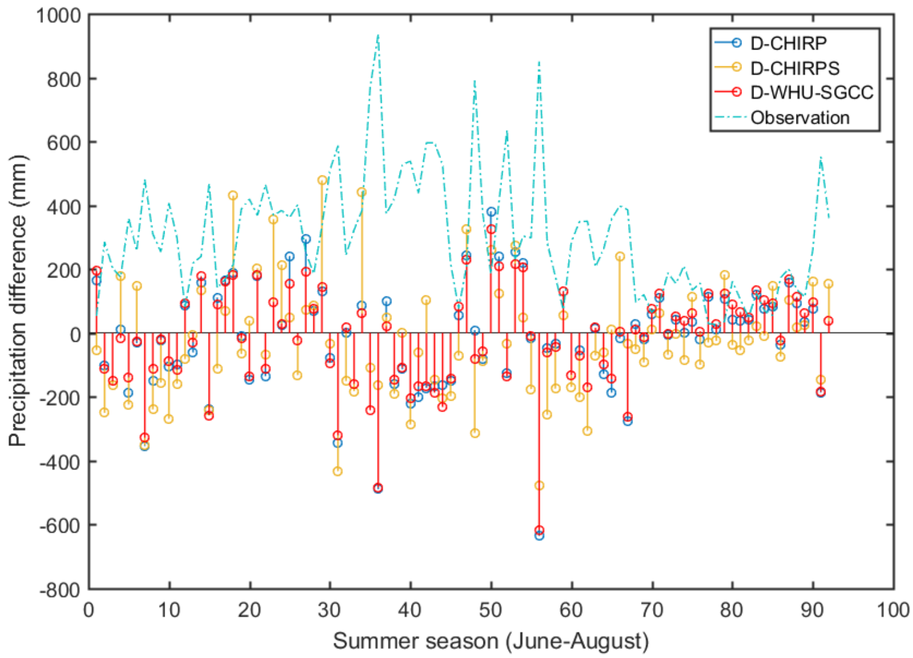


Figure 10 The daily precipitation difference between WHU-SGCC, CHIRP, CHIRPS and observations; D-CHIRP is the difference between CHIRP and observations, D-CHIRPS is the difference between CHIRPS and observations, and D-WHU-SGCC is the difference between WHU-SGCC and observations.

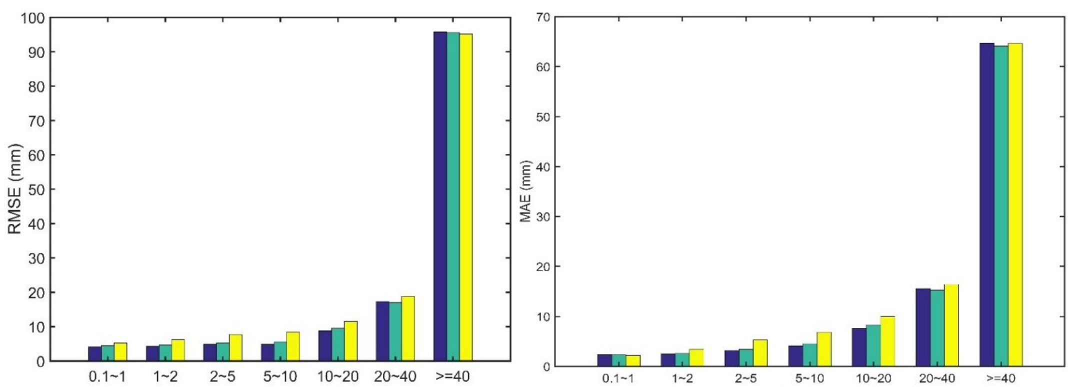
Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

4.5 Model performance for rain events

To measure the WHU-SGCC performance for different rain events, the daily precipitation thresholds of 0.1, 1, 2, 5, 10, 20, and 40 mm were considered, and the result is shown in Table 6 and Fig. 11. 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 between 1 and 20 mm, but WHU-SGCC performance for heavy rain (20-40 mm) events did not improve compared to CHIRP, though it was better than that of CHIRPS. Although the WHU-SGCC approach improved accuracy for light rain events, its behaviour for heavy rain (≥40 mm) events was not as good as CHIRP and CHIRPS, as shown in Fig. 11. These results indicate that WHU-SGCC is applicable for the detection of rainfall events with less than 20 mm precipitation, while there is insufficient observational data for the validation of WHU-SGCC performance during heavy rain events, which represented less than 4% of all observational data and were not sufficient to fully test performance of the model.

