Assessing the adequacy of bias corrected IMERG satellite precipitation estimates using extended mixture distribution mapping method over Yangtze River basin

Qiumei Ma [1], Lihua Xiong [1,a], Chong-Yu Xu [1,2], Shenglian Guo [1]

[1] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China
[2] Department of Geosciences, University of Oslo, P.O. Box 1022 Blindern, N-0315 Oslo, Norway

Abstract. Satellite precipitation estimates (SPE) product with high spatiotemporal resolution is a potential alternative to traditional ground-based gauge precipitation. However, SPE is frequently biased due to its indirect measurement, and thus bias correction is necessary before applying to a specific region. An improved distribution mapping method, i.e., Extended Mixture Distribution (EMD) of censored Gamma and generalized Pareto distributions, was established. The advantage of EMD method is that it describes both moderate and extreme values well and carries on the traditional censored, shifted Gamma distribution to combine the precipitation occurrence/non-occurrence events together. Then the EMD method was applied to the Integrated Multi-satellitE Retrievals for GPM product (IMERG) as statistical post-processing over Yangtze River basin. The Version-2 Gridded dataset of daily Surface Precipitation from China Meteorological Administration (GSP-CMA) was taken as reference. The adequacy of bias corrected IMERG precipitation was assessed and the results showed that (1) the Root Mean Squared Error and Relative Bias between bias-corrected IMERG precipitation and reference are significantly reduced relative to the raw IMERG estimates; (2) the performance of extreme values of IMERG in Yangtze River basin is enhanced since both the under- and over-estimation of the raw IMERG are compromised, due to the generalized Pareto distribution introduced in EMD which is enable to describe the extreme value distribution. This highlights the improved distribution mapping method, EMD is flexible and robust to bias correct the IMERG precipitation to obtain higher accuracy of SPE despite the coarse resolution of reference.

1 Introduction

Satellite precipitation estimates (SPE) are important alternatives to the traditional precipitation measurements and have been increasingly applied to many scientific and social-economic fields, such as, hydro-meteorological and ecological system modeling, flood forecast, water resources development and conservation, point- and nonpoint-source pollutant management [1, 2]. Global Precipitation Measurement (GPM) mission Integrated Multisatellite Retrievals (IMERG) is first GPM-era global precipitation product derived from the newly released space-borne observatory instrument [3]. Other mainstream SPE products include TRMM Multisatellite Precipitation Analysis product (TMPA), Precipitation Estimation from Remote Sensed Information Using Artificial Neural Networks-Cloud Classification System (PERSIANN-CCS), Global Satellite Mapping of Precipitation product (GSMap), and Climate Prediction Center (CPC) morphing technique product (CMORPH). Although IMERG shows better performance than its predecessors since the first release of IMERG data in April

2014 according to extensive studies on evaluating work, some bias or errors inherent in IMERG still cannot be ignored and are frequently more significant than the gauge rainfall observations [4]. To enhance the performance of SPE in various applications, some bias correction methods can serve as statistical post-processing techniques to reduce the systematic and/or random errors of SPE [5, 6].

Several bias correction methods have been applied to SPE products, and they are proved to improve the raw SPE. Habib et al. [7] employed the linear scaling method to reduce the CMORPH product bias and thus improved the precision of CMORPH-driven runoff simulation. Worqlul et al. [8] used a similar method investigating and correcting the bias of Multi-Sensor Precipitation Estimate–Geostationary (MPEG) data. This method can significantly reduce the bias while the simulated flow using bias corrected MPEG data is comparable to that using the gauge rainfall. Vernimmen et al. [9] reduced the Root Mean Square Error (RMSE) for estimates of dry season rainfall of TMPA 3B42RT estimates by applying a single empirical bias correction equation. These methods, among others, range from simple scaling approaches of

---

average, standard deviation and intensity to complicated ones employing distribution mapping or probabilistic weather generators. They were originally designed to adjust GCM data, but can also be used to correct SPE. Distribution mapping (DM) is a relatively comprehensive approach making the simulated or test precipitation values approximating to the quantiles of reference values [10]. A Censored, Shifted Gamma distribution (CSGD) was proposed to describe and map the RCM data [11], and later used to correct TMPA data [12]. CSGD offers more flexibility by considering precipitation occurrence (zero/non-zero precipitation) and precipitation mass together compared to the two-stage approach. However, Gamma distribution, although flexible enough, cannot capture the extreme values of precipitation. It is not quite adequate to model the tail of the distribution since it underestimates large values.

