Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-43-AC3, 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "A flexible two-stage approach for blending multiple satellite precipitation estimates and rain gauge observations: an experiment in the northeastern Tibetan Plateau" *by* Yingzhao Ma et al.

Yingzhao Ma et al.

yzma@colostate.edu

Received and published: 15 April 2020

Anonymous Referee #3:

The manuscript by Ma et al. Presents a very interesting study on blending multiple satellite estimates to obtain a better precipitation estimates, especially over region with complex terrain. The analysis is systematic and results support the improvement in precipitation estimates due to two-stage blending approach. During my read, on several occasion I kept searching for necessary details. Unless those details are provided,

it is hard to fully evaluate the merit of this work. Therefore I would suggest major revision of the current version of the manuscript.

Response: We thank this reviewer for the critical comments. More details of the two-stage blending (TSB) approach is attached in the supplement. We will rephrase them in the revised manuscript as pointed out by this reviewer.

Authors may want to improve the manuscript along the following lines:

[1] Please provide full details of bias adjustment and data merging stages. With help of some example dataset, Authors need to describe how Equations [1], [2a] and [2b] adjust the bias. Similarly please demonstrate with some dataset how weight parameters were obtained from Equation [3].

Response: The details of bias adjustment and data merging stages are provided in the supplemented file. As for the calculation details of Eqs. (1) to (3), please see the supplementary materials. We will also rephrase the Method section in the revised manuscript as pointed out by this reviewer.

[2] Please include plots justifying why Student's t distribution was selected. I am sure at different training sites, different distributions (Lognormal, Gamma, etc.) may show better performance.

Response: We fully agree with this reviewer that at different training sites, different distributions (Lognormal, Gamma, etc.) may show better performance. Given various SPE at different training sites, the specific probabilistic function is not limited to a certain distribution. For demonstration purposes, we herein apply the Student's t distribution, with its mean parameter expressed as a linear regression of the original SPE and terrain feature in this case. We will double check the assumption of Student's t distribution as pointed out by this reviewer.

[3] Please explain how the information from gridded data (Satellite estimates) was transferred to point locations (training and validation) sites. Did Authors apply some

downscaling approach? Bringing information from 25km grid to a point in a complex topographical region is challenging.

Response: Perhaps we didn't describe it very clearly in the original manuscript. We admit that there is a scale gap between gridded SPE and point-based gauge observations. The downscaling approach is not applied in this study. To ensure the same resolution among the original SPE, the IMERG data are resampled from 0.10° to 0.25° using the nearest neighbor interpolation to eliminate the scale difference. The rain gauge network is spatially interpolated with a 0.25° x 0.25° resolution in the region of interest on each rainy day using a bilinear interpolation approach. The 34 grid cells with the gauge sites are assumed as ground references (GR) in the blending process. We fully agree that it is challenging to bring information from 25 km grid to a point in a complex terrain region. We will rephrase the statements and add some discussions in the revised manuscript as pointed out by this reviewer.

[4] In equation [1], normalized elevation is used as a covariate. If it is not included, how it will affect the result. Can you quantify it? Was that included just because you are dealing with TP? In the discussions (Section 5), Authors mention about the importance of including other covariates related to precipitation generation mechanism.

Response: We are thankful to this reviewer for the important comment. We quantify the impact of elevation covariate on the bias-corrected and blended SPE performances as pointed out by this reviewer in Table 1. It is found that the consideration of elevation performs slightly better skill compared with the model without terrain information in this case study. We would like to admit that it is an initial exploration partly because we are dealing with the TP. We will rephrase these concerns in the revised manuscript as pointed out by this reviewer.

[5] As mentioned in Section 2, the data of only warm period from May to September 2014 has been used in this study. Since all the satellite data are available for several years, can Authors perform similar analysis for few years and validate their approach?

