Response for Reviewer #2

High-accurate precipitation data are essential for various aspects and estimating precipitation is challenging especially in complex-terrain and ungauged regions. This work produced a global precipitation dataset (0.25°, 1 hourly) by weighted average of multiple precipitation datasets. However, great improvements are needed in terms of the merging method and validation of the dataset. The main concerns are as follows:

Response: We would like to thank you for your constructive comments on our manuscript. Your insightful review has improved our manuscript considerably. Below is a point-by-point response to your comments.

1) Validation of the produced dataset is inadequate. The authors produced a global precipitation dataset. However, the validation with gauge observations was only conducted in Chinese Mainland, which does not support its accuracy in other regions of the world. I suggest the authors perform a more convincing validation using worldwide observations.

Response: Thank you for your constructive comments and insightful suggestions. We agree with the reviewer that the validation of the MGP suite should be performed on a global scale. However, an independent high-quality global ground precipitation dataset is lacking, which is a problem faced by global scholars, not limited to us. Taking an example, the MSWEP product was only validated in the United States (please see Beck et al., 2019), due to lacking high-quality independent global ground precipitation observations for the public. We really want to validate the MGP products on a global scale, which is crucial for the product application and further improvement of the quality. In the revised manuscript, we will add validation in the United States, because we are getting an independent high-quality ground precipitation dataset here. Meanwhile, we are looking for high-quality independent ground precipitation datasets from around the world. If we can get an independent high-quality ground precipitation dataset in other areas, we will add more validation in the revised manuscript.

2) Why CC was used as the weight for merging various datasets rather than other metrics like root mean square error (RMSE), Kling–Gupta efficiency (KGE), Nash-Sutcliffe efficiency (NSE), et al. In the case where a dataset is highly consistent with the observations in terms of the temporal trend but has a systematic error, this method may lead to a systematic bias in the merging output.

Response: Thank you for your comments. The correlation coefficient (CC) can accurately quantify the performance of evaluated global products, and its range of values is from 0 to 1. Thus, CC is very suitable for determining the weights of global precipitation products. Additionally, Beck et al. (2019) also verified the effectiveness of CC in determining the fusion weights. We agree with the reviewer that the method using CC to determine fusion weights may lead to a systematic bias in the merging output. We will discuss this a little bit in the next version.

We are very grateful to the reviewer for the insightful suggestions. RMSE is related to precipitation intensity. Large differences in space and time exist between satellite-based and model-based precipitation intensity. The method using RMSE to determine fusion weights can consider the dependency of precipitation errors on precipitation intensity but also might introduce an intensity-based bias in the merging output. Considering the respective strengths of CC and RMSE, we accepted the suggestion and add using RMSE to determine fusion weights in the revised manuscript for comparing the performance of these two methods in determining fusion weights. KGE and NSE might be potential weighting methods, but their ranges are negative infinity to 1. The range is too large, and it is also difficult to get a reasonable fusion weight when KGE or NSE is a negative number. These limitations are unfavorable for accurately identifying the error characteristics in each dataset. We will provide a detailed explanation in the revised manuscript of why we used CC and RMSE for weighting and discuss possible limitations with CC and RMSE in weighting based on verification results.

3) The authors repeated the merging procedures three times at different spatial and temporal scales and then spatially and temporally downscale the coarse dataset. What is the underlying logic for applying such a strategy?

Response: Thank you for the comments. The results of Tan et al. (2017) indicated that compared with spatial downscaling, the loss in performance of precipitation products is large in temporal downscaling. To ensure the high accuracy of the produced product, consequently, we first conducted spatial downscaling and then conducted temporal downscaling, for generating high-quality MGP products. We have added the explanations; the modification is represented as follows: Tan et al. (2017) indicated that the reduction of accuracy for precipitation products in temporal downscaling is greater than that in spatial downscaling. To obtain high-accuracy global merged precipitation estimates, therefore, we first conducted spatial downscaling to minimize the loss of the quality of precipitation caused by spatiotemporal downscaling.

4) The authors emphasized that considering the seasonality of errors in the merging procedures is novel. However, no evidence was provided in the manuscript to support the novelty or added value of considering the seasonality of errors in precipitation datasets.

Response: Thank you for pointing this out. Many evaluation studies found that the errors in satellite and reanalysis precipitation products have a significant dependency on seasonality (e.g., Tian et al., 2009; Jiang et al., 2021; Chen et al., 2021). In addition, Global land maps of the weights for satellite and reanalysis data have obvious differences (please see Figure.1). For instance, ERA5 in DJF and MAM seasons has large weights in high latitude and cold mountainous areas, whereas the weights of satellite (i.e., IMERG-Late) in DJF season are lower (<30%) in high latitude areas. Therefore, it is necessary to consider the dependency of precipitation errors on seasonality in the merging algorithms, which can accurately identify the error of seasonality in each dataset. The added explanations (modifications) were presented as follows: (1) Many evaluation studies found that the errors in satellite and reanalysis precipitation products have a significant dependency on seasonality (e.g., Tian et al., 2009; Jiang et al., 2021; Chen et al., 2021). Therefore, the precipitation amounts of satellite and reanalysis precipitation products were separated into four seasons (i.e., MAM (March-May), JJA (June-August), SON (September-November), and DJF (December-February)); (2) In addition, Global land maps of the weights for satellite and reanalysis data have obvious differences. For instance, ERA5 in DJF and MAM seasons has large weights in high latitude and cold mountainous areas, whereas the weights of satellite (i.e., IMERG-Late) in DJF season are lower (<30%) in high latitude areas.

