

Responses to Review Comments on HESS-2020-43

We are thankful to the three anonymous reviewers for their comments and suggestions. We have addressed the comments point by point in the revision. In the text below, comments are repeated verbatim and the corresponding responses are in blue. In addition, we have made substantial improvements to the manuscript based on the comments and suggestions.

Anonymous Referee #1:

It is of importance for the scientific community to improve the retrieval accuracy of satellite precipitation estimates over complex terrains. This study proposed a flexible two-step approach to reduce the systematic errors of currently mainstream satellite precipitation products in the northeastern Tibetan Plateau. Evaluation results show that this approach effectively reduce the errors and biases of satellite retrievals. Overall, the paper is rich in content and technically sound. It can offer insightful references for both satellite precipitation produces and data users, especially for improving the retrieval algorithm over mountain regions. I consider it is clearly written and informative, and it should be of interest to a significant subset of HESS readers. Thus, I recommend it be accepted for publication, with just a few minor revisions.

Response: We thank this reviewer for the supportive comment.

First, I wonder why the new approach can effectively reduce the biases but not change the CC values. In the text, the authors should explain this point in more details.

Response: As bias correction is performed for each SPE in the first stage, the blended SPE has a low bias compared with the original SPE. We agree that the CC index does not improve significantly compared to the RMSE and NMAE values for the blended SPE. The CC between two data sets is a measure of how well they are related. In Stage 1, the mean parameter in the Student's t distribution is expressed as a linear regression of the original SPE. A linear assumption in the proposed model might fail to expect significant difference in the correlation. Thus, the RMSE and NMAE indices are also adopted to evaluate the performance of the proposed blending approach. We have given more explanations in the revised manuscript as suggested by this reviewer (Lines 190-192).

Second, the study area is limited within a squared rectangle. In practice, it is difficult to present the application potentials of new approach using such relatively small region as study domain (only like a case study). The gauge numbers are still not enough for validation. At least, the authors should discuss this in the section of conclusion.

Response: We thank this reviewer for the important comments. Perhaps we might not describe them very clearly in the original manuscript. We have rephrased the statements in the revised manuscript as pointed out by this reviewer. Please allow us to give an additional explanation. The experiment is selected in the northeastern Tibetan Plateau in terms of the area at around 5.8×10^5 km². To verify the performance of the new method, the original, bias-corrected and blended SPE are intercompared

at the random validation grid cells in the survey region (Lines 180-207). The time series of daily rainfall estimates and rainfall accumulations in terms of the original and blended SPE is further added at a selected validation location in the warm season of 2014 (Fig. 6; Lines 210-215). To mitigate the impact of validation locations, 10 randomly test is performed for the selection of validation grids (Lines 216-229). Also, a heavy rainfall event that occurred on September 22, 2014 is examined to quantify its performance in the extreme rainfall scenario (Lines 244-266). The two-step blending (TSB) method is also compared with the existing fusion algorithms (e.g., Bayesian model averaging (BMA) and One-outlier removed (OOR)) at the validation grids (Table 6; Lines 268-275). More details can be found in the revised manuscript.

Last but not at least, this manuscript needs to be further polish before publication.

Response: Polish as suggested.

Anonymous Referee #2:

This manuscript describes a two-step methodology to combine multiple satellite precipitation products to produce a blended daily precipitation estimate. The process involves first bias correcting the individual satellite QPE products relative surface rain gauges. Then, a Bayesian weighting is applied to blend the various QPE datasets into a single product. The approach is demonstrated on a small area in the northeastern part of the Tibetan Plateau over the 2014 warm season, as well as an individual heavy rain case. Overall the manuscript needs to be checked for correct grammar and usage, and the data and methods sections could be lengthened a bit to make things clearer and therefore reproducible (some specific suggestions for this below). Generally, with a few tweaks to the writing I feel this is publishable with minor revisions.

Response: We thank this reviewer for the great comments. The manuscript is carefully checked to avoid grammar and usage typos. The data and methods sections have been rephrased in the revised manuscript as required by this reviewer (Lines 68-174).

Specific Comments:

The manuscript would be much easier to follow if consistent terminology were used to refer to original SPE, bias corrected SPE, and blended SPE throughout.

