

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/324681765>

Cross-Evaluation of Ground-based, Multi-Satellite and Reanalysis Precipitation Products: Applicability of the Triple Collocation Method across Mainland China

Article in *Journal of Hydrology* · April 2018

DOI: 10.1016/j.jhydrol.2018.04.039

CITATIONS

7

3 authors:



Changming Li
Tsinghua University

2 PUBLICATIONS 19 CITATIONS

[SEE PROFILE](#)

READS

237



Guoqiang Tang
Tsinghua University

43 PUBLICATIONS 518 CITATIONS

[SEE PROFILE](#)



Yang Hong
496 PUBLICATIONS 14,826 CITATIONS

[SEE PROFILE](#)

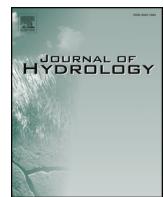
Some of the authors of this publication are also working on these related projects:



Capacity Building + Technology Transfer + Workshop Training [View project](#)



Global_Precipitation_Measurement_Validation_Applications [View project](#)



Review papers

Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: Applicability of the Triple Collocation method across Mainland China

Changming Li^a, Guoqiang Tang^{a,*}, Yang Hong^{a,b,*}

^a State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, China

^b Department of Civil Engineering Environmental Science, University of Oklahoma, Norman, OK, United States

ARTICLE INFO

This manuscript was handled by Emmanouil Anagnostou, Editor-in-Chief, with the assistance of Amir AghaKouchak, Associate Editor.

Keywords:
Triple Collocation
Precipitation
Evaluation
China
Tibetan Plateau

ABSTRACT

Evaluating the reliability of satellite and reanalysis precipitation products is critical but challenging over ungauged or poorly gauged regions. The Triple Collocation (TC) method is a reliable approach to estimate the accuracy of any three independent inputs in the absence of truth values. This study assesses the uncertainty of three types of independent precipitation products, i.e., satellite-based, ground-based and model reanalysis over Mainland China using the TC method. The ground-based data set is Gauge Based Daily Precipitation Analysis (CGDPA). The reanalysis data set is European Reanalysis Agency Reanalysis Product (ERA-interim). The satellite-based products include five mainstream satellite products. The comparison and evaluation are conducted at 0.25° and daily resolutions from 2013 to 2015. First, the effectiveness of the TC method is evaluated in South China with dense gauge network. The results demonstrate that the TC method is reliable because the correlation coefficient (CC) and root mean square error (RMSE) derived from TC are close to those derived from ground observations, with only 9% and 7% mean relative differences, respectively. Then, the TC method is applied in Mainland China, with special attention paid to the Tibetan Plateau (TP) known as the Earth's third pole with few ground stations. Results indicate that (1) The overall performance of IMERG is better than the other satellite products over Mainland China, followed by 3B42V7, CMORPH-CRT and PERSIANN-CDR. (2) In the TP, CGDPA shows the best overall performance over gauged grid cells, however, over ungauged regions, IMERG and ERA-interim slightly outperform CGDPA with similar RMSE but higher mean CC (0.63, 0.61, and 0.58, respectively). It highlights the strengths and potentiality of remote sensing and reanalysis data over the TP and reconfirms the cons of the inherent uncertainty of CGDPA due to interpolation from sparsely gauged data. The study concludes that the TC method provides not only reliable cross-validation results over Mainland China but also a new perspective for comparatively assessing multi-source precipitation products, particularly over poorly gauged regions such as the TP.

1. Introduction

Precipitation is a key component of the hydrological cycle and a primary input for hydro-meteorological and climate models (Behrangi et al., 2011). Therefore, accurate precipitation estimates are essential for studying the mechanism of the water cycle. There stands a wide suite of instruments monitoring precipitation, including rain gauge, weather radar and satellite-based sensors. In addition, reanalysis products based on meteorological or weather forecast models are also provided in recent years.

Affected by different factors depending on the measurement instrument, precipitation estimates are inevitably subject to various error

sources. For instance, the rain gauge can provide only point-scale data at special sites, and requires interpolation to obtain gridded data sets, which may cause significant sampling error over mountainous or high altitude areas where gauges are scarce. In regard to weather radars, the signal-beam interception and reduction in the radar beam power over complex terrain are the leading factors causing errors (Germann et al., 2006). In addition, although satellite-based products including infrared (IR) and microwave (MW) signals are able to retrieve precipitation information over the large scale, they contain random errors and systematic bias because of inadequate sampling and algorithm imperfections (Xie and Arkin, 1997). As for reanalysis products, like the latest version of European Reanalysis Agency Reanalysis Product (ERA-

* Corresponding authors at: Department of Hydraulic Engineering, Tsinghua University, Room 116, New Hydraulic Engineering Building, Beijing 100084, China (G. Tang). State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Room A207, Beijing 100084, China (Y. Hong).

E-mail addresses: tqq14@mails.tsinghua.edu.cn (G. Tang), hongyang@tsinghua.edu.cn (Y. Hong).

interim) and the National Center for Environmental Prediction (NCEP) reanalysis, they are the fusion of observed data and forecast model outputs. Therefore, the quality of the reanalysis products will inevitably be affected by the forcing data sets, the prediction model and the assimilation method. Previous studies demonstrate that changes in the observing system and uncertainties of the merged data are the main sources of reanalysis product error (Hodges et al., 2003; Inoue and Matsumoto, 2004; Marshall et al., 2002). Overall, the assessment of uncertainties of multi-source precipitation products is essential before application all along.

During the last two decades, a great number of studies have investigated uncertainties in different precipitation products at various spatiotemporal scales (McCollum et al., 2002; Hong et al., 2006; Goudenhoofdt et al., 2009; Tian and Lidard, 2010; Yong et al., 2010; Shen et al., 2014; Tang et al., 2016a,b). However, most studies use in situ gauge observations as the reference, which is usually regarded as approximation of the true value. This may lead to error at the global scale, especially over data missing areas like the Tibetan Plateau (TP), where reliable ground-based data sets are absent.

Triple Collocation (TC) is a method that objectively obtains error estimation of three independent products without knowing the true value. This method was originally introduced by Stoffelen (1998) to evaluate wind speed products. McColl et al. (2014) introduced the Extend Triple Collocation (ETC) method, based on the original mathematical assumptions, making TC capable of solving root-mean-square error (RMSE) and correlation coefficient (CC) between any of the three products and the unknown truth at the same time. Lately, the TC method has been widely applied in characterizing uncertainties of various geographical variables estimates, including soil moisture products (Dorigo et al., 2010; Miralles et al., 2010; Hain et al., 2011; Draper et al., 2013), ocean wind speed (Stoffelen, 1998; Portabella et al., 2009), leaf area index (LAI) (Fang et al., 2013), total water storage (van Dijk et al., 2014), sea ice thickness (Scott et al., 2014), sea surface salinity (Ratheesh et al., 2013), ocean wind and wave data (Caires, 2003).

Roebeling et al. (2012), for the first time, applied the TC method in the evaluation of precipitation estimates between three weather radars across Europe. Lately, Alejomhammad et al. (2015) characterized the error of four precipitation products over the United States, including the National Weather Service Ground-based WSR 88D radar network (NEXRAD IV), the Tropical Rainfall Measurement Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) 3B42, the Global Online Enrollment System (GOES) Precipitation Index (GPI) and the Global Precipitation Climatology Project (GPCP), using the multiplicative triple collocation (MTC) method. The MTC method is an improvement on the original version by replacing the additive error model with the

multiplicative error model to represent the relationship between observations and reference data, which is more appropriate for precipitation assessment (Hossain and Anagnostou, 2006). However, most TC-based studies in the precipitation products evaluation are generally in sparse temporal (biweekly or monthly) and spatial resolution, which can provide only limited information on their reliability and accuracy. Thus, assessment with high resolution is required for better usage of various products. Moreover, the TC method has not been applied in Mainland China yet, which contains the most part of TP. As the Third Pole of the Earth, the TP is a typical data missing area and previous studies based on the traditional evaluation method can only employ limited ground gauges in the eastern part (Gao and Liu, 2013; Wang and Zeng, 2012; Tong et al., 2014; Shen et al., 2014; Tang et al., 2016a,b; Ma et al., 2016).

The objectives of this study are twofold: (1) investigating the reliability and application prospect of the TC method in South China, where rain gauges are dense enough to act as the benchmark; (2) evaluating the uncertainties of five satellite-based products over Mainland China using the TC method, with special attention paid to the TP. The designed Triplet includes: the Gauge Based Daily Precipitation Analysis (CGDPA) (Shen and Xiong, 2016) provided by the China Meteorological Administration (CMA), the ERA-interim reanalysis product developed by the European Center for Medium Range Weather Forecast (ECMWF), and five satellite-based products: Version-7 post-real-time 3B42 (3B42V7) and near-real-time 3B42 (3B42RT) of TMPA, Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (IMERG), bias corrected data of the Climate Prediction Center Morphing Method (CMORPH-CRT), Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks for climate data record (PERSIANN-CDR).

The remaining parts of the paper are organized as follows: Section 2 introduces the study domain and data sets; Section 3 explains the basic mathematical principle of TC and discusses some considerations about this method; Section 4 shows the results of a method validation process over South China; Section 5 represents the outputs of the TC method. Finally, Section 6 concludes.

2. Study domain and data sets

2.1. Study domain

The study domain is Mainland China located between 18°–54° N latitude and 75°–136° E longitude. The digital elevation data is shown in Fig. 1. As it can be seen, the terrain of Mainland China is high in the west and low to the east forming a three-ladder pattern. The climate in Mainland China is complex due to the drastically changing terrain and

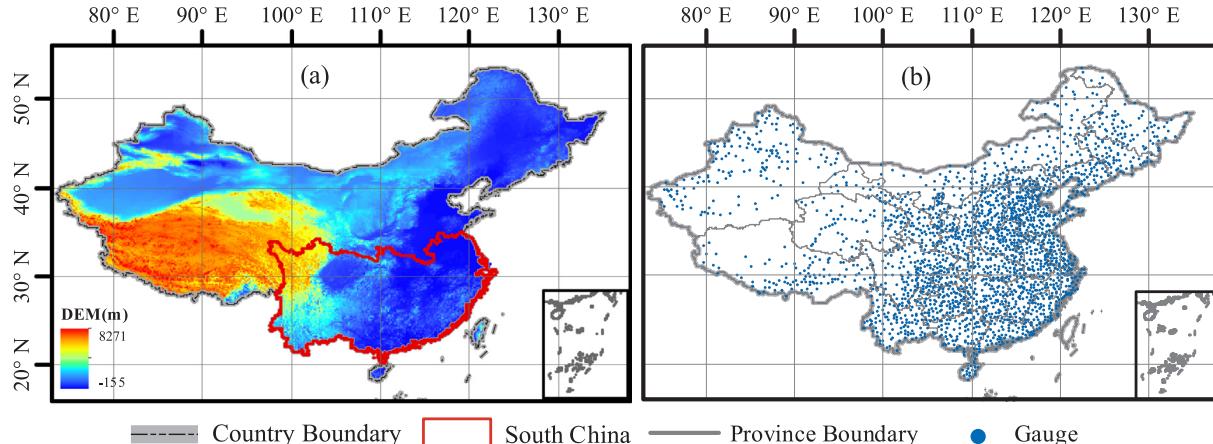


Fig. 1. Digital Elevation Model (DEM) from NOAA and distribution of about 2400 gauges across the Mainland China. The South China is marked out in dark red line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is mainly modulated by the monsoon system. According to the “2016 China Climate Bulletin”, provided by CMA, China can be divided into three main climate regions: the east monsoon region, northwest arid or semi-arid region and TP alpine plateau region. The temporal and spatial variability of precipitation is large all over China, showing a declining trend from the southeast to the northwest (Tang et al., 2016a,b). The majority of TP is located in China. As the highest plateau in the world, TP strongly influences the regional climate and weather system as well as water and energy budgets, playing an important role in the Asian monsoon establishment and maintenance (Duzheng and Youxi, 1979; Hsu and Liu, 2003; Ye and Gao, 1979; Yanai et al., 1992).

