

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.

V. Maggioni (Referee)

vmaggion@gmu.edu

Received and published: 27 June 2017

This study proposes to use a non-parametric statistical model (the Quantile Regression Forest) to merge several precipitation datasets together with ancillary information (e.g., soil moisture, air temperature, and terrain elevation), which are used as predictors to estimate a superior precipitation product. I find this work innovative and worth publication. I have a few comments, which I would like the authors to consider and which I hope will meliorate the manuscript.

C1

Minor comments:

Sometime in the text, there is confusion in the way the words “ensemble” and “members” are used. An ensemble is made of several members. Therefore, “ensembles” would refer to multiple ensembles made of several members.

Page 1, Line 12: replace “generated ensembles to force” with “to generate ensembles that force”

Page 2, Line 18: rephrase as “Satellite rainfall error models, such as SREM2D (Hosain et al., 2006), have been used to . . .”

Page 2, Line 22: replace “the error characteristics” with “errors and uncertainties”

Page 2, Line 24: replace “allow for efficient combining of” with “efficiently combine”

Figure 2: a “d” is missing in the word “and” in the top center block. This methodological framework scheme could be improved: as it is, it is unclear what each block does.

Figure 6: can the authors discuss why QRF performs so poorly for lower rain percentiles (<25%) in term of bias?

Major comments:

A better explanation on why the authors picked those predictors is needed. For instance, why including soil moisture, but not a vegetation indicator? Why including three satellite precipitation products instead of two? Was the dataset combination that produced the best results picked? Are all the predictors really needed?

What is the impact of merging datasets that are not totally independent? For instance CMORPH and TMPA 3B42 use the same MW overpasses in their algorithms, as PER-SIANN, CMORPH, and TMPA use the same IR observations. I am wondering whether there is any chance that too much weight is given to this information in the merging model.

C2