Therefore, an improved DM method, i.e., an Extended Mixture Distribution (EMD) coupling censored, shifted Gamma and generalized Pareto (GP) distribution, was established to better capture both the moderate and extreme values in this study. It couples a Gamma distribution with a GP distribution and shifts the mixture cumulative distribution function (CDF) somewhat to the left. This EMD method is used to bias correct the IMERG Early product during April of 2014 and March of 2017 over the Yangtze River basin in China.

This paper is organized as follows. In section 2, the data used was presented. In section 3, the detailed mathematical descriptions of traditional CSGD and the improved EMD were described. The results were discussed in section 4. Finally, the conclusions were presented in section 5.

2 Data and study area

2.1 IMERG product dataset

As Level 3 product of GPM, the IMERG system merges and interpolates "all" available satellite microwave precipitation estimates, together with microwave-calibrated infrared satellite estimates, ground gauge analyses of the Global Precipitation Climatology Centre (GPCC), and potentially other estimators for the TRMM and GPM eras over quasi-global range. IMERG version 5 (V05) products, used in this study, includes gridded rainfall and snowfall data with 0.1°×0.1° spatial resolution and 30 min temporal resolution. IMERG provides the near-real-time Early and Late run about 4 h and 12 h respectively after observation, and post-real-time Final run about 2 months after the observation month. The Final run incorporates gauge data from GPCC and is characterized by high accuracy but large latency. We used the daily IMERG Early run product (referred as IMERG-DE for simplicity) to avoid the problem of gauge dependence.

2.2 Rain gauge dataset

The Version-2 Gridded dataset daily Surface Precipitation

data interpolated by China Metrological Administration (GSP-CMA) based on observations of national meteorological station was selected as benchmark and used to correct the IMERG data. The IMERG-DE data was aggregated to 0.5°×0.5° cells at space when comparing with the GSP-CMA data.

2.3 Study area

Yangtze River basin, located between 24°28′-35°58′; 90°32′-122°2′, was focused on in this study, because of its representation of various climatologic and topographic conditions. As the largest river basin in China, Yangtze River basin covers a drainage area of 1.8 million km². It originates from the Tibetan Plateau and flows eastward for more than 6300 km before draining into the East China Sea, comprising nearly one-fifth of mainland China and with diverse landforms and complicated hydro-climatic conditions affected by both East and South Asian monsoon activities. Previous studies have indicated that most satellite precipitation products show significant differences between the western and eastern parts of China and application of satellite precipitation products over the western China is difficult [1, 13]. Therefore, Yangtze River basin, relatively lager and complex in space, was selected.

3 Method

The distribution mapping adjusts the probability distribution of SPE or RCM data to that of observed precipitation data by matching the CDF values of two distributions. This process can be expressed mathematically as

$$P_{cor} = F^{-1}[F(P_{SPE} \mid \theta_{SPE}) \mid \theta_{ref}], \qquad (1)$$

where F and F^{-1} are CDF of precipitation (P) and its inverse function, respectively, with parameter set θ; *cor, SPE* and *ref* in subscript represent the corrected precipitation, satellite precipitation (i.e., IMERG-DE) and reference data (i.e., GSP-CMA), respectively. This can be realized by a transfer function to shift the occurrence distributions of precipitation. Because the *Pcor* is derived with quantile of statistical distribution, the DM method was also called 'probability mapping', 'quantile–quantile mapping' (Q-Q mapping) or quantile mapping method.