In Fig. 1(a), the region marked out in dark red line is named South China for simplicity (Tang et al., 2017), which is chosen as the method validation area. This region contains 13 provinces and is mainly composed of plains. According to the “2016 China Climate Bulletin”, except for the western part of Sichuan province, which is controlled by the TP alpine plateau climate, the other areas of South China are modulated by the monsoon system with about 1500 mm annual mean precipitation. Moreover, about 1300 in situ gauges are well distributed over South China, which can provide reliable ground-based observations. Therefore, with uniform climate and adequate data, this region is suitable for method validation process.

2.2. Data sets

The TC method is based on three collocated input data sets, which are assumed to be totally independent of each other. In this study, the designed triplet is: a ground-based product, a reanalysis product and a satellite-based product.

2.2.1. Gauge precipitation observation

A daily gridded ground-based precipitation data set, CGDPA, provided by CMA, is selected. This data set is constructed at a resolution of 0.25° by optimizing the daily climatological precipitation field and amassing around 2400 gauge reports over Mainland China, using the optimal interpolation (OI) method (Shen and Xiong, 2016).

The input data of CGDPA is the manually recorded daily accumulated rainfall observations by bucket rain gauges. All the report data are subject to strict quality control including an extreme check and an internal consistency check. The minimum detectable hourly rainfall is 0.1 mm/h.

By involving a larger number of gauges, CGDPA outshines the Climate Prediction Center Unified gauge data set (CPC-UNI) in China (Shen and Xiong, 2016). However, the gauge distribution is very sparse over northwestern China and the TP. Thus, the quality of CGDPA over these areas still remains to be improved. Since it is difficult to validate the quality of ground-based data sets in absence of more accurate observations, the TC method may provide us with another prospective to obtain a first insight to the quantitative performance CGDPA in ungauged regions.

2.2.2. Reanalysis precipitation product

The ERA-interim reanalysis data is selected as the second input of the TC method. ERA-interim is a global atmospheric reanalysis product generated by ECMWF. Using the weather forecast model, the estimate of precipitation is based on temperature and humidity information (Dee et al., 2011). By completely revising the method for correcting bias in radiance data and the humidity analysis, the latest ERA-interim is an improvement on the previous ERA-40. For high-latitude regions, ERA-40 provides better depictions than GPCP (Serrze et al., 2005). As the improved version of ERA-40, it's reasonable to regard ERA-interim with the same advantage over the high-latitude areas.

The ERA-interim daily reanalysis data set is downloaded from the ECMWF website (http://apps.ecmwf.int/data_sets/data/interim/full/daily/levtype=sfc/) with a resolution of 0.25°.

2.2.3. Satellite-based products

The list of utilized satellite-based products in this study is shown below:

(1) TMPA-3B42TRMM is an integral mission of the National Aeronautics and Space Administration (NASA) Earth Science Enterprise. The inputs of the algorithm are from two different types of satellite sensors, including MW and IR. The MW data are collected by TRMM, Defense Meteorological Satellite Program (DMSP), Aqua and the National Oceanic and Atmospheric Administration (NOAA) satellites to generate passive microwave based precipitation estimates (Huffman et al., 2010). The CPC merged IR data set (Janowiak et al., 2001) is applied in the TMPA algorithm, which includes the information from Geostationary Meteorological Satellite (GMS), the European Meteorological satellite (Meteosat) series, GOES 8 and GOES 10 satellites. GPCC monthly rain gauge analysis and the Climate Assessment and Monitoring System (CAMS) monthly rain gauge observations are also employed for the validation of the results. The 3B42V7 and 3B42RT products of TMPA at 0.25° and daily resolutions are used in this study. In addition, it should be noted that the spatial extent for 3B42RT and 3B42V7 products is limited to 50°N. The data sets can be downloaded from NASA's PPS server website (<ftp://arthurhou.pps.eosdis.nasa.gov/trmmdata>).

(2) IMERG

As the successor of TRMM, the GPM Core Observatory was launched by NASA and National Space Development Agency (NASDA) on February 27, 2014. The IMERG is intended to inter-calibrate, merge, and interpolate “all” satellite microwave precipitation estimates, together with microwave calibrated IR satellite estimates, precipitation gauge observations over the entire global (Huffman et al., 2014). The PMW information is obtained from GPM, TRMM (deactivated at April 9, 2015), DMSP, Aqua and NOAA satellite series; Geo-IR information is from the same signals of TRMM sensors and GPCC rain gauge observation is also used as the monitoring data.

The Final run data of IMERG at 0.1° and half hourly resolutions is used and resampled to 0.25°. The IMERG data can be downloaded from the Precipitation Measurement Missions (PMM) website (https://pmm.nasa.gov/data_access/downloads/gpm).

(3) PERSIANN-CDR

PERSIANN-CDR is the bias corrected data set of the original PERSIANN product to address the need for a consistent, long time, high resolution and global precipitation data set. It's developed by the University of California, Irvine (Ashouri et al., 2015). The PERSIANN-CDR algorithm employs the infrared brightness temperature data from geostationary satellites to estimate rainfall. The core of this algorithm is an adaptive Artificial Neural Network (ANN) model which updates its parameters based on a pre-training process using NCEP Stage IV hourly precipitation data (Sorooshian et al., 2000). By utilizing the 2.5° monthly GPCP precipitation data to maintain a monthly total consistency of the two data sets, PERSIANN-CDR can provide a bias corrected, near global precipitation data set at 0.25° and daily resolutions.

The global IR data is available through the International Satellite Cloud Climatology Project (ISCCP) provided by NOAA National Climatic Data Center (NCDC). ISCCP B1 global geostationary observations are comprised of several observations from a number of geosynchronous satellites, including the GOES series, the Meteosat series, the GMS series, and the Chinese Fen Yung 2 (FY2) series. Gridded Satellite (GridSat B1) data are used by the PERSIANN-CDR algorithm, which is derived by merging ISCCP B1 IR data.

The PERSIANN-CDR data is available on the website of the Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine

- (<ftp://PERSIANN-CDR.eng.uci.edu/pub/PERSIANN-CDR/>).
 (4) CMORPH-CRTCMORPH uses motion vectors derived from IR information to propagate the relatively high quality precipitation estimates derived from passive microwave data ([Joyce et al., 2004](#)). IR data used in CMORPH's algorithm is based on the observation of GOES 8, GOES 10, Meteosat 8, Meteosat 5 and GMS 5 satellites. The PMW derived precipitation estimates in CMORPH are generated from NOAA polar orbiting operational meteorological satellites, the DMSP satellites and Tropical Rainfall Measuring Mission (TRMM) satellites. CMORPH bias corrected data (CRT) is used in this study. The bias correcting progress is conducted by matching probability distribution function (PDF) against daily gauge observations from the CPC real-time daily gauge analysis over the contiguous United State (CONUS) ([Xie et al., 2011](#)). The data is available on the website developed by NOAA and CPC at 0.25° and daily resolutions (ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/CRT/).

3. Methodology

In this sections, we will review the basic mathematical principle of the TC method and introduce the evaluation indicators used in this study. Finally, some considerations while applying this method will be discussed.

3.1. Mathematical description

Given an affine error model relating measurements to a geophysical variable, a standard form used in the triple collocation literature ([Zwieback et al., 2012](#)):

$$R_i = \alpha_i + \beta_i T + \varepsilon_i \quad (1)$$

where the $R_i(i\{1,2,3\})$ represents any of the three collocated estimates of true value T with observation error $\varepsilon_i(i\{1,2,3\})$, respectively. α_i and β_i are the ordinary least squares (OLS) intercept and slope, respectively.

It has been widely accepted that the multiplicative model is more appropriate to represent the relationship between observation error and the true value for precipitation estimates ([Hossain and Anagnostou, 2006](#)). [Chen et al. \(2013\)](#) compared the results based on two different models over CONUS at the daily scale, indicating that the multiplicative model has advantages in terms of predictability and adaptability. By replacing the additive error model with the multiplicative error model, [Alemohammad et al. \(2015\)](#) introduced the MTC method and draw the conclusion that additive model tends to overestimate the accuracy of rainfall products and underestimate the errors. Therefore, the MTC method is used in this study with the basic function shown below:

$$R_i = A_i T^{\beta_i} E_i \quad (2)$$

Defining $a_i = \ln(A_i)$, $r_i = \ln(R_i)$, $t = \ln(T)$, $\varepsilon_i = \ln(E_i)$ the function can be simplified as:

$$r_i = \alpha_i + \beta_i t + \varepsilon_i \quad (3)$$

The mathematical assumptions of TC method are mainly threefold and the naming of each assumption is given in previous studies ([Zwieback et al., 2012](#)): (1) The triple input products are totally independent of each other ($\rho(r_i, r_j) = 0, i \neq j$), referred as the zero cross-correlation assumption. (2) The errors for different products are independent of each other and unrelated to other products as long as the truth value ($\rho(\varepsilon_i, \varepsilon_j) = 0, \rho(r_i, \varepsilon_j) = 0, \rho(\varepsilon_i, t) = 0, i \neq j$), referred as the zero error cross-correlation assumption. (3) The expected values of the errors are zeros ($E(\varepsilon_i) = 0$), referred as the zero error expectation assumption.