Response: We thank the reviewer for this suggestion. Please allow us to explain that this study aims to develop a newly TSB algorithm on the multi-satellite precipitation data fusion in a certain time in regions of interest. Given that the larger challenge in the TP is to provide more accurate rainfall in a spatial domain, we are trying to overcome the shortage of limited rain gauge network based on the available SPE with spatial advantage using the TSB method in the NETP as a demonstration purpose. We agree that the satellite data are available for several years, but the exploration of long-term periods for the TSB method is another critical issue, e.g., the consideration of time impact on the fusion result. We will replenish this discussion in the revised manuscript. Additionally, the model performance of this new approach is demonstrated based on various aspects, and the evaluation analysis for long-term period and extended regions (e.g., TP) will be addressed in the future work.

[6] Since similar approaches have been developed previously (as mentioned at the end of second paragraph of page 2, Authors should compare the results with the existing approach. The only unique feature of the current approach is that it provides predictive uncertainty.

Response: We thank the reviewer for this great suggestion. It is very important to formally quantify the predictive uncertainty in the Bayesian analysis, which is one of the unique features for the TSB method. As suggested by this reviewer, the TSB approach is compared with two existing fusion approach, i.e., Bayesian model averaging (BMA) and One-outlier removed (OOR) in Table 2. The statistical summary of data comparison among the three fusion approaches at the validated locations are shown below. The TSB approach performs the best skill with the RMSE, NMAE and CC at 4.34 mm/d, 49.2%, and 0.606, respectively, as compared with the other two fusion methods. We will add this comparison in the revised manuscript as recommended by this reviewer.

[7] The results presented in Figures 3 and 4 homogenizes many things. In Figure 3, are you presenting the average value over all the validation sites? I am sure results will differ significantly if you look into individual sites. Also time series plots would show

more features than the bar plot. The results from blended is similar to many adjusted SPE, then can it be concluded that there is no need to blend. Simply apply the stage 1, bias adjustment, and select the best SPE.

Response: We are thankful to the reviewer for these comments. Yes, Figure.3 in the original manuscript presents the statistical error summary over all the validation sites. We agree that there are more features if looking into individual sites than overall evaluation of the validated sites. The time series plots of daily rainfall estimates and rainfall accumulations in terms of GR, original and blended SPE at a validated grid cell with a rain gauge labeled as ID 56173 is shown in Figure 1 below as a demonstration example. This rain gauge locates in 32.8° N, 102.55°E, 3484 m, and has the maximum rainfall records in the warm season of 2014 in the NETP. Visual analysis shows that the blended SPE provides reasonable rainfall as compared to the original SPE, and the blended SPE also has a better skill in terms of RMSE at 4.95 mm/d compared with the original SPE including PERCDR (10.71), 3B42V7 (9.76), CMORPH (8.0), and IMERG (10.49), respectively.

This reviewer also raises a question that why not be careful at the first place in selecting a good set of SPE, or simply apply the first stage of bias correction and then select the best SPE as the final product. To address this issue, we investigate the error differences among the best-performed SPE, i.e., BC-IME, and blended SPE before and after the removing of the worst-performed SPE, i.e., PERCDR, for 10 random verified tests in the warm season of 2014 in the NETP (Figure 2). It implies that it is beneficial to involve the second stage in the TSB method as the blended SPE performs better skills than the simply bias correction step with the best-performed SPE. The primary reason is that the blended step is designed to integrate various types of SPE, which is limited for the simple bias-corrected step. Also, both blended SPE products in Figure 2 show similar performance in terms of the RMSE, NMAE, and CC indices. The TSB approach has an advantage of not getting impacted by the poor quality SPE in the application, partly because the proposed BW model in Stage 2 can reallocate the contribution of

the SPE based on their corresponding bias characteristics.

We will rephrase the related concerns in the revised manuscript.

[8] In Figure 4, Authors conclude that the blended data have been dropped towards the gauge references but please look at the precipitation with higher values. It appears that red dots have narrow spread for the lower values but SPE is over estimating the values.

