Response to review on manuscript: [hess-2020-294]

Previous title: Global component analysis of errors in five satellite-only global precipitation estimates

Current title: Global component analysis of errors in three satellite-only global precipitation estimates

Journal: Hydrology and Earth System Sciences

Corresponding author: Bin Yong

Authors: Hanqing Chen, Bin Yong*, Pierre-Emmanuel Kirstetter, Leyang Wang, Yang Hong

Dear Editor, Graham Jewitt:

We would like to thank you for giving us the opportunity to revise our paper (**Major Revision**) and thank two anonymous reviewers for their careful reviews and their valuable comments. In the revision, we have addressed all the requests and comments of the reviewers point by point. For convenience, the responses to the comments were written down below in blue font. The modifications made in the revised text were highlighted in red font. We believe that the current version is suitable for publication in Hydrology and Earth System Sciences.

If you have any questions about this manuscript, please do not hesitate to let me know.

Sincerely yours,

Dr. Bin Yong

Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Institute for Disaster Risk Management, School of Geographical Science, Nanjing University of Information Science and Technology, Nanjing 210000, China.

Email: yongbin_hhu@126.com

Response for Reviewer #1

Understanding the error components of satellite products is very important to the algorithm developers and end-users. This manuscript presents the error analysis of five commonly used products. The finding is useful for the development on the future satellite retrieval algorithms. The topic is attractive and suitable for HESS.

Response: We would like to sincerely thank the reviewer for his/her valuable comments and suggestions.

Major comments

1. **Given that the performance of IMERG-late is very similar to the performance of IMERG-early, I suggest removing one of them.**

Response: Thanks. We have removed the IMERG-Early and GSMaP-NRT in the revised version considering their similar performance relative to their corresponding late-level products.

2. **The dataset for China and for the rest of the world are different. I suggest authors discuss the effect of this inconsistency on the conclusion.**

Response: Good suggestion. In this study, these two different reference datasets were used to explore the error components of SPPs. The calculation processes of the spatial maps of biases and systematic errors include two major steps:

(1) Global maps of biases and systematic error

a. The spatial maps of biases and systematic error computed over the global land areas use CPCU data as the benchmark. Maps of biases and systematic error over mainland China use CGDPA data as the reference.
b. The spatial maps of biases and systematic error over mainland China replace the Chinese mainland part of the global land areas.

(2) Error components of SPPs under different precipitation intensities

The satellite and ground reference precipitation values over the globe (except for mainland China) were stored in vector S1 and vector G1, respectively. Then the satellite and ground reference precipitation values over mainland China were stored in vector S2 and vector G2, respectively. Third, S1 and S2 form a new vector S ($S = [S1, S2]^T$), while G1 and G2 form a new vector G ($G = [G1, G2]^T$). Finally, the SPP error components under different precipitation intensities were computed by using equations (1)-(9).

The inconsistency of reference data has little impact on the evaluation results. We have discussed this in **lines 141-144**. The calculation processes were provided as supplementary materials, **see supplementary material**.

3. **Section 4.1 Potential for the transferability of the regional assessment results to other areas. I don't think that comparison between CONUS and China is a good example to show the transferability. Although these two regions have similarities in coverage area, and latitude range, the topography characteristics and climate regimes are very different. In addition, these two regions are very large and have dramatic heterogeneity in land surface property and climate, which is not suitable for the comparison**

Response: We thank the reviewer for this comment. We agree that the characteristics in terms of topography and climate regimes are different over CONUS and China, making them challenging for discussing the transferability. In our revision, we have selected the Chinese Fujian (FJ) and Zhejiang (ZJ) provinces to intercompare and analyze the transferability of evaluation results. These two adjacent areas are located in southeastern China and having similar topography and climate. Spatial maps of the total bias for the three SPPs over FJ and ZJ are shown in the following Fig. 1. One can see that the spatial maps of the total biases for each SPP show significant differences for all four seasons between the FJ and ZJ regions. Besides, we found that the total biases in the FJ (or ZJ) areas show evident differences. This indicates that the evaluation results between these similar areas cannot be extended to one another. It might be due to the large differences in performance between the various satellite sensors and differences in satellite samples used in the satellite precipitation retrieval systems between the areas. Also, it might be also caused by factors (e.g. characteristics of precipitation regimes, such as precipitation types) not captured by satellites or the reference. We have discussed this issue in the revised version, **see section 4.1**.