Response: Revise as suggested.

Lines 75-80: Additional information about the data used is needed: Please specify the versions of IMERG and CMORPH you are using, and whether the IMERG is the near real time early, near real time late, or research/final runs. It is also interesting that you chose to use TMPA, which is no longer being produced and is generally very similar to IMERG. Additionally, IMERG, CMORPH, and TRMM-3B42 all have daily products available - why did you choose to use the 3-h products and (presumably) accumulate to daily? Finally, what method did you use to resample the IMERG?

Response: The CMORPH V1.0 research products and the Level 3 IMERG V03 final run products

are used in this study. We agree that TMPA is similar to IMERG, but the satellite retrieval algorithm between the two products are different. Considered that TMPA 3B42V7 shows a good performance in the TP, it is selected as an individual in this blending process. It is known that the daily scale of SPE is accumulated from the 3-h (TMPA, CMORPH) or 30-min (IMERG), we admit that we can directly use the daily scale instead of accumulation again from the 3-h products as suggested. The nearest neighbor interpolation is used to resampling the IMERG data. We have clarified these issues in the revised manuscript as pointed out by this reviewer (Lines 76-81).

Line 85: If you are using CMORPH V1.0, it also corrects using GPCP.

Response: Corrected as suggested.

Line 116: This equation would be easier to read if separated into 3 lines.

Response: Separated as suggested.

Line 162-170: Some discussion of the effects of comparing point data to somewhat low resolution gridded data is needed.

Response: We have rephrased this statement in the revised manuscript as suggested by this reviewer (Lines 86-88; Lines 106-107; Lines 279-284).

Line 182-183: It seems that the scatter is reduced for the blended product, but it has induced a high bias for low rain days and a low bias on heavy rain days. It's difficult to see if the bias is improved compared to the original SPE products.

Response: We thank this reviewer for the comment. Yes, the scatter is reduced for the blended SPE. We perform an additional comparison at the validation locations based on various rainfall intensities in Fig. 5 in the revision. Based on the TSB method, the blended SPE have been effectively dropped towards GR at the validation sites (Fig. 5b), especially for the rain intensity values less than 15 mm/d (Fig. 5c). Also, 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. Overall, this TSB method has its ability to exert benefits from SPE in terms of higher performances and mitigate poor impacts from the ones with lower quality. We have also rephrased this statement in the revised manuscript as pointed out by this reviewer (Lines 200-207).

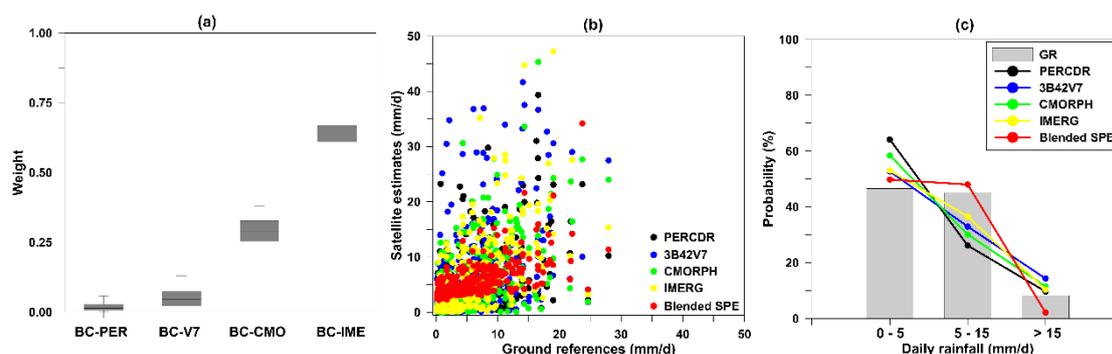


Figure 5: (a) The Box-Whisker plots of relative weights of the bias-corrected SPE in Stage 2; (b) Scatter plots between GR and various SPE (original and blended) at the validation grids in the warm season of 2014; (c) The PDF of daily rainfall in terms of the GR, original and blended SPE with various intensities at the validation grids in the warm season of 2014.