The covariance between two different input data, defined as $C_{ij}(i \neq j)$, is given by:

$$C_{ij} = \beta_i \beta_j \sigma_t^2 + \beta_j E(\varepsilon_i t) + \beta_i E(\varepsilon_j t) + \alpha_i E(\varepsilon_j) + \alpha_j E(\varepsilon_i) + E(\varepsilon_i \varepsilon_j) \quad (4)$$

Based on the zero error expectation and zero cross-correlation assumptions, all the expectation terms in Eq. (4) vanish. Therefore, the covariance value is given as:

$$C_{ij} = \text{Cov}(R_i, R_j) = \begin{cases} \beta_i \beta_j \sigma(t)^2 & (i \neq j) \\ \beta_i^2 \sigma(t)^2 + \sigma(\varepsilon_i)^2 & (i = j) \end{cases} \quad (5)$$

Defining $\theta_i = \beta_i \sigma_t$, the covariance value of each two data is given by:

$$\begin{cases} C_{11} = \theta_1^2 + \sigma_{\varepsilon_1}^2 & C_{12} = \theta_1^2 \theta_2^2 \\ C_{22} = \theta_2^2 + \sigma_{\varepsilon_2}^2 & C_{13} = \theta_1^2 \theta_3^2 \\ C_{33} = \theta_3^2 + \sigma_{\varepsilon_3}^2 & C_{23} = \theta_2^2 \theta_3^2 \end{cases} \quad (6)$$

Then we can solve RMSE between any of the three satellite products and the true value:

$$\sigma_{r_1}^2 = C_{11} - \frac{C_{12} C_{13}}{C_{23}} \quad (7)$$

$$\sigma_{r_2}^2 = C_{22} - \frac{C_{12} C_{23}}{C_{13}} \quad (8)$$

$$\sigma_{r_3}^2 = C_{33} - \frac{C_{13} C_{23}}{C_{12}} \quad (9)$$

[McColl et al. \(2014\)](#) introduced the ETC method to calculate CC in addition to RMSE. The ETC shares the same mathematical assumption of the original TC as a derivative version. In term of the correlation coefficient (CC) between any of the three products and the true (represented as $\rho_{t,i}^2(i\{1,2,3\})$), we can write:

$$\rho(r_i, t) = \frac{\text{Cov}(r_i, t)}{\sqrt{\sigma_{r_i}^2 \sigma_t^2}} = \frac{\beta_i \sigma_t}{\sqrt{\beta_i^2 \sigma_t^2 + \sigma_{\varepsilon_i}^2}} = \frac{\theta_i}{\sqrt{C_{ii}}} \quad (10)$$

Thus, the equations to solve CC of each collocated data set are shown below:

$$\rho_{t,1}^2 = \frac{C_{12} C_{13}}{C_{11} C_{23}} \quad (11)$$

$$\rho_{t,2}^2 = \frac{C_{12} C_{23}}{C_{22} C_{13}} \quad (12)$$

$$\rho_{t,3}^2 = \frac{C_{13} C_{23}}{C_{33} C_{12}} \quad (13)$$

It should be noted that the result of RMSE is absolute ([A Gruber et al., 2016](#)), and can only reflect limited information about the estimates because all kinds of accepted and unaccepted noise have been taken into account. Compared to RMSE, CC are relative and unbiased to noise, representing the similarity of two signals more directly. [McColl et al. \(2014\)](#) proposed that CC can provide more important information about the accuracy of three collocated estimates with respect to the absolute variance. Therefore, we focus on CC, and RMSE is taken as a verification to the reliability of the results.

3.2. Evaluation indicators

Five metrics are chosen to evaluate the performance of satellite products comprehensively, which can be divided into three categories ([Yong et al., 2010](#)). CC describes the agreement with two data sets and it's classified into the first category. The second category includes RMSE, representing the bias and error of satellite estimates with ground-based observations.

The final category includes the probability of detection (POD), the false alarm ratio (FAR) and the critical success index (CSI). POD describes the fraction of real rainfall events captured by satellite among all actual rainfall events while FAR is the opposite ([Ebert, 2007](#)). CSI is a more balanced score, which combines the information about false alarm and missed events, which can be regarded as the fraction of POD

Table 1
Evaluation Indicators.

| Statistic Metrics | Equation | Perfect Value |
|-------------------|---|---------------|
| CC | $CC = \frac{N(\sum_n^f m) - (\sum_n^f)(\sum_n^r)}{\sqrt{[N \sum_n^f 2 - (\sum_n^f)^2][N \sum_n^r 2 - (\sum_n^r)^2]}}$ | 1 |
| RMSE | $RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (f_n - r_n)^2}$ | 0 |
| POD | $POD = n_{11} / (n_{11} + n_{01})$ | 1 |
| FAR | $FAR = n_{10} / (n_{11} + n_{10})$ | 0 |
| CSI | $CSI = n_{11} / (n_{11} + n_{01} + n_{10})$ | 1 |

Notation: n represents the number of samples, both lowercase and uppercase; f represents the reference value and r represents the observed value. In addition, n_{11} means the number of actual rainfall events capture by both gauge and satellite while n_{10} represents the number of the false alarm with n_{01} the opposite.

and FAR (Gerapetritis and Pelissier, 2004). All equations and perfect values of the metrics are listed in Table 1.

3.3. Considerations about the application of Triple Collocation to precipitation estimates

3.3.1. Design of input triplet

As the prerequisites of TC method, the results are based on the independence assumption, error orthogonality and zero cross correlation assumptions. The collocated input of the triplets should be independent of each other both for the estimate data and relative error. What's more, we assume that the errors of each collocated data have zero mean. Therefore, the error of TC method is mainly caused by the breach of the precondition (Stoffelen, 1998).

Yilmaz and Crow (2014) conducted an evaluation of assumptions in soil moisture using the TC method which witnessed an underestimation of the true random error of soil moisture products because of the reason mentioned above. Moreover, results demonstrate that more independent between any two collocated estimates are, the smaller the error of TC based results will be. Since the required TC method assumptions of zero error across covariance do not generally hold for typical precipitation estimates, efforts must be made to minimize the impact of this problem. The collocation of triplets should consider the relevance of input data to minimize the influence to the output.

In this study, the triplets used in TC are chosen as CGDPA, ERA-interim plus one of the five satellite products. For remote sensing precipitation estimates, both the satellite-based and reanalysis products take the ground-based observation as reference in general. PERSIANN-CDR, TMPA-3B42, IMERG and ERA-interim products choose monthly GPCC global data set as the benchmark for validation (Huffman et al., 2007; Dee et al., 2011; Huffman and Bolvin, 2015; Ashouri et al., 2015), which includes some gauges that are mostly located in the southeastern part of Mainland China. Similar to that, CMORPH-CRT utilizes CPC gauge data over CONUS as reference (Joyce et al., 2004). Therefore, satellite-based and reanalysis products are not completely independent of each other to some extent. To reduce this influence, we choose CGDPA data set as the ground precipitation observation other than GPCC or CPC, since it amassed much more sites than GPCC with better accuracy and higher resolution. By choosing this combination, we are meant to reduce the impact of the zero across covariance error to some degree.

3.3.2. Treatment of zero rainfall

Since the multiplicative error model is used in this study, a log transform data preprocessing is required. Therefore, the problem has been arisen about how to deal with the zero rainfall data since it will influence the accuracy of assessment results. In the previous study over CONUS, Alemdohammad et al. (2015) utilized the biweekly accumulated rainfall product and removed the remaining zeros; Roebeling et al. (2012) merged satellite and radar products and assigning zeros to 10^{-3} ,

Massari et al. (2017) simply removed the zeros. Nevertheless, he also indicates that removing zeroes for obtaining log transformed rainfall shortens the sample size, thus providing less robust TC results. All present TC-based studies about precipitation assessment are at sparse resolution both temporally and spatially, and the influence of this problem is reduced to some degree.

In view of the “Atlas of Climate of people’s republic of China” (1979) provided by CMA, the weather over West China (72° – 100° E), is mainly modulated by Temperate Continental Climate in Xinjiang Province and Alpine plateau climate in the TP, where the climate is perennial drought. Based on the CGDPA data set from 1998 to 2014, the ratio of zeros to all records is quite high across West China with the mean value up to 30% and up to 70% during winter at the daily scale. Therefore, in this study, simply removing zeros will cause unacceptable data loss in such regions, influencing the accuracy of TC results. For the above reasons, a replacement approach for zero data should be taken into consideration.

To investigate the feasibility and influence of the replacement scheme, TC results, on different replacement values, among 10^{-9} , 10^{-6} and 10^{-3} , are compared based on the triples: CGDPA, ERA-interim and CMORPH-CRT.

As seen in the RB distribution maps of Fig. 2, though slight change in replacement value will result in big change in the log-transformed value due to steep gradient of log function, three results, three replacement values generate nearly identical results with average relative bias (RB) value between any two outputs about 2.3% over most part of Mainland China. This may benefit from the low ratio of zero rainfall presence. However, this ratio increases rapidly over some grid cells with larger difference (RB value over 30%), which are mainly located around the boundary of the TP. Previous studies demonstrate that complex topography causes the underestimation of satellite precipitation products (Dinku et al., 2007; Gao and Liu, 2013), which may be the main reason why the impact of zero replacement scheme is magnified. In addition, higher replacement value seems to result in better CC results, bringing in overestimation of the product. Meanwhile, since the closer selected alternative is, the more consistent with the observed rainfall, the replacement value is chosen as 10^{-9} in this study.

4. Validation of the triple collocation method

South China is chosen to validate the effectiveness of the TC method. In addition, the performance of additive and multiplicative errors in TC is also compared. To avoid the influence of interpolation, only grid cells with at least one gauge are involved during the process. First, the quality of five satellite products over South China is investigated based on the traditional method via comparing satellite estimates and CGDPA directly. Then, the reliability of the TC method is examined by comparing the CC and RMSE derived from TC and traditional methods.

Fig. 3 shows the distribution of POD, FAR and CSI for IMERG and CMORPH-CRT over South China, taking CGDPA as reference. As the magnitude and spatial distribution of TMPA-3B42 and PERSIANN-CDR are similar to those of IMERG, only results of IMERG and CMORPH-CRT are shown. To avoid the influence of interpolation during the calculation of the three indicators (POD, FAR and CSI), the threshold for daily gauge rain is set as 0.5 mm/d since the precision of rain gauge measurements in China is 0.1 mm/h, as suggested by previous studies (Takido et al., 2016; Ward et al., 2011; Dinku et al., 2011; Zhou et al., 2008; Yang et al., 2006). It should be noted that this threshold is only used here.