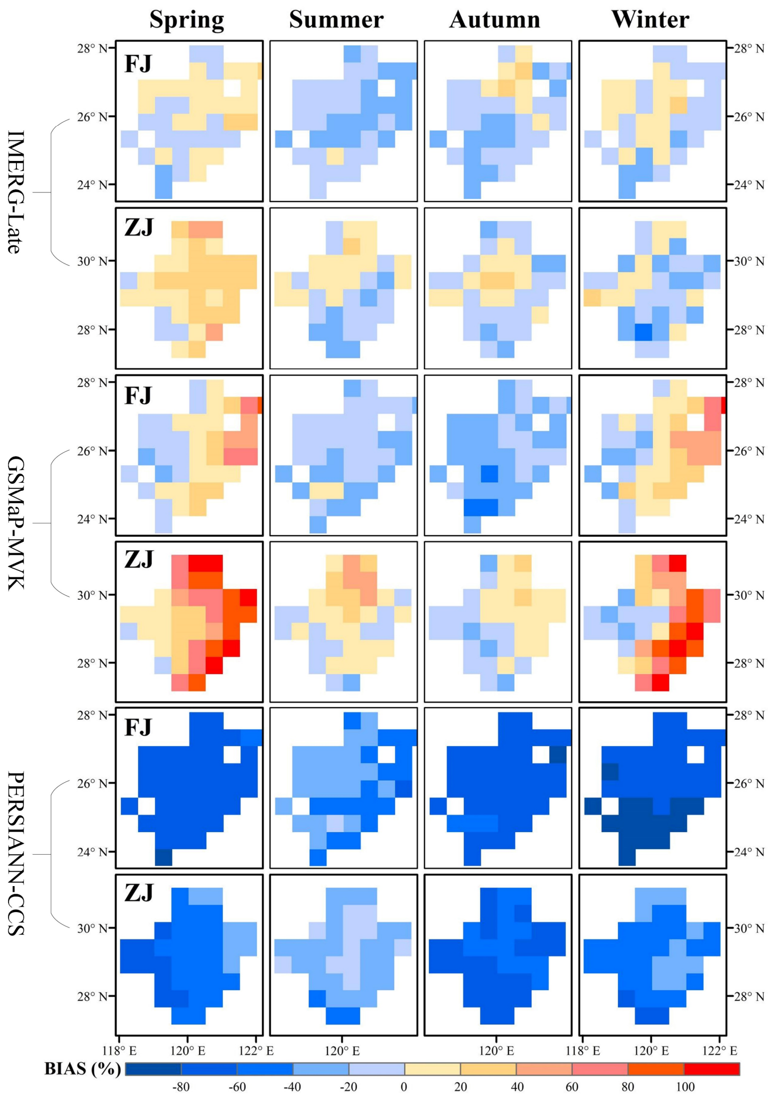

Fig. 1 Spatial maps of total bias for the three SPPs for four seasons over the Fujian (FJ) and Zhejiang (ZJ) provinces, respectively.

4. Figure 9: What is the sample size of each category? What is the uncertainty of these results?

Response: Thanks. We provided the gauges available for each category as shown in the following Figure 2. One can see that the gauge number in each category is over 135, which suggests that the sample size in each class is sufficient to support the reliability and robustness of the results. We have added the related descriptions in section 3.3, **see lines 319-321**.

