

Discussion comment 3

I concur with previous reviewers that the paper is well written and highly of interest for the hydrological community. I believe that the idea of merging satellite-based, gaugebased and reanalysis precipitation data for taking advantage of the benefits of each product is good, and surprisingly not largely investigated so far. I have only two short comments that, in my opinion, should be addressed.

We sincerely thank Dr. Brocca for his remarks, to which we respond below.

1) The gauge-based evaluation of precipitation datasets is carried out by using the GHCN –D database. I am fully aware that it is difficult to determine the quality of precipitation datasets, but I have not seen in the paper any comments about the reliability and accuracy of the gauge-based dataset.

It is hard to make any generalizations about the quality of gauge observations since the “true” precipitation is not known. Nevertheless, GHCN-D data have been subjected to stringent quality control procedures (see <https://www.ncdc.noaa.gov/oa/climate/ghcn-daily/index.php?name=quality>), so we expect the data to be as reliable as possible.

For instance, the spatial representativeness of point measurements might be low even when the average of multiple stations is done. This is particularly problematic at 0.5° resolution.

There is indeed a scale discrepancy between the point scale of gauge observations and the grid-cell scale of the datasets, but the quantification of the implications of this discrepancy for our results is confounded by the lack of knowledge on the “true” spatio-temporal precipitation pattern. While approaches have been proposed to correct for this scale discrepancy (see, e.g., Pegram and Bardossy, 2013) they inherently involve many assumptions and uncertainties. Moreover, if we had applied such corrections, it is unlikely the weight maps would have changed much, if at all, as they are based on correlation coefficients (i.e., they are based solely on the temporal dynamics of the data).

Pegram, G, and A. Bardossy (2013), Downscaling Regional Circulation Model rainfall to gauge sites using recorrelation and circulation pattern dependent quantile–quantile transforms for quantifying climate change, *Journal of Hydrology* 504, 142–159, doi:10.1016/j.jhydrol.2013.09.014.

For instance, if satellite datasets were compared with gridded-based dataset (just for checking, not for the merging procedure), what are the results? I believe that a different picture can be obtained (but I could be totally wrong). Can the authors add some additional discussions on this aspect?

We would not expect a markedly different picture if we had used a gridded gauge-based dataset (e.g., CPC Unified) instead of the GHCN-D station data. This is because in grid cells with gauges, the gridded datasets would generally follow nearly exactly the temporal dynamics of the station data.

2) As hydrologist, I was very interested from the analysis for evaluating MSWEP dataset through hydrological modelling. Particularly, Figure 9a highlights that the overall performance of satellite-only MSWEP dataset is significantly lower than the reanalysis-only and gauge-only MSWEP dataset. Do the authors have an explanation for these large differences?

Thank you. We have added the following statement to the revised m/s: “The overall low performance of the satellite-only MSWEP reflects the lack of tropical catchments in our catchment set.”

In Figure 8b, satellite-only performance are better than the reanalysis in the tropical region, but (nearly) always lower than the gauge-based product. As the paper represents also one of the first studies performing a comprehensive assessment, on a global scale, of the three sources of precipitation (satellite, reanalysis, raingauge), it would be interesting to extend the discussion of the obtained results through hydrological modelling. I believe it would be highly of interest to the HESS readership.

We thank the reviewer for his encouraging comment and agree this is the first comprehensive assessment of its kind. We mention in the m/s that the good performance of the gauge-only MSWEP is somewhat misleading and happens because the large majority of the catchments are located in regions with dense P measurement networks (page 17 lines 5–7). However, to further address this comment we added the following text to the revised m/s: “These results once again confirm the complementary nature of satellite and reanalysis data clearly demonstrated in the previous evaluation of P datasets using gauge observations (see Section 4.2).”