IMERG performed better than CMORPH-CRT as a whole with higher POD and lower FAR. In the northwestern part of Sichuan province, both POD and FAR are low for CMORPH-CRT, indicating that the uncertainty of this product is high in this region. In view of CSI, CMORPH-CRT outshines IMERG in the southwest part, which is caused by low POD and high FAR of CMORPH-CRT. The average values of the three

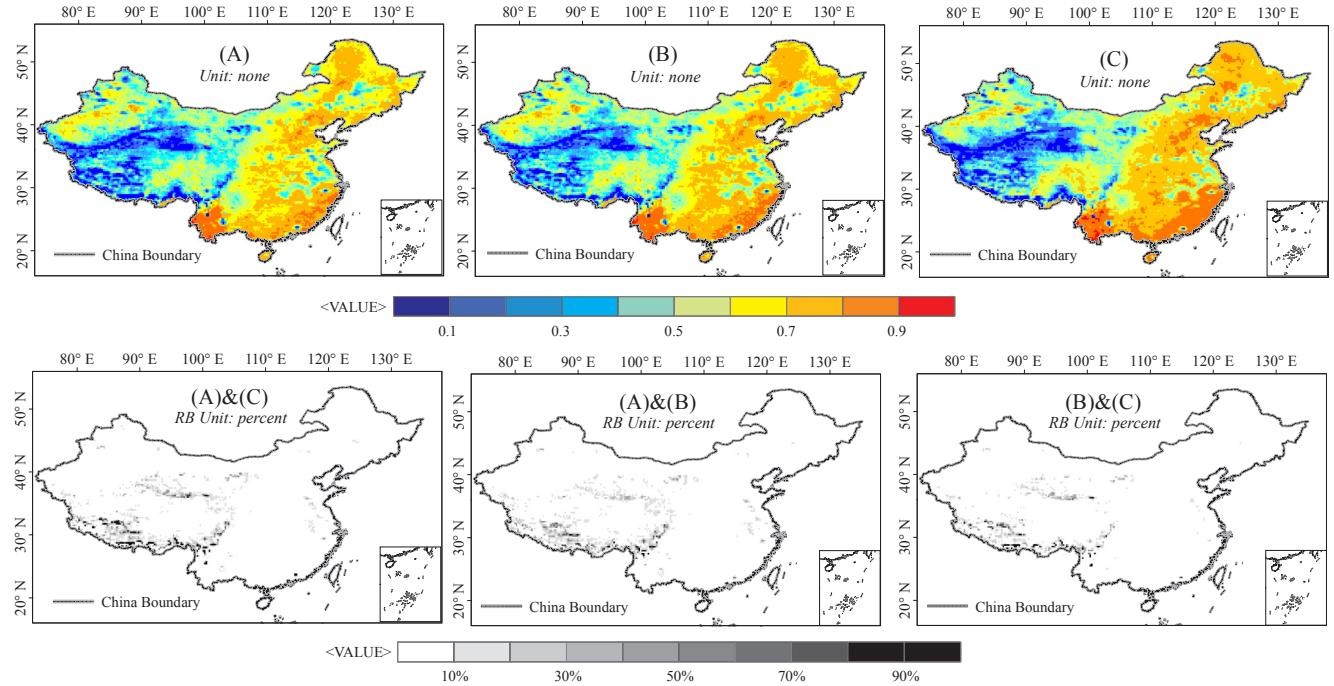


Fig. 2. The CC values of CMORPH-CRT are shown in the first row, using three replacement values: 10^{-9} for (A), 10^{-6} for (B) and 10^{-3} for (C), respectively. The relative bias (RB) of each two CC results is also shown in the second row. Title “(A) & (C)” manifests that the value is given by $\frac{|A - C|}{A}$ in percent unit. By that analogy, “(A)&(B)” refers to $\frac{|A - B|}{A}$, “(B) & (C)” refers to $\frac{|B - C|}{B}$.

indicators are shown in Fig. 4. Results illustrate that all satellite data sets show high POD (> 0.6) and low FAR (< 0.4). IMERG shows better performance than the others. Overall, the quality of satellite data sets over South China is good.

The validation results are shown in Table 2. After filtering out grid cells without gauges, there remain about 1100 grid cells over South China. Both the additive and multiplicative error models are applied in the validation process to assess the robustness of the TC method. To measure how well the results (CC and RMSE) of TC approximate that of the traditional method, the goodness of fit (also shown in CC) between results of two methods for both error models are also given. TC and traditional methods exhibit similar mean CC and RMSE values with each other. Nevertheless, by applying the multiplicative model, the goodness of fit of results between TC and traditional methods is

significantly improved compared with the additive model. For example, the value associated with CC for PERSIANN-CDR is improved from 0.84 to 0.96; the value associated with RMSE for IMERG is improved from 0.82 to 0.90. Results based on the multiplicative model are more consistent with the traditional method than the additive model for all products. Therefore, the multiplicative model is more suitable in this study. However, it should be noted that the TC method using the multiplicative model tends to generate better CC and RMSE compared with the traditional method, leading to the risk of overestimating the performance of satellite products. As a whole, the TC method has good reliability and can be applied in the assessment of various satellite products over Mainland China.

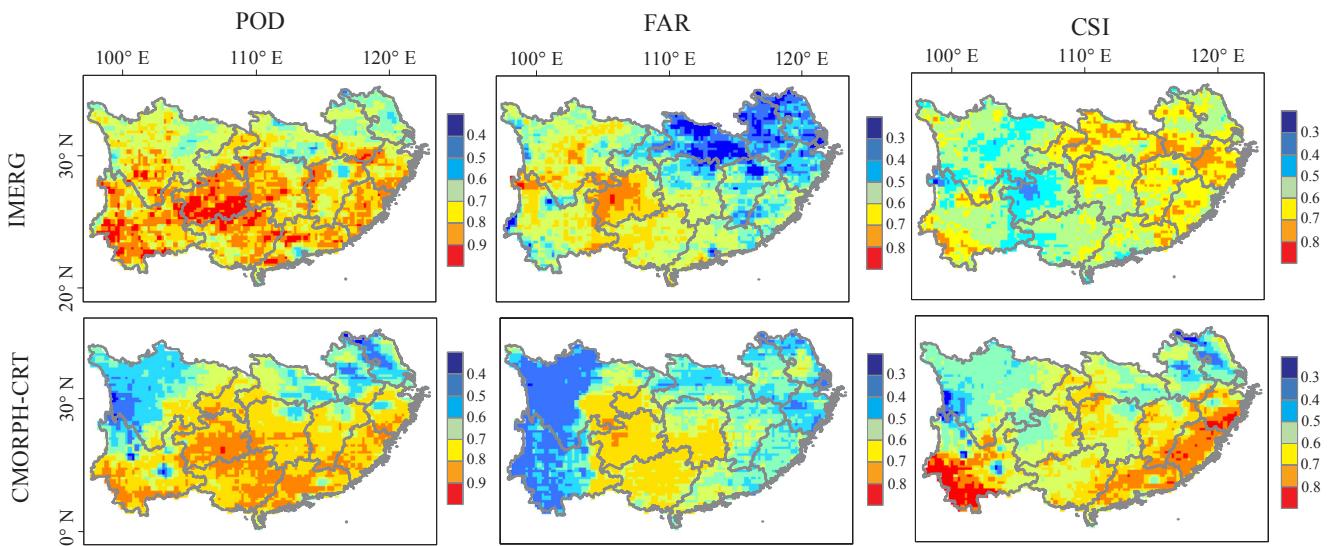


Fig. 3. The distribution of POD, FAR and CSI for IMERG and CMORPH-CRT over South China.

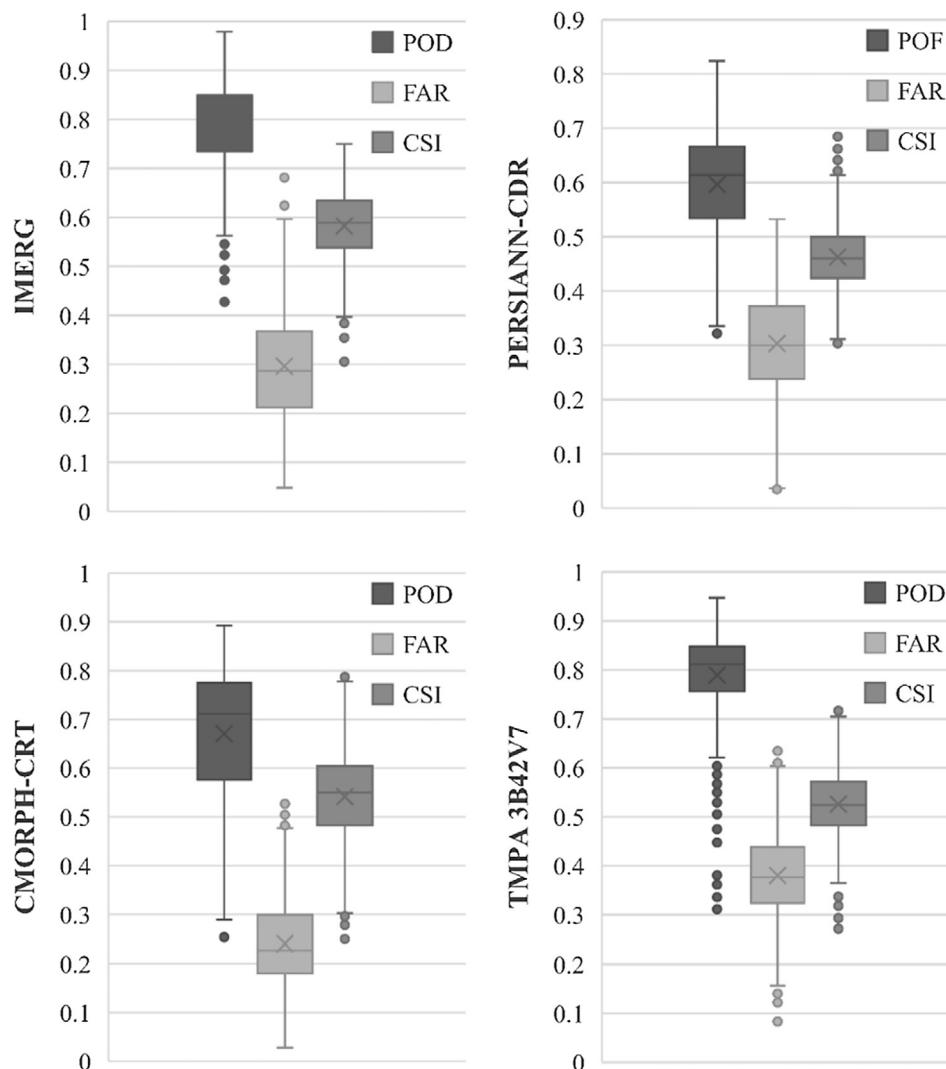


Fig. 4. The box-chart for the three indicators (POD, FAR and CSI) of four satellite products (IMERG, CMORPH-CRT, PERSIANN CDR and TMPA 3B42) in reference to CGDPA over South China. Only grids with at least one gauge are involved.