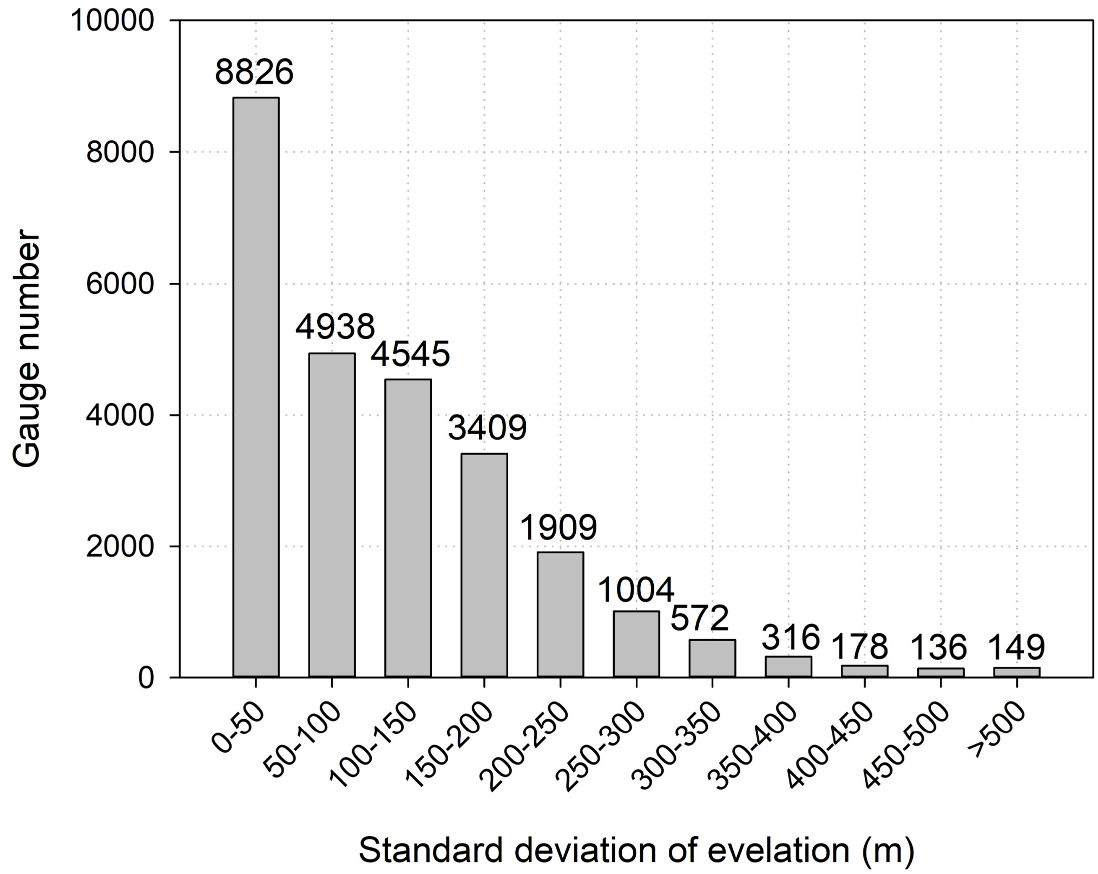

Fig. 2 Gauge number for each topography class.

5. **Figure 10 shows that the relation between the normalized error component (NEC) and elevation is very similar to that between the system error and elevation. What is the added value of NEC? What is significance of the index? Please clarify this.**

Response: Good questions. The relationship between the normalized error component (NEC) and topography is very similar to that between the system error and topography.

It is due to close values of mean precipitation in all topography categories. This phenomenon can be explained by the following equation for NEC (also see equation (10) in the manuscript)

$$NEC = \left(\sum_{i=1}^{n}(\hat{S} - G)^2\right) / \left(\left(\sum_{i=1}^{n}(S - G)^2\right) \times \bar{G}\right)$$

The relationship between NEC and topography degenerates into the relationship between systematic error and topography when the mean precipitation (i.e., $\bar{G}$) of all topography classes is very close. However, the high similarity does not mean that the NEC score is meaningless for investigating the impact of topography on the systematic error. In practice, the purpose of the NEC is to exclude the impact of other factors such as precipitation intensity on the systematic error. It works only when the rainfall intensities are significantly different. We have clarified the significance of this metric in section 2.3.2, **see lines 211-214**.

Minor comments

1. **Line 17: IMERG-late, not MERG-late**

Response: Thanks, we have revised this mistake, **see lines 20**.

2. **Line 333: "false bias" should be "miss bias".**

Response: Thanks, we have changed 'false bias' to 'miss bias', **see line 299**.