3.1 Traditional Censored, Shifted Gamma Distribution (CSGD) mapping method

Normally, precipitation is characterized by right-skewed distribution in temporal process. To some extent, it can be assumed that Gamma distribution family is suitable for fitting precipitation events depending on a shape parameter k and a scale parameter ς. On the other hand, precipitation occurrence/nonoccurrence obeys binormial distribution parameterized by probability of precipitation (POP). The CSGD was proposed to jointly modeling precipitation occurrence/nonoccurrence and amount using

a continuous distribution that permits negative values, and left-censoring it at zero [11, 14]. The CDF of CSGD abovementioned can be written as left-shifted Gamma CDF:

$$F(x) = \begin{cases} F_{k,\varsigma}(x+\delta) & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}, \qquad (2)$$

where x is the depth of precipitation P. The corresponding probability density function (PDF) of CSGD when x greater or equal to zero is written

$$f(x) = \frac{1}{\Gamma(k)\varsigma^k}(x+\delta)^{k-1}e^{-\frac{(x+\delta)}{\varsigma}}; \qquad (3)$$

$$x \geq 0; \quad k, \varsigma > 0$$

where $\Gamma(k)$ represents Gamma function.

3.2 Improvement using extended mixture distribution

Although the Gamma distribution model is frequently used to fit precipitation events, the extremely high part probably fails to capture by Gamma model because of its statistical property. For this reason, some extreme value distributions (e.g. generalized Pareto distribution used in this study) were recommended to be merged into the initial Gamma distribution to adequately consider both low and moderate and high precipitation intensities. The Gamma distribution, F(x) in Equ. (2) was then replaced by a two-component mixture distribution, i.e., Gamma and generalized Pareto distributions, with an auxiliary weighting parameter, λ :

$$f(x) = (1-\lambda)\times f_1[(x+\delta)\,|\,k,\varsigma]+\lambda\times f_2[(x+\delta)\,|\,\xi,\sigma];$$

$$k,\varsigma,\xi,\sigma,\delta > 0, \text{ and } 0 \leq \lambda \leq 1$$
$$, \qquad (4)$$

where $f_2()$ is the PDF of generalized Pareto:

$$f_2(x) = \frac{1}{\sigma}[1+\frac{\xi(x+\delta)}{\sigma}]^{(-\frac{1}{\xi}-1)}. \qquad (5)$$

3.3 Performance metrics

To evaluate the performance of raw and corrected IMERG precipitation vs. reference GSP-CMA data comprehensively, seven widely used statistical metrics were selected, including Pearson correlation coefficient (r), Root Mean Squared Error (RMSE) and Relative Bias (RB), for quantifying differences in performance between the raw and corrected IMERG runs. Their definitions are as follows:

$$r = \frac{\text{cov}(P_{SPE}, P_{ref})}{\sqrt{\text{var}(P_{SPE})}\sqrt{\text{var}(P_{ref})}}, \qquad (6)$$

where $\text{cov}()$ and $\text{var}()$ denote the covariance between P_{SPE} and P_{ref}, and variance of P_{SPE} or P_{ref}.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_{SPE}-P_{ref})^2}{n}}, \qquad (7)$$

$$RB = \frac{\sum_{i=1}^{n}(P_{ref}-P_{SPE})}{\sum_{i=1}^{n}P_{ref}}, \qquad (8)$$

where P_{SPE} and P_{ref} are the respective precipitation intensity from IMERG (including raw and corrected data) and reference precipitation at *ith* time step with n pairs of data.