Table 2
Comparison of CC and RMSE based on TC and traditional methods in South China.

| Satellite Products | PERSIANN-CDR | CMROPH-CRT | IMERG | 3B42RT | 3B42V7 |
|---------------------|---|-------------|-------------|-------------|-------------|
| Mean CC | Traditional Method 0.61 TC method | 0.70 | 0.72 | 0.67 | 0.70 |
| Additive | 0.54 | 0.71 | 0.70 | 0.63 | 0.62 |
| <i>CC</i> | <i>0.84</i> | <i>0.86</i> | <i>0.84</i> | <i>0.80</i> | <i>0.82</i> |
| Multiplicative | 0.70 | 0.78 | 0.80 | 0.74 | 0.78 |
| <i>CC</i> | <i>0.96</i> | <i>0.90</i> | <i>0.91</i> | <i>0.91</i> | <i>0.88</i> |
| RMSE (Unit: mm/day) | Traditional Method 7.23 TC Method | 7.54 | 5.43 | 6.41 | 6.72 |
| Additive | 6.81 | 7.22 | 5.29 | 6.43 | 5.87 |
| <i>CC</i> | <i>0.82</i> | <i>0.85</i> | <i>0.82</i> | <i>0.84</i> | <i>0.80</i> |
| Multiplicative | 6.54 | 7.00 | 4.93 | 5.93 | 5.74 |
| <i>CC</i> | <i>0.88</i> | <i>0.92</i> | <i>0.90</i> | <i>0.87</i> | <i>0.84</i> |

Notation: The CC and RMSE shown here are the mean result value of all about 1100 grid cells over South China. The goodness of fit (also shown in CC) between the result of additive and multiplicative models of TC and traditional methods is given in the table in italics.

5. Multi-source products evaluation using triple collocation

Considering the availability of different satellite products, two study time periods are designed. The longer period is from March 2013 to March 2015, during which 3B42RT, 3B42V7, CMORPH-CRT, and PERSIANN-CDR are available. The shorter period is from March 2014 to March 2015 because IMERG is released from March 2014.

5.1. Evaluation over Mainland China

TC-based CC and RMSE of five satellite products along with CGDPA and ERA-interim are shown in Fig. 4. Based on functions (7)–(9) and functions (11)–(13), larger CC is accompanied by smaller RMSE, which is in consistence with the results shown in the first two columns. Over Mainland China, all five satellite products share relative similar distribution of CC and perform reasonably well in the east over the west. Overall, IMERG shows the best performance of the five satellite products with highest mean CC value of 0.72, followed by 0.62 for CMORPH-CRT, 0.60 for 3B42V7, 0.58 for 3B42RT and 0.44 for PERSIANN-CDR.

3B42RT and 3B42V7 exhibit similar CC and RMSE in most parts of China with a slight difference over the west. This is reasonable since these two products share the same remote sensing inputs and only have differences in the producing process (Huffman et al., 2007). In addition,

by comparison with neighboring grid cells, we can find some anomalously lower CC values over about 350 grid cells along with higher RMSE and average daily precipitation amount for 3B42RT, 3B42V7 and CMORPH-CRT. Further investigation indicates that 94% of irregular grid cells are inland water bodies. Tian et al. (2007) found TRMM 3B42V6 and CMORPH tend to overestimate rainfall amount over inland water bodies with about twice as many raining days than land over the southeastern United States (SEUS). Tang et al. (2016a,b) reported the systematical overestimation over inland water bodies of 3B42RT and 3B42V7 in high mountain Asia (HMA), while IMERG products eliminate the problem. In this study, raining days of these grid cells over water bodies are found about 1.5 times more than those over land, causing a high false alarm rate over water bodies. This is the main reason to the anomalous CC values over certain grid cells for 3B42RT, 3B42V7 and CMORPH-CRT.

PERSIANN-CDR shows a relatively poorer performance than the other satellite products. The mean CC value is 0.56 over Mainland China. Moreover, in the western and northwestern regions, the agreement between average daily rainfall of CGDPA and PERSIANN-CDR is weaker. Previous studies found that PERSIANN-CDR generally tended to underestimate the rainfall amount over arid or semi-arid regions, such as the TP and Taklamakan desert, that is mainly due to the inherent uncertainty of the employed IR data (Ashouri et al., 2015; Miao et al., 2015). While, other satellite products concerned in this study are generally motivated by PMW data with higher accuracy than the IR data (Huffman and Bolvin, 2015), which may be the main factor for the relatively poorer performance of PERSIANN-CDR. Besides, the advantage of PERSIANN-CDR lies in climate studies with a long time period (usually up to five years) compared to that with a short time period (Zhu et al., 2016a,b; Miao et al., 2015).

In addition, an interesting feature that can be found in the results is a significant reduction of CC in the red box marked area in Fig. 5. The marked region is located at the plateau area of Sichuan province on the map, which is the south part of the Hengduan Mountains where the second ladder of China terrain is laid. Therefore, the area is characterized by drastic topography change, which contributes to the significant reduce of CC for all products, including ground observations (CGDPA) and the reanalysis product (ERA-interim). In addition, CC of all products is significantly lower alongside the upper boundary of TP than the other area, where the Kunlun Mountains are laid. It can be concluded that all products fail to provide high-quality precipitation estimates or observations in the very complex mountains. Such phenomenon can hardly be identified using traditional evaluation methods because there are generally few gauges in regions with drastic elevation change, and the ground-based product self also suffers from the problem.

ERA-interim suffers less to terrain change compared with satellite products. The reanalysis data involved both temperature and humidity information, reducing the impact of topography to some degrees (Dee et al., 2011). Moreover, ERA-interim presents much better performance over high latitude region, like the Takalamaham desert compared with satellite products and CGDPA. This further proves the high accuracy of ERA-interim reanalysis product over high latitude regions (Bosilovich et al., 2008). Su et al. (2006) conclude that reanalysis products are valuable for studying the high-latitude water cycle, due to the employment of historical data.

5.2. Stability of the Triple Collocation method

The TC method requires the inputs of three collocated data sets. However, different combinations may influence the outputs (CC and RMSE) for the same product belonging to different triplets. If the uncertainty is large, the effectiveness of the TC method could be diminished. Here, using the variable control method, we investigate the stability of the TC method through comparing CC of CGDPA based on four different combinations by changing the input satellite product

(Fig. 6). Triplet (A) is CGDPA, ERA-interim plus IMERG; Triplet (B) is CGDPA, ERA-interim plus CMORPH-CRT; Triplet (C) is CGDPA, ERA-interim plus PERSIANN-CDR; Triplet (D) is CGDPA, ERA-interim plus TMPA-3B42V7.

Regardless of the triplets, TC results shown in Fig. 6 indicate that CGDPA is the most accurate product with mean TC-based CC close to 0.70 over Mainland China which further illustrates that CGDPA can be used as a benchmark to satellite estimates. Its correlation pattern performs well in eastern parts where rain gauges are denser whereas the TC-based correlation is significantly lower in the western part of TP. This is reasonable given that CGDPA still contains errors due to the limitation of rain gauge density while TC should theoretically provide the correlation with respect to the truth.

Meanwhile, the agreement of TC results for the four triplets is high in the eastern part of Mainland China both in spatial distribution and value. However, the results vary significantly in West China, especially over the upper boundary of the TP. In West China, various factors, such as higher altitude and much more complex topography than the relative homogeneous eastern part of Mainland China, jointly contribute to the degraded quality of all related products. The mean CC values over this region are 0.52 for triplet (A), 0.43 for triplet (B), 0.32 for triplet (C) and 0.36 for triplet (D), respectively. The difference in the results may have something to do with the quality of satellite products since CGDPA and ERA-interim are employed in all triplets. Results also represent that the input satellite product with better quality outputs better TC-based correlation for CGDPA.

Based on Eqs. (11)–(13), the results of CC for CGDPA should be identical regardless of the input satellite products. While this conclusion is made only when the error orthogonality and zero error cross-covariance assumptions are met, which does not generally hold for typical remote sensing products. Yilmaz and Crow (2014) reported that the breach of zero error cross-covariance contributes significantly to TC bias while this influence can be reduced via proper rescaling based on ground observations. In this study, the grid cell is set as square 0.25° which is in consistent with the spatial resolution of CGDPA to reduce the bias.

5.2.1. Comparison over the Tibetan Plateau

As Earth's third pole, the TP and adjacent mountain regions play an essential role in the downstream hydrology and climate, as a controlling factor for the Asia monsoon system. In addition, it's also a typical region where rain gauge distribution is extremely sparse. Previous studies that evaluate the performance of satellite or reanalysis products are subject to various sources of uncertainty due to the lack of sufficient ground observations (Wang and Zeng, 2012; Tong et al., 2014; Ma et al., 2016). In contrast, TC provides an alternative without the need for true observations and thus is born for such ungauged regions. In this section, the uncertainties of different precipitation products over the TP is further investigated based on the TC method. The number of gauges significantly decreases in the TP with only 254 gauges, just 9% to the total gauges in the whole China. Therefore, we evaluate the performance of multi-source products over the grid cells with and without gauges, respectively.

To further ensure the reliability of the TC method over the TP, CC and RMSE of five satellite products and ERA-interim based on TC and traditional methods are compared over grid cells with at least one gauge from the same angle in the validation section. The validation results are shown in Table 3. Analogously, TC and traditional methods exhibit approximate mean CC and RMSE values. The CC between the result of TC and traditional methods of all products is lower than that in South China, which is mainly due to the decrease in the length of data as only 248 grid cells contain gauge stations. While it still illustrates the high agreement between results of the two methods. In sum, the TC method has good reliability and can be applied over the TP (Table 4).

To specify the distribution of CC and RMSE, the cumulated distribution function (CDF) curves for all precipitation products in TP over