Line 213: I disagree with this statement. PRECDR is clearly very different from the others, and to this point in the manuscript has shown very litter value to be kept in consideration, and I think it is worth acknowledging this, then using the case study to point out that PRECDR can in fact be informative and on a case by case basis.

Response: We fully agree with this reviewer for the comments. We have added the statements in the revised manuscript as pointed by this reviewer (Lines 205-207).

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 TSB approach have been added in the revised manuscript as pointed out by this reviewer (Lines 96-174).

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: We thank this reviewer for the important comment. In the revised Method section, the process of bias adjustment and weight parameter estimation are explicitly described. The full details of bias adjustment and data merging stages are provided in the revised manuscript as suggested by this reviewer (Lines 96-174).

[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 purpose, 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. The goodness-of-fit of the Student's t distribution for the bias between GR and SPE is examined graphically by using a quantile-quantile

plot at the training grids (Fig. 3). It is found that all of them are close to the diagonal red line. It indicates that the selection of Student's t distribution is basically acceptable. We have also rephrased this statement in the revised manuscript (Lines 114-119).

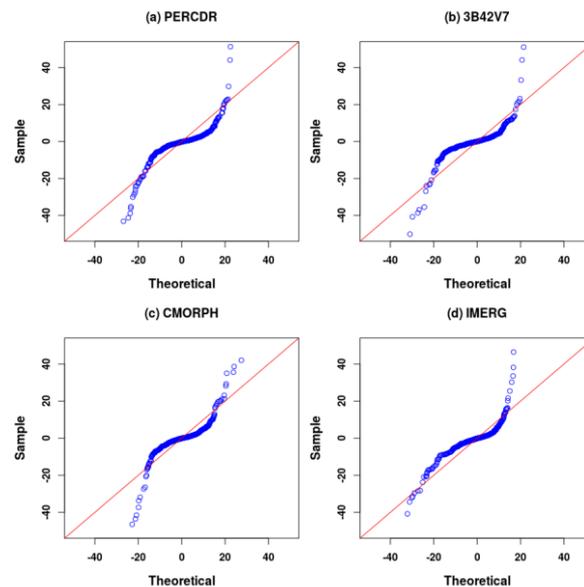


Figure 3. Quantile-quantile plots at training grid cells for the bias between GR and SPE, where (a) to (d) shows PERCDR, 3B42V7, CMORPH, and IMERG, respectively.

[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: We fully agree that it is challenging to bring information from 25 km grid to a point in a complex terrain region. Perhaps we didn't describe it very clearly in the original manuscript. 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 $0.25^\circ \times 0.25^\circ$ 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 admit that there is a scale gap between SPE and gauge observations. We have clarified the statements and add some discussions in the revised manuscript as pointed out by this reviewer (Lines 79-88).

[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 comment. We quantify the impact of elevation covariate on the bias-corrected and blended SPE performances as pointed out by this reviewer (Table 7). It is found that the consideration of elevation feature performs slightly better skill compared with

the model without terrain in this case study. We would like to admit that it is an initial exploration partly because we are dealing with the TP. We have rephrased the concerns in the revised manuscript as pointed out by this reviewer (Lines 279-284).

Table 7. Statistical error indices (i.e., RMSE, NMAE, and CC) of the bias-corrected and blended SPE before (No Terrain) and after (Terrain) the consideration of terrain information at the validated grid locations in the warm season of 2014 over the NETP.

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

[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.

Generally, the model performance of this new approach has been demonstrated based on various aspects in the revised manuscript. Please allow us to repeat them below: To verify the performance of the new method, the original, bias-corrected and blended SPE are intercompared at the random validation grid cells in the survey region (Lines 180-207). The time series of daily rainfall estimates and rainfall accumulations in terms of the original and blended SPE is further added at a selected validation location in the warm season of 2014 (Fig. 6; Lines 210-215). To mitigate the impact of validation locations, 10 randomly test is performed for the selection of validation grids (Lines 216-

229). Also, a heavy rainfall event that occurred on September 22, 2014 is examined to quantify its performance in the extreme rainfall scenario (Lines 244-266). The TSB method is also compared with the existing fusion algorithms (i.e., BMA and OOR) at the validation grids (Table 6; Lines 268-275). We thus consider that the evaluation analysis for long-term period and extended regions (e.g., TP) will be performed in a future study.