4 Results and discussion

4.1 Statistical evaluations on the raw IMERG data

The daily average precipitation of raw IMERG and GSP-CMA data over Yangtze River basin was showed in Fig. 1(a) and (b). The spatial pattern of raw IMERG is similar with that of reference with increasing precipitation from west to east roughly, which was demonstrated earlier [1, 2]. Overall, the range of areal average IMERG was 2.76 mm/d, slightly less than that of reference (3.09 mm/d). The difference between the raw IMERG vs. reference was described in Fig. 1(c) and (d) using quantitative indices defined in Section 3.3. The mean and median values of RMSE are 0.78 and 0.66 mm/d respectively. There was some difference between these two datasets, and the bias varies depending on the spatial location. In general, the RMSE shows larger values in the west of Yangtze River basin, compared to the middle and east parts, while the RB shows more positive values in the west than in the middle and east. There was a significant tendency that IMERG data estimates more low-intensity rain events (drizzle) in the west and middle parts of Yangtze River basin, while generally overestimates the precipitation intensity over the east region (Fig. 1(d)). This result can be partly attributed to greater uncertainty resulting from the spare intensities of national meteorological stations across the upstream area of Yangtze River due to high altitude. Another region that the bias shows larger values is around the boundary of study area. This is associated with techniques of analyzing data which results in not all the IMERG grids around boundary can match to a GSP-CMA grid.
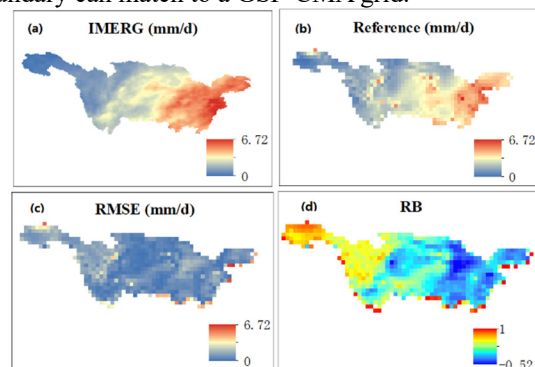

Fig. 1 The daily precipitation of the raw IMERG (a) vs. reference (GSP-CMA) data (b); RMSE and RB indices between the raw IMERG and reference data for (c) and (d) respectively.

4.2 Statistical evaluations on bias corrected IMERG data

Fig. 2 shows the spatial pattern of corrected IMERG and indices between corrected IMERG vs. reference. The range of areal average for the corrected IMERG was 2.81 mm/d, slightly high than that of the raw (2.76 mm/d) and reaching closely to the reference. The biases were found similar to those of the raw. The mean and median values of RMSE are 0.71 and 0.53 mm/d respectively, which are significantly reduced compared to the corresponding indices of raw IMERG (0.78 and 0.66 mm/d respectively).
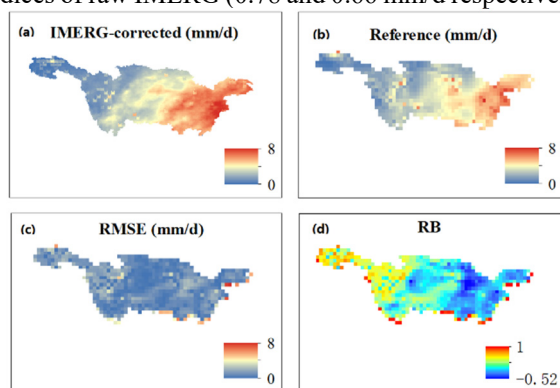

Fig. 2 The daily precipitation of the corrected IMERG (a) vs. reference (GSP-CMA) data (b); RMSE and RB indices between the corrected IMERG and reference data for (c) and (d) respectively.

In addition, a comparison of corrected IMERG data to the raw data was conducted. The variability of indices of two products against reference in space average was presented by boxplot in Fig. 3. The interquartile ranges of RMSE and RB are both reduced after the bias correction process. For the RMSE index, the interquartile range related to the corrected data is 0.24—0.91 mm/d, narrower than that of the raw (0.29—1.01 mm/d). While for the RB index, the interquartile range related to the corrected data is -0.086—0.316 mm/d, also narrower than that of the raw (-0.090—0.398 mm/d). It's obvious from Fig. 1(d) and Fig. 2(d) that the overestimates in the raw IMERG data across the west part of study area are compromised by the bias correction. From the view of r index, there is no significant difference between corrected and raw IMERG vs. reference (both near to 0.6 of r values). From the basic statistics of IMERG estimates and reference data summarized in Fig. 3, the improved bias correction method, EMD performs well for enhancing the performance of SPE data.
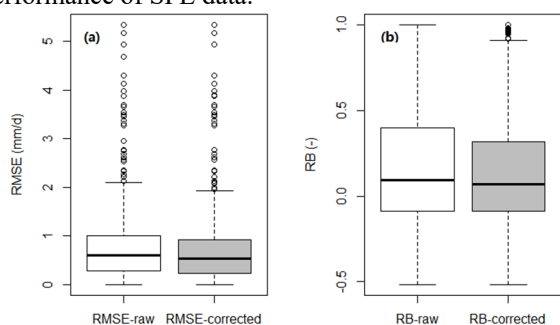

Fig. 3 The spatially aggregated spread of RMSE and RB indices for the raw (a) and corrected IMERG (b) cases.