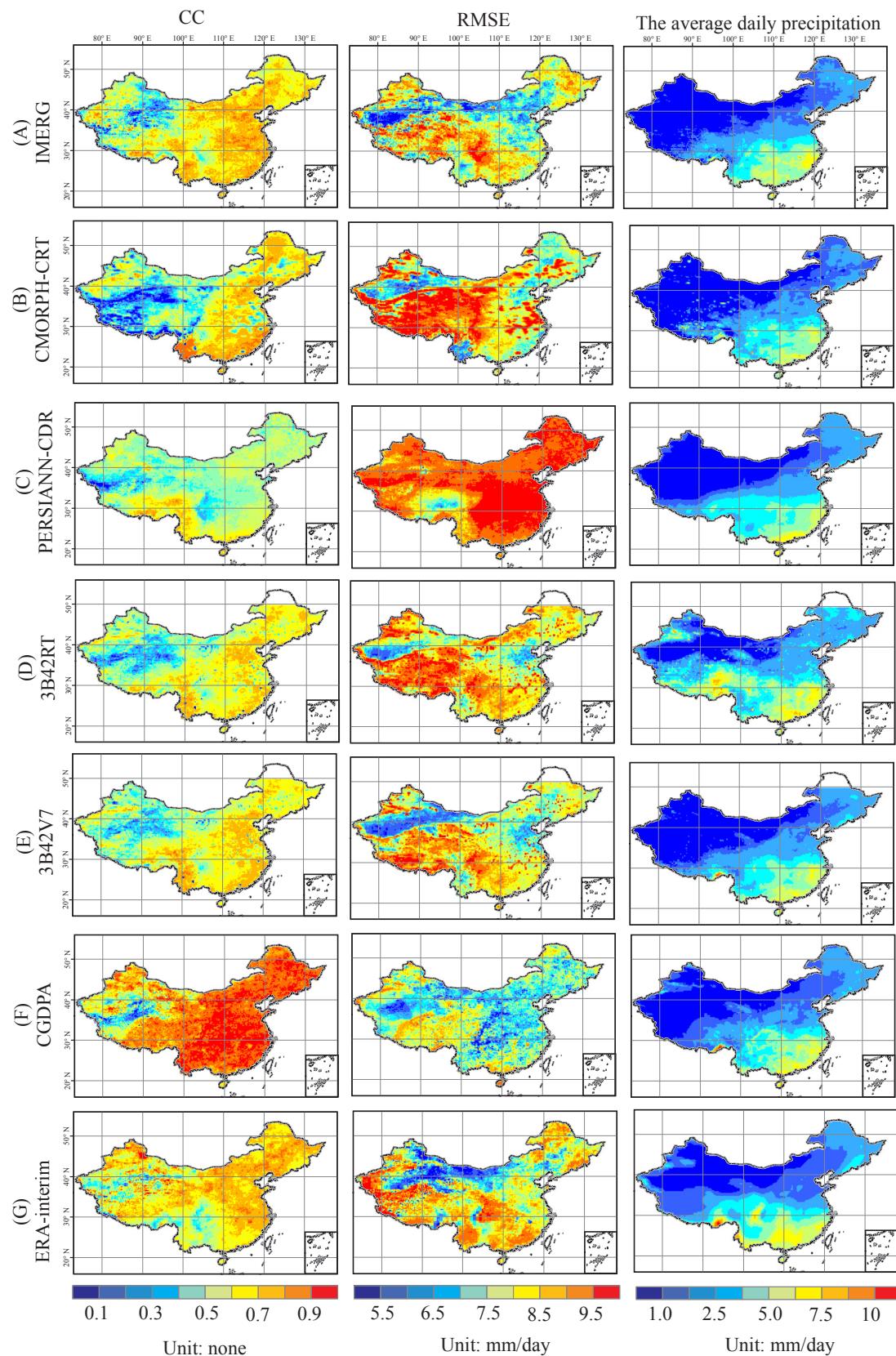
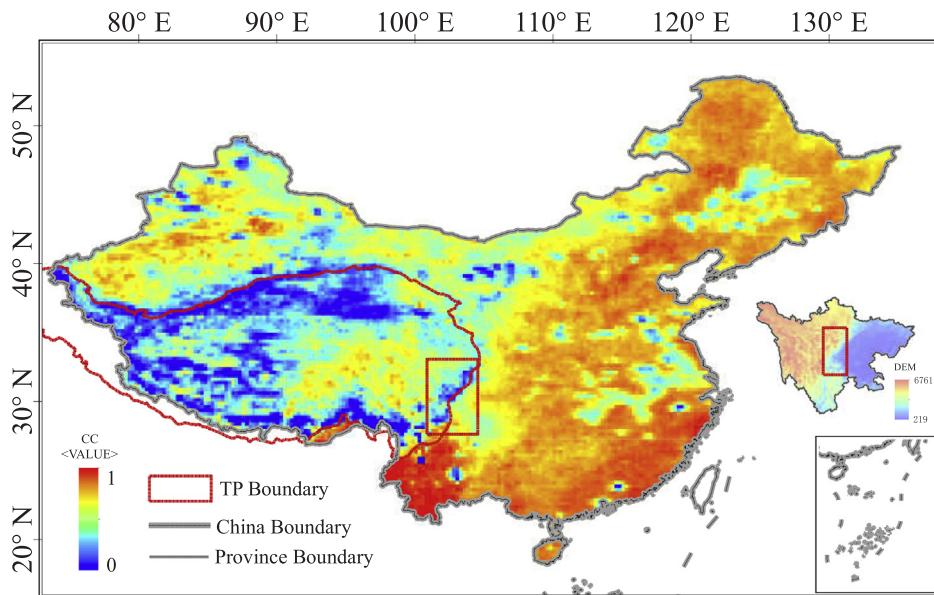


Fig. 5. The outputs (CC and RMSE) of the TC method and gridded daily mean precipitation of five satellite products along with CGDPA and ERA-interim. The first column shows the CC for each product and the second column includes the RMSE. The average daily precipitation maps are given in the third column for comparison. Each three graphs in the same row represents the result of one.



gauged and ungauged regions are calculated and shown in Fig. 7 with black and gray dashed line representing CGDPA and ERA-interim. Meanwhile, the CDF of average daily rainfall for all products are also presented. In statistics, x values (CC) corresponding to 60% probability can be chosen as the indicators for evaluation (Tuller and Brett, 1984), referred as $x_{60\%}$ hereafter. In the TP, number of gauged grid cells (254) is far smaller than that of ungauged ones (3833). Therefore, it is reasonable that the CDF curves over ungauged cells (Fig. 7a and b) are more continuous and smoother than the gauged ones (Fig. 7c and 7d). Since the results of RMSE is consistent with those of CC, we only discuss the CDF curve for CC (Fig. 7a and b).

Over gauged region, the highest CC of $x_{60\%}$ is 0.83 for CGDPA, followed by ERA-interim (0.65), IMERG (0.65), 3B42V7 (0.63), 3B42RT (0.62) and PERSIANN-CDR (0.60) with the lowest 0.40 for CMORPH-CRT (Fig. 7b). Overall, CGDPA presents better performance than the other products. However, its quality drastically degrades over ungauged region. The probability of CC under 0.5 increases from 8% to about 40% along with RMSE of $x_{60\%}$ varying from 6.2 mm/d to 8.9 mm/d. The performance of CGDPA is also much poorer in the west and mid-west part (see in Fig. 4f) where the Kunlun Mountains are laid. Results reveal the inherent uncertainty of the ground-based observation CGDPA over TP, which is mainly due to the lack of sufficient gauge data and the influence of topography. Furthermore, the main rain type varies largely from orographic in the south to convective in the central and north of TP. The interpolation method applied in CGDPA does not well reflect the spatial variability of precipitation which may be the leading factor to its degraded quality (Fig. 8).

Table 3
Comparison of CC and RMSE based on TC and traditional methods over the TP.

| Satellite Products | IMERG | CMORPH-CRT | PERSIANN-CDR | 3B42RT | 3BRTV7 | ERA |
|--------------------|---|-------------|--------------|-------------|-------------|-------------|
| CC | Traditional Method 0.63 TC method | 0.30 | 0.50 | 0.48 | 0.51 | 0.60 |
| Multiplicative | 0.60 | 0.33 | 0.56 | 0.57 | 0.59 | 0.60 |
| <i>CC</i> | <i>0.91</i> | <i>0.92</i> | <i>0.90</i> | <i>0.86</i> | <i>0.88</i> | <i>0.89</i> |
| RMSE | Traditional Method 8.41 TC Method | 9.83 | 8.94 | 9.40 | 8.77 | 8.45 |
| Multiplicative | 8.33 | 9.80 | 8.79 | 9.10 | 8.69 | 8.24 |
| <i>CC</i> | <i>0.88</i> | <i>0.91</i> | <i>0.88</i> | <i>0.90</i> | <i>0.85</i> | <i>0.88</i> |

Notation: The CC and RMSE shown here are the mean result value of 248 grid cells with at least one gauge over the TP. The CC between the result of TC (using multiplicative error model) and traditional methods is given in the table in italics.

Fig. 6. The CC of CMORPH-CRT as an example to show the reducing region for all products. The region is marked out in red dashed line in the graph. The small figure to the right is the DEM of Sichuan province with the inside red dashed frame corresponding to the same one in CC distribution map. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

IMERG surpasses other satellite products with CC of $x_{60\%}$ to be 0.65, followed by 3B42V7 (0.63), 3B42RT (0.62), PERSIANN-CDR (0.60) and CMORPH-CRT (0.40). CMORPH-CRT shows the relative poorest performance with 70% of the grid cells containing CC value under 0.4. The quality of satellite products also degrades over the ungauged regions while the gap between CGDPA is reduced.

For the reanalysis product, ERA-interim shows a good performance next to CGDPA. The difference in results under two conditions is not much, which is reasonable since the independence between CGDPA and ERA-interim is better than that of the satellite products. Nevertheless, the CC value of ERA-interim even outshines CGDPA over the upper boundary of the TP. ERA-interim only employs temperature and humidity information to retrieve precipitation. Thus, the influence of topography is reduced to some degree so that it's able to better characterize the variability of precipitation in TP.

In regard to the CDF of average daily rainfall, the performance of all products shown here is in consistent with that based on the CDF of CC and RMSE. Over gauged region where the quality of CGDPA can be better guaranteed, bias between CGDPA and other products is mainly found over grid cells with light rain (under 2 mm/d). This contributes as the main error source for other products. While, over ungauged region, IMERG, ERA-interim and CGDPA exhibit similar CDF curve with higher difference over slightly heavier rainfall (over 2 mm/d) areas. Moreover, the rainfall measured by CMORPH-CRT is generally lower than that of other products, which may be one of the many reasons why CMORPH-CRT performs the relative lowest CC shown in Fig. 7(a).

Table 3 shows the mean CC value of all precipitation products over

Table 4

The mean CC over gauged and ungauged regions in the Tibetan Plateau.

| | IMERG | CMORPH-CRT | PERSIANN-CDR | 3B42RT | 3B42V7 | ERA | CGDPA |
|----------|---------------------|------------|--------------|--------|--------|------|-------|
| Gauged | 0.60 | 0.33 | 0.56 | 0.57 | 0.59 | 0.60 | 0.80 |
| Ungauged | 0.63 | 0.30 | 0.50 | 0.47 | 0.51 | 0.61 | 0.58 |
| | RMSE (unit: mm/day) | | | | | | |
| Gauged | 8.33 | 9.80 | 8.79 | 9.10 | 8.69 | 8.24 | 5.93 |
| Ungauged | 8.42 | 9.83 | 8.95 | 9.42 | 8.78 | 8.46 | 8.47 |

