

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/hess-2021-332-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.



Comment on hess-2021-332

Anonymous Referee #2

Referee comment on "Easy-to-use spatial Random Forest-based downscaling-calibration method for producing precipitation data with high resolution and high accuracy" by Chuanfa Chen et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-332-RC2>, 2021

Easy-to-use spatial Random Forest-based downscaling-calibration method for producing high resolution and accurate precipitation data

This study used a random forest machine learning algorithm to downscale the GPM satellite precipitation measurements and calibrated them with gauge observations and the aforementioned high-resolution variables. The study is interesting and the authors presented some interesting results, while the context is hard to follow and the writing needs significant improvement. The tense of the article is very confused. The present tense and past tense are confused throughout the article, and there are also many grammatical mistakes. There are too many mistakes to point out here. It is suggested that the paper should be revised by someone who is good at English.

Major comments:

- You have the daily measured precipitation data, and you can also get daily GPM precipitation data. Why don't you analyze the downscaling model on the daily scale? In fact, there have been many daily downscaling models based on Random Forest. In general, the novelty of the article is not enough.
- The author did not give a clear explanation why the environment variables could be used in RF model to derive the correlations. The relationships between these variables and precipitation should be explained in detail, not just by listing the reference papers that use these variables. The correlations between each environmental variables (NDVI, LST, DEM etc.) with precipitation and their contribution to the prediction of the precipitation should be fully discussed.
- The IMERG data were fused into precipitation products from the satellite observations and gauge data. As the gauge data had been applied in the IMERG data generation (Level3), how could you reconcile the errors in SRF with the gauge data at ground used in your study?
- Evaluation of the model was performed by considering rain gauge data as observations at ground. However, it assumed that the rain gauge measurements were representative values at their respective grid-cells. Although this was widely used in other literatures, the authors should discuss this issue to support their decision.

- This paper analyzed the accuracy of various precipitation data sets on a monthly, seasonal, and annual. There is no need to spend a lot of space in the paper to analyze the differences between seasonal and annual results, which makes the paper look more like a technical report without scientific insights. You should compare with previous literature that states if the results are in agreement or not with other studies in the field in your Discussion part. However, the present part of the discussion is rather empty, and it mainly analyzes the research results of others, and lacks the systematic analysis and discussion of the proposed method in this paper

Minor comments:

- Line 158-160: "Overall, the high spatial and temporal variability of precipitation with the complex topography makes the study site ideally suitable for the evaluation of satellite-based precipitation estimates." Can you give the reasons or some references as evidences?
- Section 5.1. Please indicate the optimized hyper parameters (i.e., number of trees, depth of the tree, and number of features) for the Random Forest model.
- Line 269-271: "the spherical model was used since it shows better results than the others in the experiments." You should give the analysis results (data, charts, etc.) to prove the reason for choosing the spherical model.
- Line 315-316¼□"Our monthly-based estimation method was compared with the annual-based SRF fraction disaggregation method (termed as SRFdis)" Please explain SRFdis in detail.
- As can be seen from Figure 5 and Table 2, the proposed method and Bi SRF have very similar performance in most accuracy index, such as CC. However, the bilinear interpolation downscaling method is obviously easier to operate than SRF downscaling method. Is it necessary to use a more complex downscaling method to improve the CC value of 0.003?
- Line 404¼□Change "our method" to "the proposed method"
- Line 432¼□Why do you say that "This is because this year has the largest precipitation (Fig. 2)." Can you explain why this is the reason for the worst results in 2018?
- Line 488: "Table 1"? Table 1 is the detailed information of the datasets used in the study.
- Line 499-500¼□What are the reasons that the day-night land surface temperature difference was used in this study?