Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC4
https://doi.org/10.5194/hess-2021-332-RC4, 2021

Comment on hess-2021-332

Anonymous Referee #4

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-RC4, 2021

The paper applies a machine learning technique for downscaling and calibration of precipitation based on remotely sensed inputs that also aims to incorporate the spatial structure of rainfall using spatial autocorrelation. The idea of paper is interesting and it also has a organized structure which is generally well-written. However, based on the methods applied and discussion of the results, the paper has several shortcomings that need to be addressed and further explained prior to publication.

Major comments:

1. Several aspects of the OK based interpolated maps at 1k and 10k resolutions are not fully convincing. First, the accuracy of the OK-derived maps should be reported in order to determine reliability of the maps. Errors in the interpolated maps are going to be propagated to the errors in the spatial RF model because it is one of the covariates used, so they are important. It would be interesting to see if the large RMSE's in the middle part of the study area in fig.7 also show up with large errors or variance in the OK maps.

Related to this, the authors also need to further clarify the interpolation of a 1km image based on a 10km IMERG images using OK, which is a raster-to-raster interpolation performed (lines 273-284). A coarse to fine raster-based interpolation seems unusual, so that authors need to further describe this step.

2. The parameters tested and chosen for all the models, including the semi-variogram should be reported otherwise the study is not reproducible.

3. It appears that the sRF model (also for the other ML techniques applied) did not include a separate testing phase . This is a standard approach applied when assessing the accuracy of a ML methods. I would suggest to also validate the models using an independent test that is not used in the training phase. Or re-configure the ML methods to split the total data into a training and a test set.

4. Discussion of the results focuses more on the positive aspects of using sRF but the authors do not give a balanced view by providing a critical analysis of the results of sRF.

For instance, the accuracy metrics presented highlight that sRF performs well compared to the other models. However, visual comparison of the boxplots of these metrics alone in figs. 8-9 shows comparable accuracies all the models based on their range and median. Significant differences between the accuracies obtained, particularly in relation to sRF, should be reported to provide gravitas on the authors claim that sRF outperforms the other models.

5. There is an underestimation of precipitation values regardless of the model used based on fig 5 . This should be further elaborated in addition to the three accuracy metrics provided, so the bias of the estimates should also be reported. Furthermore, for very high precipitation values (e.g. >400mm), the scatter of the points in fig.5 becomes larger, indicating that all the models tested perform poorly at v. high rainfall amounts. It could be insightful to assess separately how the models compare for v. high rainfall conditions, since prediction of these extreme cases need to be generally improved.

6. It is unclear how the importance measures are calculated from fig. 13, so this should also be included in the methodology of the paper. Furthermore, discussion of the rankings could be made more in depth by determining whether they agree or deviate (and why they do) from known controls on rainfall distribution.

7. The authors already indicate that there is a delayed response of vegetation to rainfall. It is perhaps expected that the NDVI is one of the least important factors in the sRF model. But actually, this also provides an opportunity to also explore the lagged values of the predictors (and not only NDVI) with known delayed responses to rainfall.

Minor comments:

1. The captions of the figures need to be improved. Some of the features in multi-plot figures are hard to understand because of the captions are highly simplified.

2. The final version of the manuscript will benefit for another round a English check as some sentences a phrased a bit vaguely (e.g. line 150-151)