Response: We are thankful to the reviewer for the important comment. We also notice that there is an overestimation for the original SPE compared to GR, and the blended SPE shows different spread at various rainfall intensities. To address this issue, we perform an additional analysis of probability density function of daily rainfall with various intensities at the validated locations in Figure 3c. There is an overestimation for the original SPE but an underestimation for the blended SPE as the daily rainfall is more than 15 mm, partly because the BC process might over-correct the original SPE on the heavy rainfall in this case. We will rephrase this statement in the revised manuscript as pointed out by this reviewer.

[9] Authors claim that the two-stage approach has advantage of not getting impacted by the poor quality SPE. Based on Figure 4a, it can be argued that why to include those SPE which has very low weight. Please justify. Furthermore, Figure 6 shows improvement ratio, of course the SPE with very low weight will show high value here. Why not be careful at the first place in selecting a set of SPE?

Response: We thank this reviewer for the critical comment. It is well known that the SPE are obtained from different satellite retrieval algorithms, and each of them can provide various rainfall information. The over-performed SPE would provide more information, and the poor-performed ones give less value. It is thus necessary to integrate all kinds of SPE so as to reduce the predictive uncertainty in the domain. The proposed TSB approach has an advantage of integrating various SPE information and not getting impacted by the poor quality of SPE, partly because the Bayesian weight model in Stage 2 is can reallocate the contribution of the SPE based on their corresponding bias characteristics.

To address this issue, we also investigate the statistical error difference among the best-performed SPE, and blended SPE before and after the removing of the worst-performed SPE, i.e., PERCDR in this case, for 10 random verified tests in the warm season of 2014 in the NETP (Figure 2). It is found that the blended SPE performs better skill than the simply bias correction with the best-performed SPE (e.g., BC-IME in this case), and both blended SPE products show similar performance with the RMSE, NMAE, and CC indices. It proves that it is beneficial to involve the second stage in the TSB method.

We will rephrase the related expressions in the revised manuscript.

[10] Authors talk about CC for a rainfall event (Sept 22, 2014)? Given that the analysis is performed on daily data, how do you obtain CC?

Response: There are 27 rain gauge sites in total that has a rainfall record on Sep 22, 2014 in the regions of interest. The CC index is calculated based on the data sets from the 27 grid cells.

[11] Title says ‘A flexible two-stage approach...’ In the second paragraph of Section 6, Authors talk about what is the flexibility here. The statement is very general that it is capable of involving a group of multi-SPE. Is that so unique? Please look into it and accordingly modify the title.

Response: We thank this reviewer for the important comment. The word of “flexible” is removed in the title. We will replace the title with “A two-stage approach for blending multiple satellite precipitation estimates and rain gauge networks: An experiment in the northeastern Tibetan Plateau” in the revised manuscript.

[12] Figure 8a is quite different from Figures 7a to 7d. By blending, higher values disappeared from the map except in Southwest corner. Please explain.

Response: Thank you for this specific comment. Fig.8a is the spatial map of the blended result which is weight summation of the original SPE from Figs. 7a to 7d. There is an overestimation for most of the original SPE in the NETP in this experiment, the bias of the blended SPE is reduced based on the TSB approach. Thus, higher values disappear from the map except in southwest corner. Because daily mean rainfall is the highest in southwest corner for each SPE, higher value exists after the blending process. We will explain this issue in the revised manuscript as pointed out by this reviewer.

[13] The blending product will be extremely beneficial for the areas where there is no or very few rain gauges (specially in mountain area). However the study area was carefully selected in such a way that the rain gauge intensity is high. Can the results be extrapolated from the training and validations sites to get the improved blended gridded product, the way Authors have done in Figure 8? If yes, then there must be some guideline how many minimum training sites do I need to apply this two-stage approach in other complex regions.