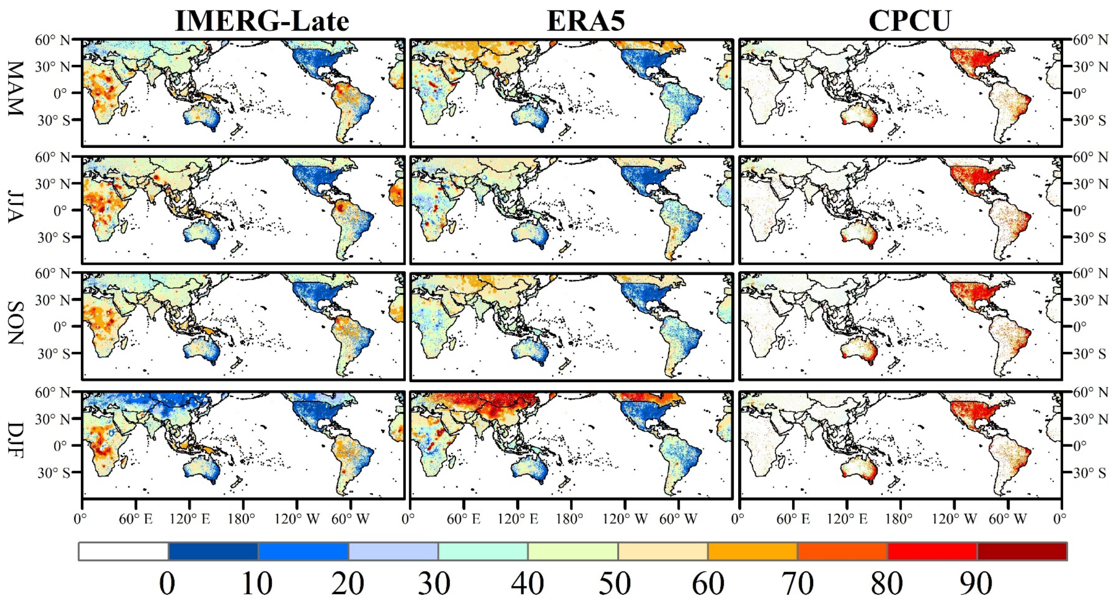

Figure 1. Global land maps of the weights designed to satellite, reanalysis, and CPCU in four seasons (i.e., MAM, JJA, SON, and DJF).

References

Tian, Y., Peters-Lidard, C.D., Eylander, J.B., Joyce, R.J., Huffman, G.J., Adler, R.F., Hsu, K.L., Turk, F.J., Garcia, M., and Zeng, J.: Component analysis of errors in satellitebased precipitation estimates. J. Geophys. Res.-Atmos. 114 (D24), https://doi.org/10.1029/2009JD011949, 2009.

Jiang, Q., Li, W., Fan, Z., He, X., Sun, W., Chen, S., Wen, J., Gao, J., and Wang, J.: Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. J. Hydro. 595, 125660, https://doi.org/10.1016/j.jhydrol.2020.125660, 2021.

Chen, H., Yong, B., Kirstetter, P. E., Wang, L., and Hong, Y.: Global component analysis of errors in three satellite-only global precipitation estimates. Hydrol. Earth Syst. Sci. 25(6), 3087-3104, https://doi.org/10.5194/hess-25-3087-2021, 2021.

5) The comparison between MGP-6P and MGP-3P makes no sense. If another group of datasets was selected for the merging, the results may differ evidently from those in the manuscript. The key to this problem is not the number of datasets that are applied, but whether the merging method can accurately identify the error characteristics in each dataset.

Response: Many thanks for your comments. The comparison between MGP-3P and MGP-6P was used to answer the question of which is crucial for the quality of merged precipitation estimates: the number of input products or the quality of input products? Therefore, we designed six input sources for merging scheme 1 to generate MGP-6P, without considering the quality of input products, whereas we considered the quality of input products in merging scheme 2 for generating a research level precipitation product MGP-3P, merging the satellite-only IMERG-Late and ERA5 reanalysis precipitation products. We agree with the reviewer that it is important that the merging method can accurately identify the error characteristics in each dataset. However, most algorithms often combine the advantage of the strengths of different data with their shortcomings into the merging output. At this point, the quality of the input products is also very important to that of merging output. That is why we performed the comparison between MGP-3P and MGP-6P. In the revised version, we will delete merging scheme 1 and keep merging scheme 2, and hence the comparison between MGP-6P and MGP-3P will be removed.