3. **Lines 365: "increase" should be "decrease".**

Response: Thanks, we have replaced the word 'increase' with 'decrease', **see line 325**.

4. **Figure 1: add the unit for density of rain gauge, add the source of precipitation data (1891-2018), and delete the legend on distance.**

Response: We have revised the Figure 1 according to this comment. The source of precipitation data (1891-2018) has provided in the description, **see Fig.1 and lines 689-690**.

5. **Figures 2-5: the objective of this study is to compare bias among different products. It could be more visually distinct to put the same bias category (for example, total bias) of different products in a column or row.**

Response: We have completely revised the Figures 2-5 according to the reviewer's suggestion, **see Figs. 2-5 in the revised version**.

6. **Figure 7: delete "false bias" in the description, as no false bias is presented in the figure.**

Response: We have deleted 'false bias' in the description, **see line 717**.

In addition, we did the experiment again because we found that the previous results in mainland China were from mismatched satellite-ground observations (CGDPA) pairs, and all figures and tables have been revised. Other revisions were highlighted in red font.

Thanks for your review, we look forward to your positive response.

Response for reviewer #2

Although many publications already exist in the literature, errors and uncertainties associated with global satellite precipitation products are still difficult to characterize. Thus, I believe this work could be of interest to the HESS readership and worth publication. However, I have a few comments that I would like the authors to consider before accepting the manuscript for publication.

Response: Many thanks for the important comments and suggestions from this reviewer.

1. **First off, I recommend revising the language, since there are a few grammatical mistakes**

Response: The manuscript has been carefully checked and revised to avoid all grammatical mistakes.

2. **Second, I suggest clarifying the main goal of this work. There are currently 5 goals mentioned at the end of the Introduction that read more like tasks. I think that a more focused article would be more effective. In other words, is the goal to validate the SPPs? Is it to model their errors/uncertainties? Is it to investigate what are the factors causing more errors/uncertainty in one location/product/season vs. another? Is it to inform users and algorithm developers on how to use/improve such products?**

Response: Thank you for the suggestions. We have further clarified the main goal of this work at the end of the Introduction. The revised content shows:

The objectives of this study are four-fold: (1) to reveal the major components of errors (including total bias and total mean squared error) for three SPPs including IMERG Late run (IMERG-Late), GSMaP Microwave-IR Combined Product (GSMaP-MVK), and PERSIANN-CCS for four seasons across global land areas; (2) to investigate the potential for the transferability of the regional assessment results to other similar regions; (3) to investigate what are the factors causing large systematic errors; (4) to inform users and algorithm developers on how to improve these satellite precipitation products.

The detailed revision can be found in **lines 109-116.**

3. **In this regard, the abstract should be more concise and highlight the main goals and findings of this work**

Response: We have further refined the abstract to highlight the main goals and findings of this work, **see Abstract**.

4. **Is including both IMERG Early and Late really necessary? The algorithm is same and – as expected – their performance very similar. Same goes for GSMaP-NRT and GSMaP-MVK. This may help with my comment above of a more focused article.**

Response: Thank you. In fact, the performance of IMERG-Early is similar to that of IMERG-Late. According to the reviewer's suggestion, we have removed the IMERG-Early and GSMaP-NRT in our revised manuscript.

5. **The section on "transferability of the regional assessment results to other areas" is weak and not well justified. However, this part of the study is also one of the most interesting, since improving our knowledge of how SPPs perform in regions of the world where no ground observations are available can be extremely useful (e.g., hydrologic applications). Ground observations are mainly available in plain areas (and sparse vegetation). Thus, how can we generalize such results to densely vegetated and highly complex regions?**

Response: Thanks. We entirely agree with the reviewer's opinion. The discussion on transferability of the evaluation results seems weak and is not well justified. However, this part of the study is rather important. Therefore, in the revision we chose two suitable areas (i.e., Chinese Fujian (FJ) and Zhejiang (ZJ) provinces) for comparing and investigating the potential for transferability of the regional assessment results to other areas. These two selected adjacent areas, which are located in southeastern China, have similar topography and climate. The spatial distribution of the total bias for each SPP over FJ and ZJ are given in the following Fig.1. The spatial distributions of the total biases exhibit significant differences between FJ and ZJ, suggesting that the evaluation results between similar areas cannot be extended to one another. This indicates that the evaluation results between these similar areas cannot be extended to one another. It might be due to the large differences in performance between the various satellite sensors and differences in satellite samples used in the satellite precipitation retrieval systems between the areas. Also, it might be also caused by factors (e.g. characteristics of precipitation regimes, such as precipitation types) not captured by satellites or the reference. We have discussed this issue in the revised version, **see section 4.1**.