5 Conclusions

An improved distribution mapping method of bias correction, i.e., an Extended Mixture Distribution method (EMD) combining censored, shifted Gamma with generalized Pareto distributions, was proposed and applied to adjust the IMERG satellite precipitation estimates over Yangtze River basin. The performance metrics exhibit different spatial patterns. Overall the IMERG data underestimates precipitation in the west region of study area compared to the reference data, while overestimates the precipitation in the east part. The under- and over-estimation of IMERG are compromised after using EMD to bias correct. In addition, the mean, median and interquartile range of performance metrics between the corrected IMERG vs. reference are all smaller than those between the raw IMERG vs. reference.

The EMD bias correction method is proved to reduce the bias of IMERG data, especially for adjusting the extreme value bias. The advancement of EMD method was seen in this study, which can operate on both moderate and extreme precipitation bias correction by introducing Gamma and generalized Pareto distributions in the mixture distribution. This method can also be used to bias correct other satellite precipitation products and reproduce spatiotemporal precipitation information with high precision. Further work is need to evaluate the hydrological utility of a bias corrected product of IMERG Early runs and other version dataset in the Yangtze River basin.

Acknowledgements

References:

1. Wang, C., et al., Global intercomparison and regional evaluation of GPM IMERG Version-03, Version-04 and its latest Version-05 precipitation products: Similarity, difference and improvements. Journal of Hydrology, **564**: p. 342-356, (2018)

2. Tang, G., et al., Similarities and differences between three coexisting spaceborne radars in global rainfall and snowfall estimation. Water Resources Research, **53**(5): p. 3835-3853, (2017)

3. Foelsche, U., et al., Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria. Hydrology and Earth System Sciences, **21**(12): p. 6559-6572, (2017)

4.   Worqlul, A.W., et al., Comparison of rainfall estimations by TRMM 3B42, MPEG and CFSR with ground-observed data for the Lake Tana basin in Ethiopia. Hydrology and Earth System Sciences, **18**(12): p. 4871-4881, (2014)

5.   Yang, Z., et al., Bias adjustment of satellite-based precipitation estimation using gauge observations: A case study in Chile. Journal of Geophysical Research: Atmospheres, **121**(8): p. 3790-3806, (2016)

6.   Condom, T., P. Rau and J.C. Espinoza, Correction of TRMM 3B43 monthly precipitation data over the mountainous areas of Peru during the period 1998–2007. Hydrological Processes, **25**(12): p. 1924-1933, (2011)

7.   Habib, E., et al., Effect of bias correction of satellite-rainfall estimates on runoff simulations at the source of the Upper Blue Nile. Remote Sensing, **6**(7): p. 6688-6708, (2014)

8.   Worqlul, A.W., et al., Performance of bias corrected MPEG rainfall estimate for rainfall-runoff simulation in the upper Blue Nile Basin, Ethiopia. Journal of Hydrology, **556**: p. 1182-1191, (2018)

9.   Vernimmen, R., et al., Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia. Hydrology and Earth System Sciences, **16**(1): p. 133-146, (2012)

10. Shin, J., et al., Bias correction of RCM outputs using mixture distributions under multiple extreme weather influences. Theoretical and Applied Climatology, **2018**: p. 1-16, (2018)

11. Scheuerer, M. and T.M. Hamill, Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. Monthly Weather Review, **143**(11): p. 4578-4596, (2015)

12. Wright, D.B., D.B. Kirschbaum and S. Yatheendradas, Satellite Precipitation Characterization, Error Modeling, and Error Correction Using Censored Shifted Gamma Distributions. Journal of Hydrometeorology, **18**(10): p. 2801-2815, (2017)

13. Ning, S., et al., Error analysis and evaluation of the latest GSMap and IMERG precipitation products over Eastern China. Advances in Meteorology, **2017**. (2017)

14. Baran, S. and D. Nemoda, Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting. Environmetrics, **27**(5): p. 280-292, (2016)