[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. In the revised manuscript, the TSB approach is compared with two existing fusion approach, i.e., BMA and OOR. 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 as compared with the other two fusion methods. We have added this comparison in the revised manuscript as kindly suggested by this reviewer (Lines 268-275).

Table 6. Statistical error indices (i.e., RMSE, NMAE, and CC) of three blending approach (i.e., OOR, BMA, and TSB) at the validated grid locations in the warm season of 2014 over the NETP.

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

[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 presents the statistical error summary over all the validation grids. We agree that there are more features if looking into individual sites than overall evaluation of the validated sites. The time series plot of daily rainfall estimates and rainfall accumulations of GR, original and blended SPE at a validated grid cell with a rain gauge labeled as ID 56173 is shown in Figure 6 as a demonstration example in the revised manuscript. This rain gauge, which is located at (32.8° N, 102.55°E, 3484 m), has the maximum rainfall record in the warm season of 2014 in the NETP. Visual analysis shows that the blended SPE provides reasonable rainfall and 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.

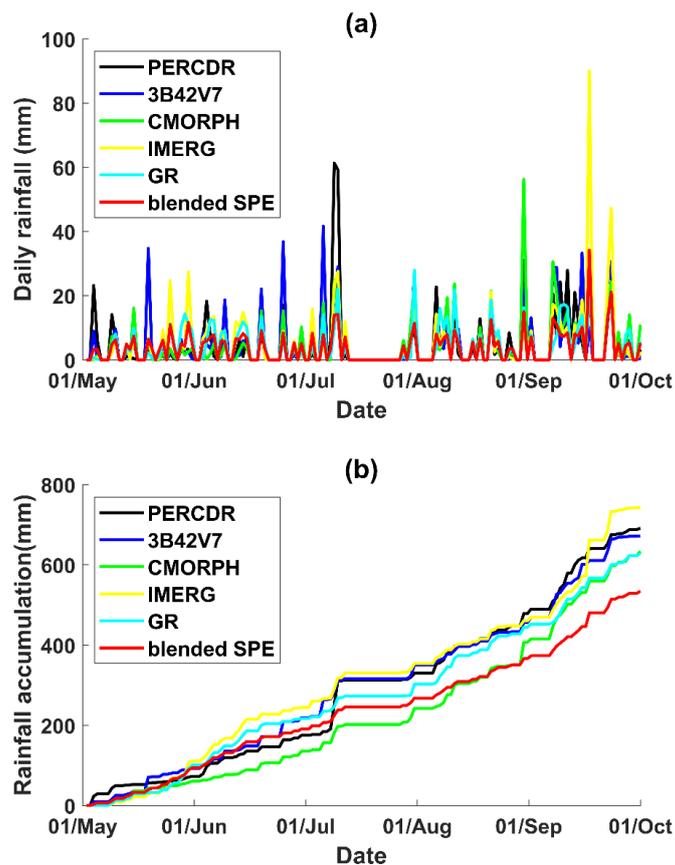


Figure 6. Time series of daily rainfall estimates and rainfall accumulations at a selected validation grid with the maximum rainfall record in the warm season of 2014: (a) daily rainfall estimates, and (b) rainfall accumulations.

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 bias-corrected SPE, i.e., BC-PER, for 10 random verified tests in the warm season of 2014 (Fig. 13). It shows that it is beneficial to involve the Stage 2 in the TSB method because the blended SPE performs better skill than BC-IME in the Stage 1 process. The primary reason is that the BW model is designed to integrate various types of bias-corrected SPE, which is limited in the BC model. Also, both blended SPE in Figure 13 show similar performances of the RMSE, NMAE, and CC indices. It implies that the TSB approach has an advantage of not impacted by the poor quality individuals (e.g., BC-PER), partly because the BW model can reallocate the contribution of the bias-corrected SPE based on their corresponding bias characteristics.

We have also rephrased the expressions in the revised manuscript as pointed out by this reviewer (Lines 200-215; Lines 286-295).