Response: We are thankful to this reviewer for the comment. As pointed out by this reviewer, it is helpful to give some guideline that how many minimum training sites are needed to apply the TSB approach in a region with complex terrain and limited ground observations. The sensitivity analysis of the number of training grid cells on the performance of blended SPE at the validated sites is explored in Figure 5. As the number of training sites is increasing, there is a decreasing trend for the RMSE and NMAE values, but a slight increasing trend for the CC value. Except for an anomaly with No. 23, the performance of the blended SPE becomes similar as the number of training sites increases to 21 in this case. Also, it is noted that if more useful information is provided from the involved SPE and rain gauges, it is more beneficial for the blended gridded product in the region of interest. We will rephrase this critical issue in the discussion part in the revised manuscript.

[14] The manuscript should be thoroughly checked for grammar and usages.

Response: To be thoroughly checked as suggested.

Please also note the supplement to this comment:
https://www.hydrol-earth-syst-sci-discuss.net/hess-2020-43/hess-2020-43-AC3-supplement.pdf

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-43, 2020.

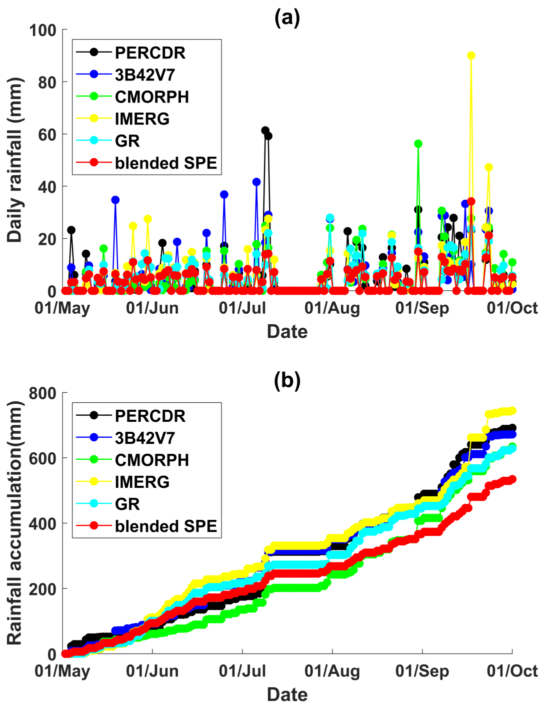

Fig. 1. Time series of (a) daily rainfall estimates and (b) rainfall accumulations at a selected validation grid location with the maximum rainfall records in the warm season of 2014.

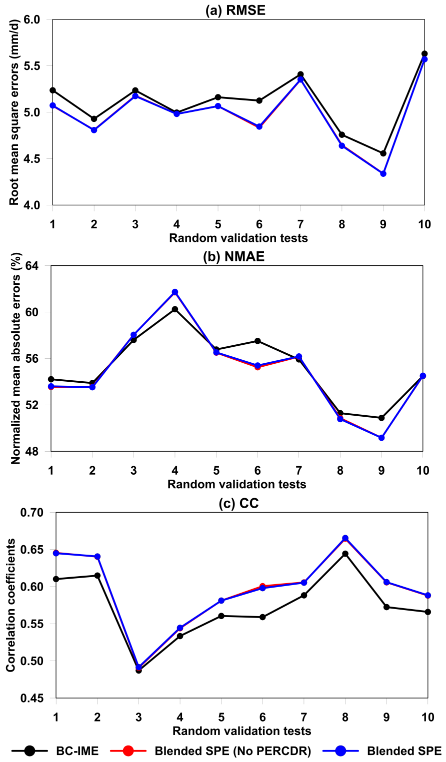

Fig. 2. Statistical error indices of the best-performed bias-corrected SPE (i.e., BC-IME, black) and blended SPE before (red) and after (blue) the removing of the worst-performed PERCDR for 10 random tests