Table 6 Accuracy assessment on wet precipitation events for JJA 2016

	RMSE			MAE			BIAS		
Rain Event	WHU-SGCC	CHIRP	CHIRPS	WHU-SGCC	CHIRP	CHIRPS	WHU-SGCC	CHIRP	CHIRPS
[0.1,1)	4.1609	4.5077	5.2762	2.3569	2.2940	2.2187	4.8423	4.9153	4.7541
[1 , 2)	4.2658	4.7385	6.2943	2.4820	2.5563	3.3707	1.3491	1.8199	2.3996
[2 , 5)	4.8378	5.2392	7.7315	3.2026	3.4011	5.2681	0.2808	1.0023	1.5525
[5 , 10)	4.8765	5.5616	8.4619	4.0646	4.5505	6.8346	-0.2292	0.6315	0.9485
[10,20)	8.8240	9.5254	11.5381	7.5957	8.3153	10.0287	-0.4627	0.6142	0.7408
[20,40)	17.3305	17.0107	18.8758	15.5649	15.2646	16.4080	-0.6035	0.6011	0.6461
≥40	95.8157	95.5185	95.2107	64.6789	64.1252	64.6337	-0.8850	0.8774	0.8844



Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

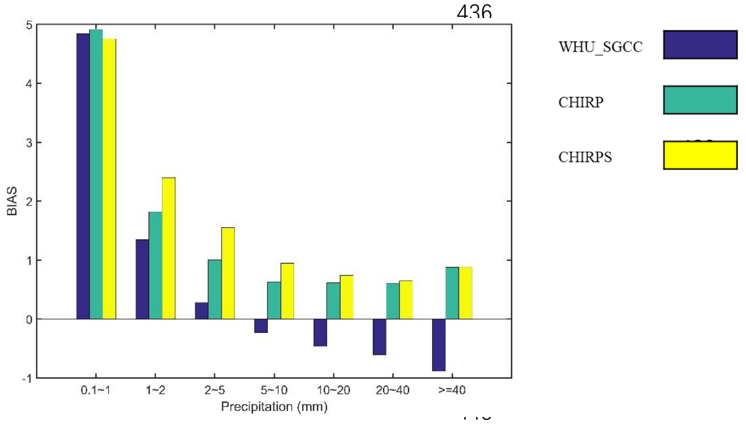


Figure 11 Accuracy assessment based on daily observations for the estimations of WHU-SGCC, CHIRP and CHIRPS for wet precipitation events in JJA 2016: a) root mean square error b) mean absolute error, and c) relative bias.

5 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 Jinsha River Basin in summer 2016 can be downloaded in TIFF format.

6 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 in JJA 2016, and the adjusted precipitation estimates were compared to CHIRP and CHIRPS. WHU-SGCC aims to reduce systematic and random errors in CHIRP over the region 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 SCC, RMSE, MAE, BIAS, NSE, POD, FAR and CSI. Particularly, the SCC statistic was improved by 17.38% and 39.62% compared to CHIRP and CHIRPS, respectively, and all measured errors were reduced. 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 the precipitation region resulted in a limited improvement in accuracy, with SCC values less than 0.3, especially during short rainstorms. However, higher SCC 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 complicated mountainous region. (3) The WHU-SGCC validations for daily rain events confirmed that the model was effective in the detection of precipitation events less than 20 mm. 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 1-20 mm. Specifically, compared with CHIRP, the RMSE value was reduced by approximately 9%, the MAE value by 2.91% ~ 10.68%, and the BIAS value by 1.49% ~ 175.33%; compared with CHIRPS, the RMSE and MAE values were reduced by 20% ~ 40%, and the BIAS value by 43.78%

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

~ 162.46%.

Therefore, the WHU-SGCC approach can help adjust the biases of daily satellite-based precipitation estimates over Jinsha River Basin, the complicated mountainous terrains with sparse rain gauges, particularly for precipitation events with less than 20 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 reducing random errors in all seasons; 2) introduction of more topographic and long time series climatic factors to achieve a more accurate spatial distribution of precipitation; and 3) investigation of the performance over other areas.

7 Acknowledgments

This work was supported by the National Nature 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.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

Open Access Earth System Science Data Discussions

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.

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, 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.

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.

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Bo, E. R., 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.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods, Hydrology and Earth System Sciences, 16, 3383-3390, doi:10.5194/hess-16-3383-2012, 2012.

Herold, N., Behrangi, A., and Alexander, L. V.: Large uncertainties in observed daily precipitation extremes over land, J. Geophys. Res.-Atmos., 122, 668-681, doi:10.1002/2016jd025842, 2017.

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.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 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.

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.

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.

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 (DOI registration in progress), 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.

Themessl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, Int. J. Climatol., 31, 1530-1544, doi:10.1002/joc.2168, 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,

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 2 January 2019

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.

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.