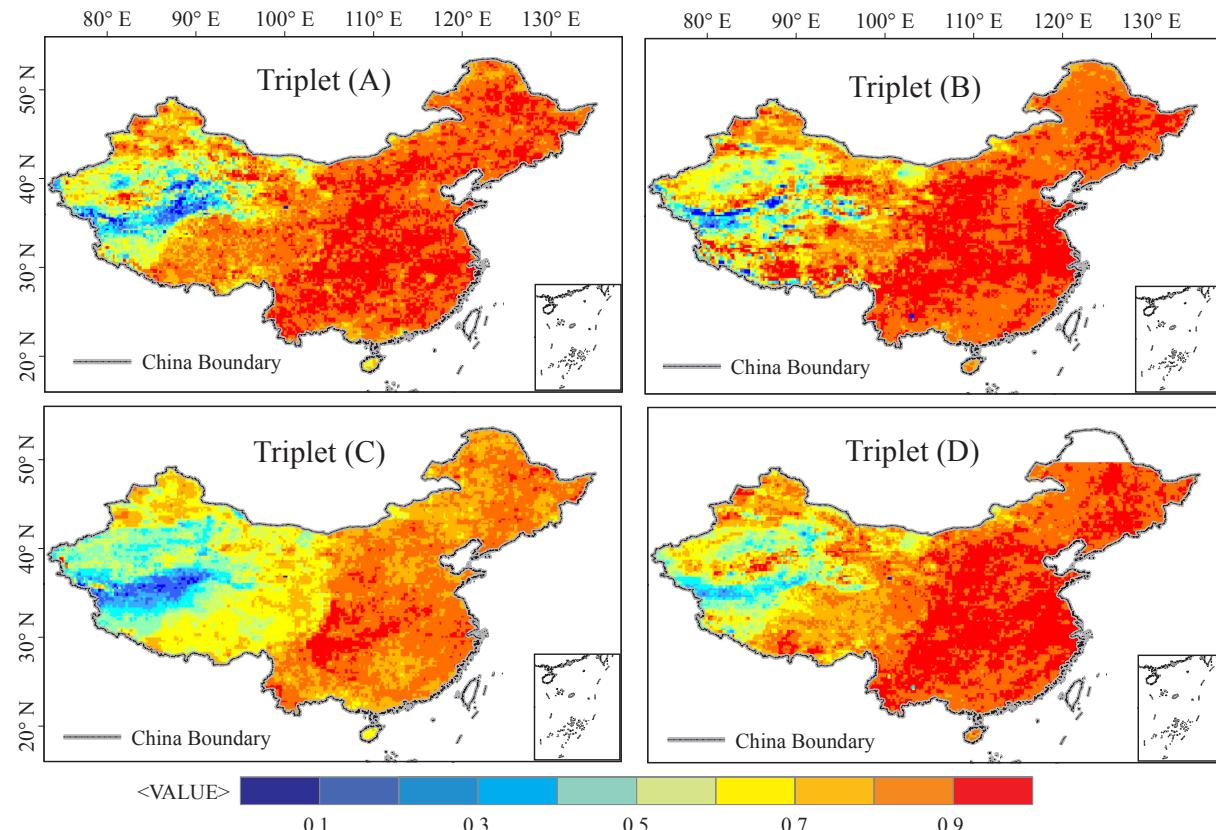


Fig. 7. CC of CGDPA based on four triplets. CGDPA and ERA-interim are employed in all triplets. The third product is IMERG for Triplet (A), CMORPH-CRT for Triplet (B), PERSIANN-CDR for Triplet (C), and TMPA-3B42V7 for Triplet (D).

gauged and ungauged regions in the TP, including ground-based CGDPA and reanalysis ERA-interim. Over gauged regions, CGDPA is the best product with the highest CC (0.80) and lowest RMSE (5.93 mm/d). However, over the ungauged regions, the indicators for CGDPA drops to 0.58 for CC and 8.47 mm/d for RMSE.

IMERG along with ERA-interim slightly outshine CGDPA with a mean CC value of 0.63 and 0.61, respectively. In spite of the potential overestimation due to the usage of multiplicative model, results still demonstrates the expected usage of satellite-based and reanalysis products and the inherent uncertainty of ground observation over ungauged regions due to the interpolation using sparsely gauge data. It reminds us that evaluation of satellite precipitation products only based on interpolated data may suffer from large uncertainties when the gauge density is low.

6. Summary and conclusions

In this study, we apply the TC method to evaluate five satellite-based precipitation products, and assess uncertainties of satellite-based, ground-observed and model reanalysis data sets over Mainland China for the first time. In summary, we can draw the conclusions that:

1. The TC method is effective in evaluating precipitation products through comparison with the traditional evaluation method based in South China. TC-based CC and RMSE agree well with the metrics calculated using ground-based rain gauges with the mean relative differences of 9% and 7% for CC and RMSE, respectively. In addition, results based on the multiplicative error model are better than the additive error model but probably overestimate the performance of products used in TC.
2. All the five satellite products perform better in East China than West China mainly due to the topography and climate differences. Overall, IMERG performs the best with the highest mean CC (0.72) over Mainland China, followed by 0.62 for CMORPH-CRT, 0.60 for 3B42V7, 0.58 for 3B42RT and 0.44 for PERSIANN-CDR. Satellite, ground, and reanalysis data sets all exhibit degraded quality in regions with drastic topography change according to the TC method, which can hardly be revealed by traditional evaluation methods due to sparse gauges.
3. Grid cells containing inland water bodies have systematic anomalies of precipitation overestimation for TMPA-3B42, 3B42RT and CMORPH-CRT. The TC method can reveal this phenomenon in the absence of the ground observations, which in turn verifies the

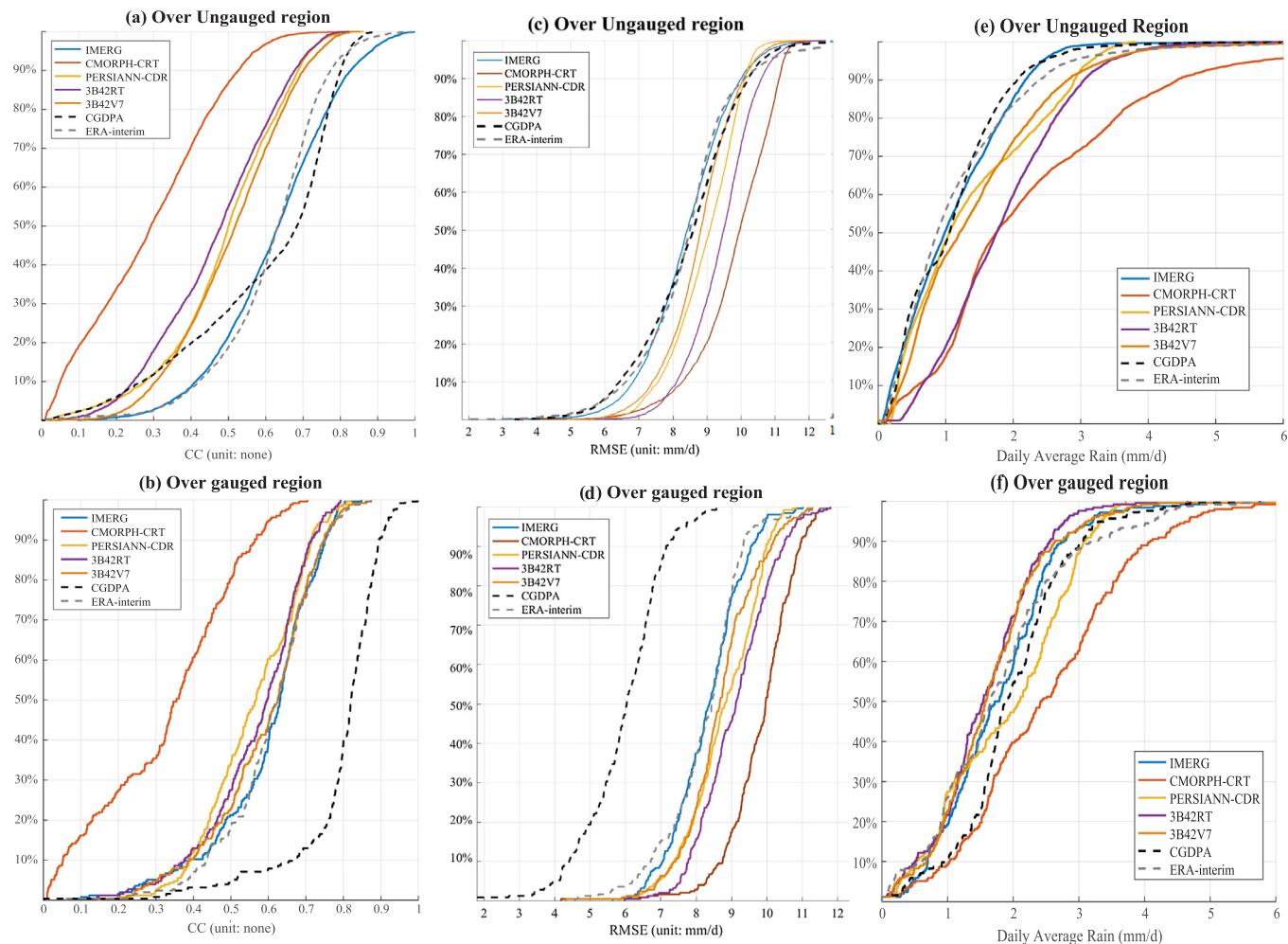


Fig. 8. The Cumulated Distribution Function (CDF) curve of CC (a and b), RMSE (c and d) and Daily Average rain amount (e and f) in the TP over gauged and ungauged regions, respectively.

effectiveness of TC.

4. The quality of CGDPA significantly degrades over grid cells without gauges in comparison to those with gauges. The probability of CC under 0.5 increases from 8% to about 30% for CGDPA. While ERA-interim performs nearly identical results in both conditions. Overall in the TP, IMERG and ERA-interim slightly outshine CGDPA over ungauged grid cells with mean CC value of 0.60 and 0.63, indicating the potential of remote sensing and reanalysis products over sparsely gauged regions.

The TC method provides us with a new perspective for the evaluation of precipitation estimates without the ground truth. Thus, the application of TC-based results for merging multi-source products is promising and will be addressed in the future. Also, new strategies for reducing the bias of TC-based results due to the violation against its mathematical assumptions should be considered.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grant NO.7146101701) and China National 973 Project (Grant NO.2013CB036406). The authors' great gratitude is extended to the China Meteorological Administration for providing ground-based rainfall data. The efforts of research communities are also highly appreciated for making all the data available for international users.

