Hydrology and
Earth System
Sciences
Discussions

Interactive comment on "A Nonparametric Statistical Technique for Combining Global Precipitation Datasets: Development and Hydrological Evaluation over the Iberian Peninsula" *by* Md Abul Ehsan Bhuiyan et al.

Anonymous Referee #2

Received and published: 4 July 2017

This study is focused on developing and evaluating a non-parametric statistical method to generate an ensemble of precipitation estimates to better capture the uncertainty in global precipitation estimates. The inputs to the algorithm include three global satellite-based precipitation estimates, a reanalysis product (rainfall and air temperature), elevation, and remotely-senses soil moisture.

While the manuscript is written very well, and the topic will be of interest to the HESS readership, the following caveats needs to be addressed before publication of the manuscript:

-Major Comments:

1) The authors provide detailed information about the products used in the QRF method and explain the method itself very well. But no information is provided on how the training and validation of the method is performed. How much of the data is used for training? How much used for validation and testing? Please also include the temporal coverage of the data.

2) There is also no information on avoiding overfitting. One of the challenges in data-driven methods is overfitting (i.e. the method is so fine tuned to the training data, and has larger errors when applied to new datasets). I don't see any discussion of this in the paper. For example how did you choose to use 1000 trees in the model? Are there noticeable differences between the performance of the method during training and validation?

3) How are the ensembles generated? No information is provided on how each ensemble member is initialized and generated using the QRF trained on the data.

4) The results provided in section 4 needs to be clarified whether they are based on the data used in training or the data used in validation, or a mixture of both.

5) The low value of NCRMSE for the small basins report in Page 11, Line 6 is a signal of overfitting in the algorithm. This is another indication that overfitting should be analyzed in depth.

6) Page 6, Lines 10-18: Please clarify if different trees are developed for the three groups that you introduce at the beginning of the paragraph. You have introduced four groups at the end (warm-high, warm-low, cold-high and cold-low) but there is no reference to the categorization of products based on their rain detection (group 1-3 in lines 11-12).

-Minor Comments:

1) Why did you choose to use PERSIANN product instead of the newer version

PERSIAN-CCS?

2) In section 2.3, please include details on how you have downscaled the 0.5 degree reanalysis product to 0.25 degree to be consistent with other products.

3) In section 2.4, please include the version number of the ESA-CCI product.

4) Page 8, Line 2: What does actual uncertainty mean? Do you mean uncertainty in the reference product? If so, please explain how a UR=1 will provide the best estimate of the uncertainty in the reference product.