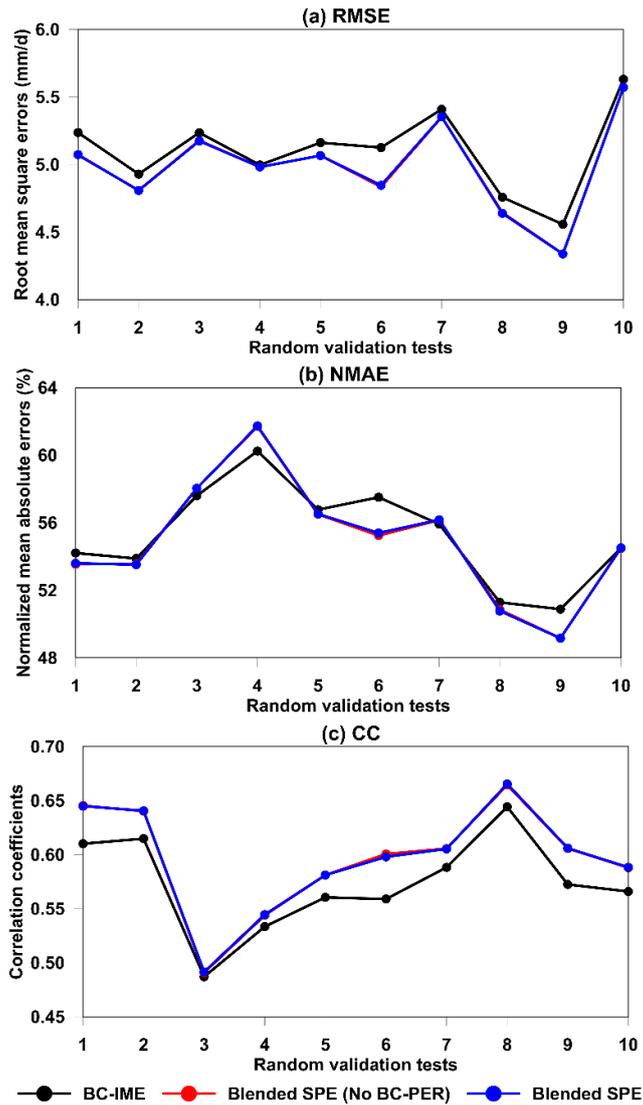


Figure 13. Statistical error indices (i.e., RMSE, NMAE, and CC) of the best-performed bias-corrected SPE (i.e., BC-IME, black) and blended SPE before (red) and after (blue) the removing of the worst-performed BC-PER for 10 random tests in the warm season of 2014 in the NETP.

[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 blow (Fig. 5c). 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 have rephrase this statement in the revised manuscript as pointed out by this reviewer (Lines 200-205).

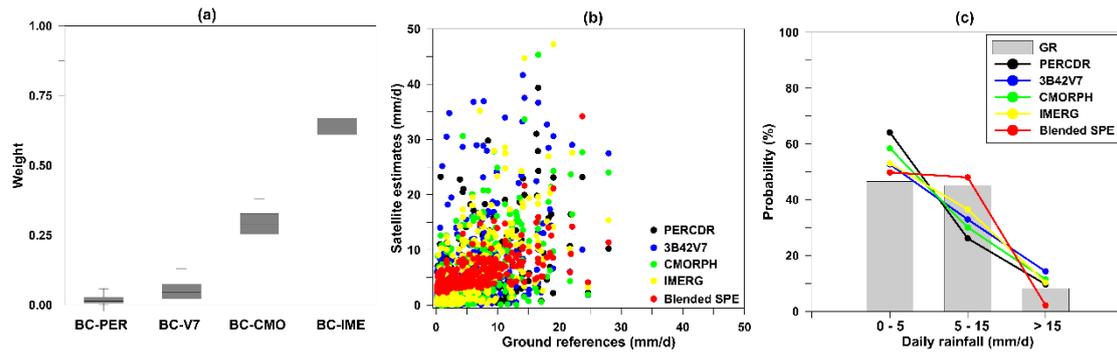


Figure 5: (a) The Box-Whisker plots of relative weights of the bias-corrected SPE in Stage 2; (b) Scatter plots between GR and various SPE (original and blended) at the validation grids in the warm season of 2014; (c) The PDF of daily rainfall in terms of the GR, original and blended SPE with various intensities at the validation grids in the warm season of 2014.