References

- Alemohammad, S.H., et al., 2015. Characterization of precipitation product errors across the United States using multiplicative triple collocation. *Hydrol. Earth Syst. Sci.* 19 (8), 3489–3503.
- Ashouri, H., et al., 2015. PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Am. Meteorol. Soc.* 96 (1), 69–83.
- Behrangji, A., et al., 2011. Hydrologic evaluation of satellite precipitation products over a mid-size basin. *J. Hydrol.* 397 (3–4), 225–237.
- Bosilovich, M.G., et al., 2008. Evaluation of global precipitation in reanalyses. *J. Appl. Meteorol. Climatol.* 47 (9), 2279–2299.
- Caires, S., 2003. Validation of ocean wind and wave data using triple collocation. *J. Geophys. Res.* 108 (C3).
- Chen, S., et al., 2013. Evaluation of the successive V6 and V7 TRMM multisatellite precipitation analysis over the Continental United States. *Water Resour. Res.* 49 (12), 8174–8186.
- Dee, D.P., et al., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137 (656), 553–597.
- Dinku, T., Ceccato, P., Grover-Kopec, E., et al., 2007. Validation of satellite rainfall products over East Africa's complex topography[J]. *Int. J. Remote Sens.* 28 (7), 1503–1526.
- Dinku, T., Ceccato, P., Connor, S.J., 2011. Challenges of satellite rainfall estimation over mountainous and arid parts of east Africa. *Int. J. Remote Sens.* 32 (21), 5965–5979.
- Dorigo, W.A., et al., 2010. Error characterisation of global active and passive microwave soil moisture datasets. *Hydrol. Earth Syst. Sci.* 14 (12), 2605–2616.
- Draper, C., et al., 2013. Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sens. Environ.* 137, 288–298.
- Duzheng, Y., Youxi, G., 1979. Meteorology of the Tibetan Plateau. Scientific Publication Agency Google Scholar, Peking, China.
- Ebert, E.E., 2007. Methods for verifying satellite precipitation estimates. In: *Measuring precipitation from space*. Springer, pp. 345–356.
- Fang, H., Jiang, C., Li, W., et al., 2013. Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and

- theoretical uncertainties[J]. *J. Geophys. Res.: Biogeosci.* 118 (2), 529–548.
- Gao, Y.C., Liu, M.F., 2013. Evaluation of high-resolution satellite precipitation products using rain gauge observations over the Tibetan Plateau. *Hydrol. Earth Syst. Sci.* 17 (2), 837–849.
- Gerapetritis, H., Pelissier, J., 2004. The critical success index and warning strategy. In: 17th Conference on Probability and Statistics in the Atmospheric Sciences, Seattle, 2004.
- Germann, U., et al., 2006. Radar precipitation measurement in a mountainous region. *Q. J. R. Meteorol. Soc.* 132 (618), 1669–1692.
- Goudenhoofdt, E., Delobbe, L., 2009. Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrol. Earth Syst. Sci.* 13 (2), 195–203.
- Gruber, A., et al., 2016. Recent advances in (soil moisture) triple collocation analysis. *Int. J. Appl. Earth Obs. Geoinf.* 45, 200–211.
- Hain, C.R., et al., 2011. An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling. *J. Geophys. Res.* 116 (D15).
- Hodges, K.I., Hoskins, B.J., Boyle, J., et al., 2003. A comparison of recent reanalysis datasets using objective feature tracking: Storm tracks and tropical easterly waves[J]. *Mon. Weather Rev.* 131 (9), 2012–2037.
- Hong, Y., et al., 2006. Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. *Water Resour. Res.* 42 (8).
- Hossain, F., Anagnostou, E.N., 2006. A two-dimensional satellite rainfall error model. *IEEE Trans. Geosci. Remote Sens.* 44 (6), 1511–1522.
- Hsu, H.-H., Liu, X., 2003. Relationship between the Tibetan Plateau heating and East Asian summer monsoon rainfall. *Geophys. Res. Lett.* 30 (20).
- Huffman, G.J., Bolvin, D.T., 2015. Real-time TRMM multi-satellite precipitation analysis data set documentation[J]. NASA Tech. Doc.
- Huffman, G.J., et al., 2014. NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD), NASA/GSFC, Greenbelt, MD, USA.
- Huffman, G.J., et al., 2010. The TRMM multi-satellite precipitation analysis (TMPA). In: Satellite Rainfall Applications for Surface Hydrology. Springer, pp. 3–22.
- Huffman, G.J., et al., 2007. The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* 8 (1), 38–55.
- Inoue, T., Matsumoto, J., 2004. A comparison of summer sea level pressure over East Eurasia between NCEP–NCAR reanalysis and ERA-40 for the period 1960–99. *J. Meteorol. Soc. Jpn. Ser. II* 82 (3), 951–958.
- Janowiak, J.E., Joyce, R.J., Yarosh, Y., 2001. A real-time global half-hourly pixel-resolution infrared dataset and its applications. *Bull. Am. Meteorol. Soc.* 82 (2), 205–217.
- Joyce, R.J., et al., 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.* 5 (3), 487–503.
- Ma, Y., et al., 2016. Similarity and error intercomparison of the GPM and its predecessor TRMM multisatellite precipitation analysis using the best available hourly gauge network over the Tibetan Plateau. *Remote Sens.* 8 (7).
- Marshall, G.J., 2002. Trends in Antarctic geopotential height and temperature: A comparison between radiosonde and NCEP–NCAR reanalysis data. *J. Clim.* 15 (6), 659–674.
- Massari, C., Crow, W., Brocca, L., 2017. An assessment of the accuracy of global rainfall estimates without ground-based observations. *Hydrol. Earth Syst. Sci. Discuss.* 1–24.
- McColl, K.A., et al., 2014. Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophys. Res. Lett.* 41 (17), 6229–6236.
- McCollum, J.R., et al., 2002. Evaluation of biases of satellite rainfall estimation algorithms over the continental United States. *J. Appl. Meteorol.* 41 (11), 1065–1080.
- Miao, C., et al., 2015. Evaluation of the PERSIANN-CDR Daily Rainfall Estimates in Capturing the Behavior of Extreme Precipitation Events over China. *J. Hydrometeorol.* 16 (3), 1387–1396.
- Miralles, D.G., Crow, W.T., Cosh, M.H., 2010. Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations. *J. Hydrometeorol.* 11 (6), 1423–1429.
- Portabella, M., Stoffelen, A., 2009. On Scatterometer Ocean Stress. *J. Atmos. Oceanic Technol.* 26 (2), 368–382.
- Ratheesh, S., et al., 2013. Assessment of satellite-derived sea surface salinity in the Indian Ocean. *IEEE Geosci. Remote Sens. Lett.* 10 (3), 428–431.
- Roebeling, R.A., et al., 2012. Triple collocation of summer precipitation retrievals from SEVIRI over Europe with gridded rain gauge and weather radar data. *J. Hydrometeorol.* 13 (5), 1552–1566.
- Scott, K.A., Buehner, M., Carrieres, T., 2014. An assessment of sea-ice thickness along the Labrador coast from AMSR-E and MODIS data for operational data assimilation. *IEEE Trans. Geosci. Remote Sens.* 52 (5), 2726–2737.
- Serreze, M.C., Barrett, A.P., Lo, F., 2005. Northern high-latitude precipitation as depicted by atmospheric reanalyses and satellite retrievals. *Mon. Weather Rev.* 133 (12), 3407–3430.
- Shen, Y., Xiong, A., 2016. Validation and comparison of a new gauge-based precipitation analysis over mainland China. *Int. J. Climatol.* 36 (1), 252–265.
- Shen, Y., et al., 2014. Uncertainty analysis of five satellite-based precipitation products and evaluation of three optimally merged multi-algorithm products over the Tibetan Plateau. *Int. J. Remote Sens.* 35 (19), 6843–6858.
- Sorooshian, S., et al., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Am. Meteorol. Soc.* 81 (9), 2035–2046.
- Stoffelen, A., 1998. Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys. Res. Oceans* 103 (C4), 7755–7766.
- Su, F., et al., 2006. Evaluation of surface water fluxes of the pan Arctic land region with a land surface model and ERA-40 reanalysis. *J. Geophys. Res.* 111 (D5).
- Takido, K., Valeriano, O.C.S., Ryo, M., et al., 2016. Spatiotemporal evaluation of the gauge-adjusted global satellite mapping of precipitation at the basin scale. *J. Meteorol. Soc. Jpn. Ser. II* 94 (2), 185–195.
- Tang, G., Long, D., Hong, Y., 2016a. Systematic anomalies over inland water bodies of high mountain asia in TRMM precipitation estimates: no longer a problem for the GPM Era? *IEEE Geosci. Remote Sens. Lett.* 13 (12), 1762–1766.
- Tang, G., et al., 2017. Can near-real-time satellite precipitation products capture rainstorms and guide flood warning for the 2016 summer in South China? *IEEE Geosci. Remote Sens. Lett.* 14 (8), 1208–1212.
- Tang, G., et al., 2016b. Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. *J. Hydrol.* 533, 152–167.
- Tian, Y., Peters-Lidard, C.D., 2010. A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.* 37 (24).
- Tian, Y., Peters-Lidard, C.D., Choudhury, B.J., et al., 2007. Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications [J]. *J. Hydrometeorol.* 8 (6), 1165–1183.
- Tong, K., et al., 2014. Evaluation of satellite precipitation retrievals and their potential utilities in hydrologic modeling over the Tibetan Plateau. *J. Hydrol.* 519, 423–437.
- Tuller, S.E., Brett, A.C., 1984. The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis[J]. *J. Clim. Appl. Meteorol.* 23 (1), 124–134.
- van Dijk, A.I.J.M., Renzullo, L.J., Wada, Y., et al., 2014. A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble[J]. *J. Hydrol. Earth Syst. Sci.* 18 (8), 2955.
- Wang, A., Zeng, X., 2012. Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau. *J. Geophys. Res.* 117 (D5).
- Ward, E., Buytaert, W., Peaver, L., et al., 2011. Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. *Adv. Water Resour.* 34 (10), 1222–1231.
- Xie, P., Arkin, P.A., 1997. Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Am. Meteorol. Soc.* 78 (11), 2539–2558.
- Xie, P., et al., 2011. Bias-corrected CMORPH: A 13-year analysis of high-resolution global precipitation. *Geophys. Res. Abstr.*
- Yanai, M., Li, C., Song, Z., 1992. Seasonal heating of the Tibetan Plateau and its effects on the evolution of the Asian summer monsoon. *J. Meteorol. Soc. Japan. Ser. II* 70 (1B), 319–351.
- Yang, C., Chandler, R.E., Isham, V.S., et al., 2006. Quality control for daily observational rainfall series in the UK. *Water Environ. J.* 20 (3), 185–193.
- Ye, D.Z., Gao, Y.X., 1979. Tibetan Plateau Meteorology[J]. Science, Beijing, pp. 89–101.
- Yilmaz, M.T., Crow, W.T., 2014. Evaluation of assumptions in soil moisture triple collocation analysis. *J. Hydrometeorol.* 15 (3), 1293–1302.
- Yong, B., et al., 2010. Hydrologic evaluation of Multisatellite Precipitation Analysis standard precipitation products in basins beyond its inclined latitude band: A case study in Laohahe basin, China. *Water Resour. Res.* 46 (7).
- Zhou, T., Yu, R., Chen, H., et al., 2008. Summer precipitation frequency, intensity, and diurnal cycle over China: A comparison of satellite data with rain gauge observations. *J. Clim.* 21 (16), 3997–4010.
- Zhu, Q., Xuan, W., Liu, L., et al., 2016a. Evaluation and hydrological application of precipitation estimates derived from PERSIANN-CDR, TRMM 3B42V7, and NCEP-CFSR over humid regions in China. *Hydrol. Process.* 30 (17), 3061–3083.
- Zhu, Q., et al., 2016b. Evaluation and hydrological application of precipitation estimates derived from PERSIANN-CDR, TRMM 3B42V7, and NCEP-CFSR over humid regions in China. *Hydrol. Process.* 30 (17), 3061–3083.
- Zwieback, S., et al., 2012. Structural and statistical properties of the collocation technique for error characterization. *Nonlinear Processes Geophys.* 19 (1), 69–80.