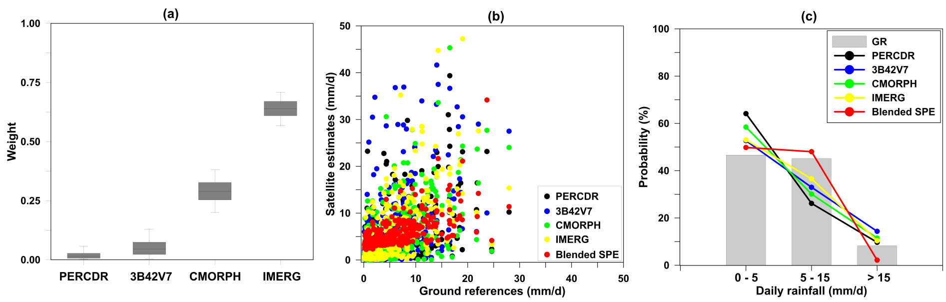

Fig. 3. (a) The Box-Whisker plots of relative weights of SPE in Stage 2; (b) Scatter plots between GR and various SPE; (c) The PDF of the GR, original and blended SPE with various rainfall intensities

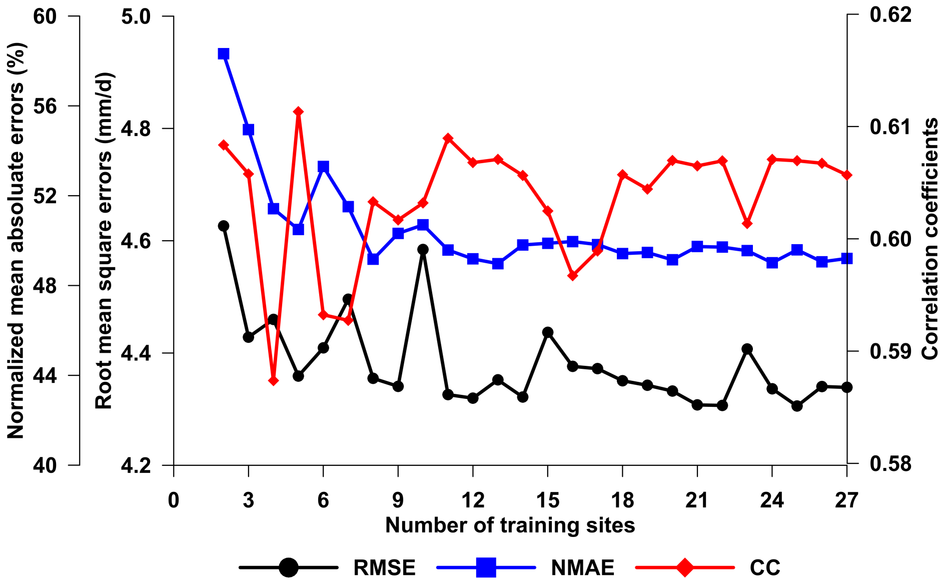

Fig. 4. Statistical error indices of the blended SPE at the validated grid locations in terms of different number of training sites in the warm season of 2014 in the NETP

Table 1: Summary of statistical error indices (i.e., RMSE, NMAE, and CC) in terms of bias-corrected and blended SPE with and without consideration of terrain feature as a covariate in the TSB method at the validated grid cells of NETP in the warm season of 2014.

Product	Type	RMSE (mm/d)	NMAE (%)	CC
BC-PER	No Terrain	5.03	58.9	0.416
	Terrain	5.02	58.7	0.418
BC-V7	No Terrain	5.08	58.0	0.403
	Terrain	5.06	57.5	0.410
BC-CMO	No Terrain	4.83	55.0	0.493
	Terrain	4.81	54.6	0.497
BC-IME	No Terrain	4.58	51.4	0.568
	Terrain	4.56	50.9	0.572
Blended SPE	No Terrain	4.36	49.7	0.603
	Terrain	4.34	49.2	0.606

Table 2: Summary of statistical error indices (i.e., RMSE, NMAE, and CC) in terms of three fusion methods (i.e., OOR, BMA, and TSB) at the validated grid cells of NETP in the warm season of 2014.

Method	RMSE (mm/d)	NMAE (%)	CC
OOR	5.63	59.2	0.547
BMA	5.44	57.6	0.595
TSB	4.34	49.2	0.606