[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 comments. 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 impacted by the poor quality of SPE, partly because the BW model in Stage 2 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 bias-corrected SPE (i.e., BC-IME), and blended SPE before and after the removing of the worst-performed bias-corrected SPE (i.e., BC-PER) in this case, for 10 random verified tests in the warm season of 2014 in the NETP (Fig. 13). It is found that the blended SPE performs better skill than the simply bias correction with BC-IME, and both blended SPE products show similar performances of the RMSE, NMAE, and CC indices. It proves that it is beneficial to involve the Stage 2 process in the TSB method.

We have rephrased the related expressions in the revised manuscript as pointed out by this reviewer (Lines 286-295).

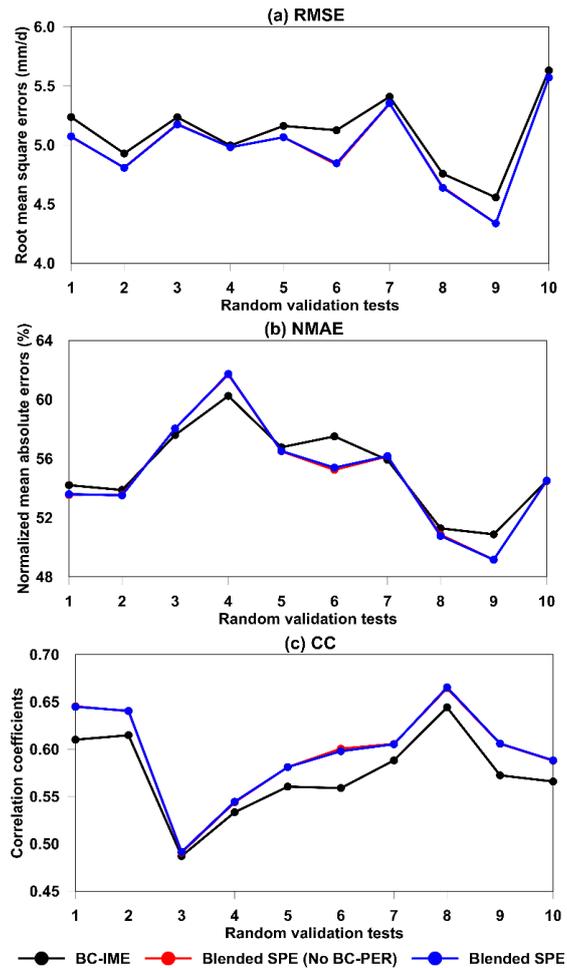


Figure 13. Statistical error indices (i.e., RMSE, NMAE, and CC) of the best-performed bias-corrected SPE (i.e., BC-IME, black) and blended SPE before (red) and after (blue) the removing of the worst-performed BC-PER for 10 random tests in the warm season of 2014 in the NETP.

[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 September 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 have replaced 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. Figure 8a (i.e., Fig. 10a in the revised manuscript) is the spatial map of the blended result which is weight summation of the original SPE from Figures. 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 have explained this issue in the revised manuscript as pointed out by this reviewer (Lines 238-242).

[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 training sites are needed to apply the TSB approach in a region with complex terrain and limited GR. The sensitivity analysis of the number of training grid cells on the performance of blended SPE at the validated sites is explored in Figure 14. 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 have rephrased this issue in the revised manuscript as pointed out by this reviewer (Lines 299-308).

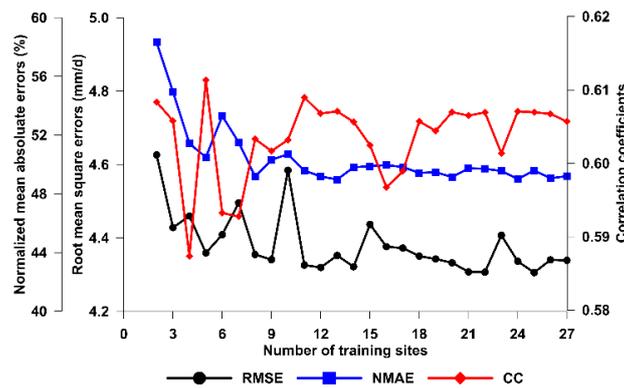


Figure 14. Statistical error indices (i.e., RMSE, NMAE, and CC) of the blended SPE at the validated grid locations in terms of different number of training sites in the warm season of 2014 in the NETP.