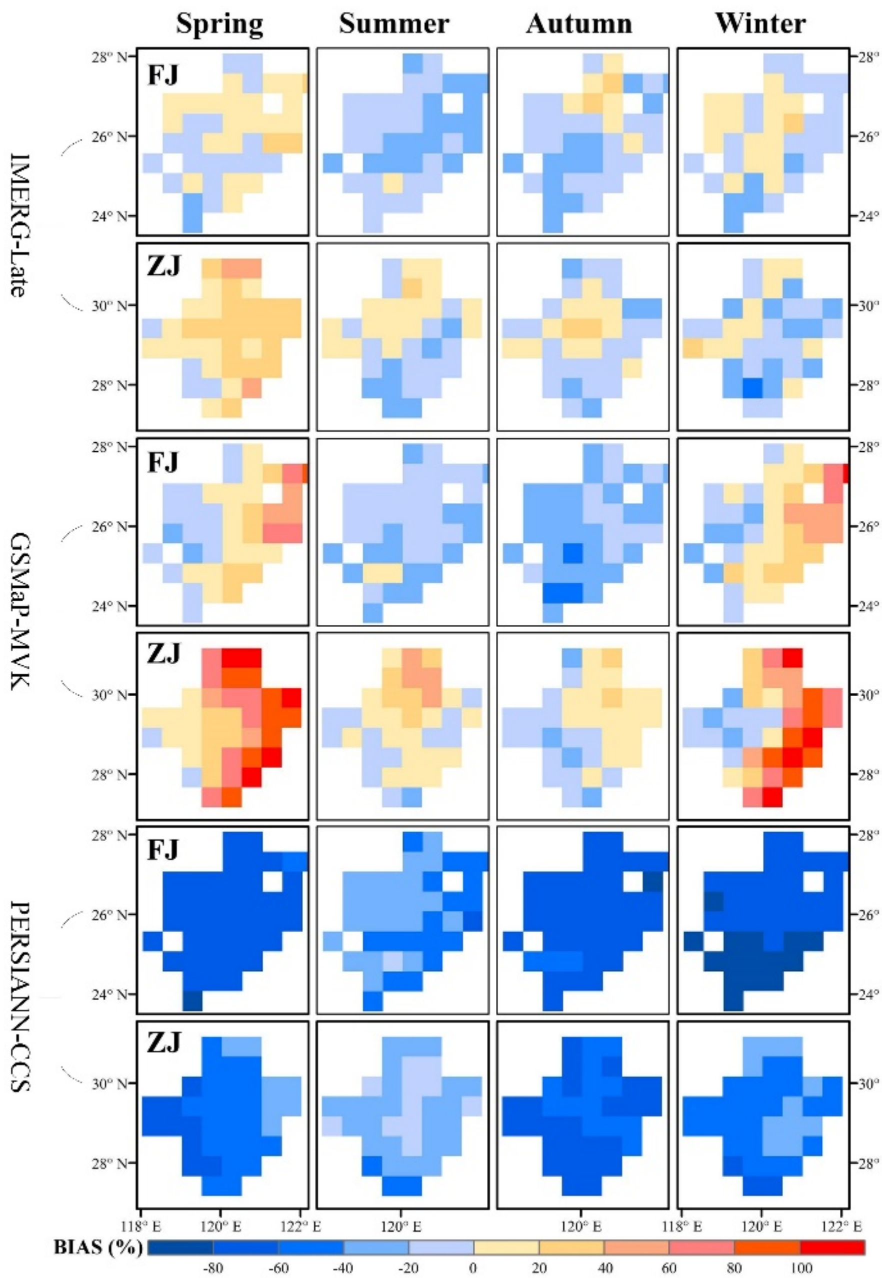

Fig. 1 The spatial maps of the total biases of three SPPs for four seasons over the Fujian (FJ) and Zhejiang (ZJ) provinces, respectively.