6) Given that a large number of global datasets have been available with a spatial resolution of 0.1° or finer, the produced dataset with a spatial resolution of 0.25°may not be competitive in the science community.

Response: Thank you for your comments. In version 1 of MGP, the accuracy of MGP is the primary consideration. Considering the finest spatial resolution of ERA5 is 0.25°, the purpose of the spatial resolution at 0.25° is to ensure the accuracy of the MGP-3P product. If the spatial resolution of MGP is 0.1°, only IMERG-Late can be used as input in the spatial downscaling when the spatial resolution is scaled down from 0.25°to 0.1°, which could not guarantee the accuracy of MGP-3P. We agree with the reviewer that on the premise of ensuring the accuracy of MGP, the finer the resolution of MGP, the more competitive it becomes. In the revised manuscript, we will try to use ERA5-Land (0.1°) instead of ERA5 (0.25°) as one of the input products for the MPDF algorithm for generating MGP with a finer spatial resolution (0.1°, hourly).

Some specific comments:

7) Line 74-75, "the quality of gauge observations is extremely dependent on the spatial density of the rain gauges": the quality of gridded data interpolated from gauge observations, rather than the gauge observation itself, depends on gauge density.

Response: Thanks. We have revised this sentence as follows: "Nevertheless, the quality of gridded data interpolated from gauge observations is extremely dependent on the spatial density of the rain gauges.".

8) Line 143-145, "spatial interpolation was proved to be an effective method in improving the quality of global satellite and reanalysis precipitation estimates": There is no definitive relationship between spatial interpolation and improving the quality of precipitation datasets.

Response: Thank you for pointing this out. We have removed this incorrect description.

9) There are many confusing sentences in the manuscript, e.g. Line 177-182, Line 242-243, Line 273-275, Line 467-469, Line 526-527.

Response: Thank you for pointing those confusing sentences out. Line 177-182 was revised as follows: "The quality of the MSWEP product mainly depends on that of the ground-based precipitation observations as the weights designed for ground-based precipitation observations are significantly larger than those designed for satellite- and reanalysis-based precipitation estimates."; Line 242-243 was revised as follows: "The spatial distribution of rain gauges used in the CPCU system can be seen in Fig. 2a."; Line 273-275 was revised as follows: "The precipitation estimates at a grid cell come from CPCU precipitation observations when rain gauges captured precipitation occurrences at the same grid cell but satellite and reanalysis did not."; Line 467-469 was revised as follows: "the MSWEP estimates were derived from high-quality ground-based precipitation observations by using the spatiotemporal downscaling technique (Beck et al., 2017, 2019)."; Line 526-527 was revised as follows: Notably, false precipitation was not observed from evaluated global precipitation products as the rainfall intensities were from the ground references.

We will carefully check the manuscript to eliminate all issues.

10) The used data are not fully introduced, especially for the CPCU dataset.

Response: Thank you for pointing this issue out. We have added the introduction of the used data. The modifications were presented as follows: Three different global precipitation products, i.e., satellite-only IMERG-Late (0.1°, 0.5 hourly; Huffman et al., 2019), ERA5-Land reanalysis (0.1°, hourly; C3S, 2017; Hersbach et al., 2020), and ground-based Climate Precipitation Center unified (CPCU; 0.5°, daily; Xie et al., 2007; Chen et al., 2008), were used as inputs for the new algorithm. IMERG-Late is a multi-satellite merged precipitation product that blends microwave and infrared data. ERA5-Land is a model-based precipitation dataset, with finer spatiotemporal resolutions (0.1°, hourly) compared with other model-based precipitation products. In terms of CPCU, it was generated using optimal interpolation based on > 17000 gauges, and its precipitation estimates cover quasi-global (50°S-50°N) land areas (Xie et al., 2007; Chen et al., 2008). The reasons for the selection of these precipitation products are that IMERG-Late performed better than other satellite-only precipitation products in most areas of the world (Chen et al., 2020b, 2021), whereas ERA5-Land has the finest spatiotemporal resolution compared with other model-based reanalysis precipitation products.

11) Line 412-414: it is not easy to see the differences from Fig. 3. It is suggested to present the differences between these global datasets and the produced dataset.

Response: Thank you for your insightful suggestion. We accepted your suggestion, and the maps of differences between these global datasets and the produced dataset will be provided in the revised manuscript. Notably, the analysis based on new results will be provided.

12) Section 4.2. Comparing the quality of precipitation datasets on different temporal scales makes no sense, because it is well acknowledged that precipitation product performs worse on a shorter temporal scale than on a longer scale.

Response: Thank you for the comments. We agree with the reviewer. We have deleted Section 4.2.