6. **It would be useful to see how many gauges are available for each stdev class shown in Fig. 10.**

Response: Thanks. We provided the gauges available for each category as shown in the following Figure 2. One can see that the gauge number in each category is over 135, which suggests that the sample size in each class is sufficient to support the reliability and robustness of the results. We have added the related descriptions in section 3.3, **see lines 319-321**.

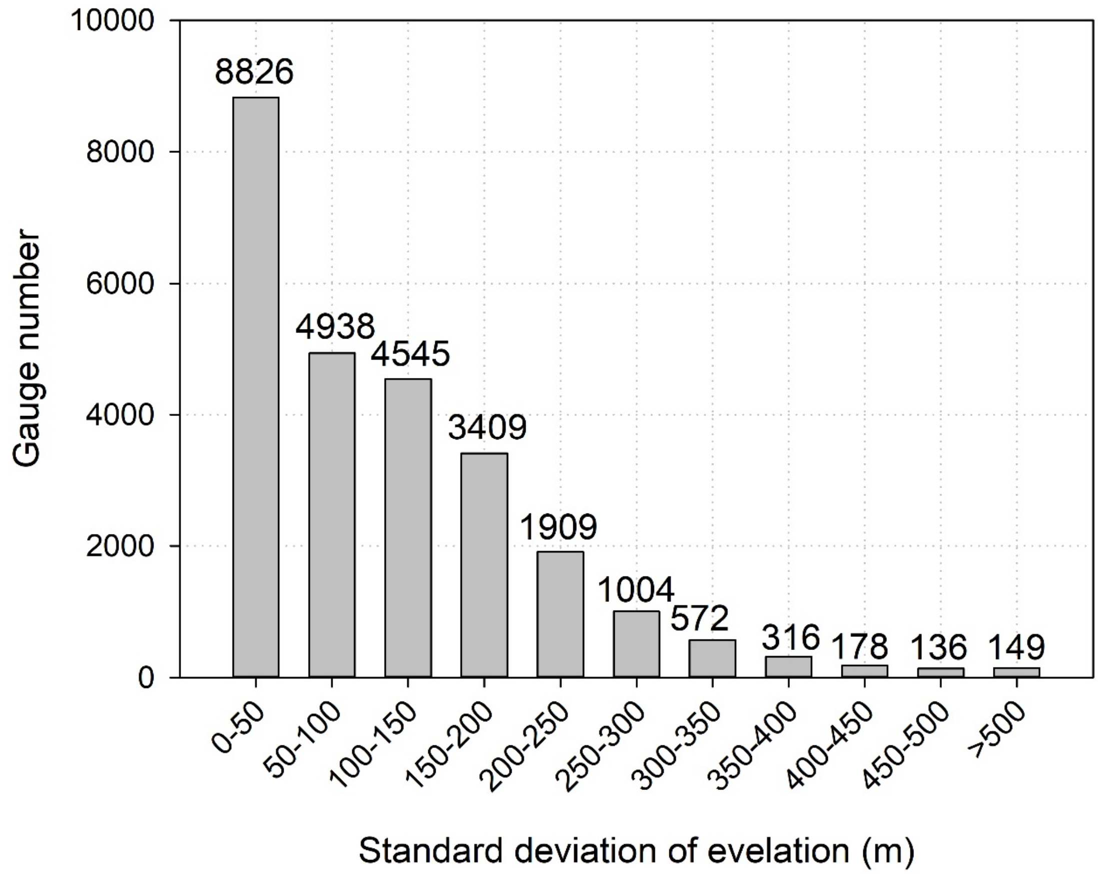

Fig. 2 The gauge number of each topography class.

In addition, we did the experiment again because we found that the previous results in mainland China were from mismatched satellite-ground observations (CGDPA) pairs, and all figures and tables have been revised. Other revisions were highlighted in red font.

Thanks for your review, we look forward to your positive response.