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

Response: Thoroughly checked as suggested.

A flexible two-stage approach for blending multiple satellite precipitation estimates and rain gauge observations: an ~~an~~ experiment in the northeastern Tibetan Plateau

Yingzhao Ma¹, Xun Sun^{2,3}, Haonan Chen^{1,4}, Yang Hong⁵, Yinsheng Zhang^{6,7}

¹Colorado State University, Fort Collins, CO 80523, USA

²Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China

³Columbia Water Center, Earth Institute, Columbia University, New York, NY 10027, USA

⁴NOAA/Physical Sciences Earth System Research Laboratory, Boulder, CO 80305, USA

⁵School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK 73019, USA

⁶Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, 100101, China

⁷CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing, 100101, China

Correspondence to: Xun Sun (xs2226@columbia.edu)

Abstract. Substantial biases exist in the ~~Satellite-satellite~~ ~~Precipitation-precipitation~~ ~~E~~estimates (SPE) over complex terrain regions and it has always been a challenge to quantify and correct such biases. The combination of multiple SPE and ~~rain ground-gauge observations-observations~~ would be beneficial to improve the ~~gridded~~ precipitation estimates. In this study, a ~~flexible two-step~~ ~~page blending (TSB)~~ approach is proposed, ~~which-by~~ firstly ~~ly~~ ~~reducesing~~ the systematic errors of each SPE ~~using rain-gauge observations as references based on a Bayesian correction model~~, and then ~~mergesing~~ the ~~improved-multi-bias-corrected~~ SPE with a Bayesian ~~weighting~~ model. In the ~~1st-first~~ stage, ~~the~~ ~~gauge-based~~ ~~references-observations~~ are assumed as a generalized regression function of SPE and terrain feature. In the ~~2nd-second~~ stage, the SPE weights are calculated based on the associated performances relative to ~~gauge-references~~ ~~ground references~~. ~~This proposed blending-TSB~~ method has the ability to exert benefits from ~~multi-the bias-corrected~~ SPE in terms of higher performance, and mitigate negative impacts from the ones with lower quality. In addition, Bayesian analysis is applied in the two phases by specifying the prior distributions on ~~the~~ model parameters, which enables to produce ~~the~~ posterior ensembles associated with their predictive uncertainties. The performance of the ~~two-step-blending-TSB~~ ~~approach-method~~ is ~~assessed-evaluated using-with~~ independent ~~rain-gauge observations-validation grids during-in~~ the warm season of 2014 in the northeastern Tibetan Plateau. Results show that the blended ~~multi-SPE~~ is significantly improved compared to the original ~~individuals~~ ~~SPE~~, especially ~~during-in the~~ heavy rainfall events. This study can also be expanded as a data fusion framework in the development of high-quality precipitation products in ~~high-cold-any~~ regions ~~of interest-characterized-by-complex-terrain~~.

Formatted: Superscript

1 Introduction

High-quality precipitation data is fundamental to understand the regional and global hydrological processes. However, it is still difficult to acquire accurate precipitation information in the mountainous regions, e.g., Tibetan Plateau (TP), due to limited ground sensors (Ma et al., 2015). The satellite sensors ~~can~~are capable of providing precipitation estimates at a large scale (Hou et al., 2014), but performances of available satellite products vary among different retrieval methods and climat~~ie~~ areas (Yong et al., 2015; Prat and Nelson, 2015; Ma et al., 2016). Thus, it is suggested to incorporate precipitation estimates from multiple sources into a fusion procedure with a fully consideration of the strength of individual members and associated uncertainty.

40 Precipitation data fusion was initially reported by merging radar-gauge rainfall in the mid-1980s (Krajewski, 1987). The Global Precipitation Climatology Project (GPCP) was an earlier attempt for satellite-gauge data fusion, which adopted a mean bias correction method and an inverse-error-variance weighting approach to develop a monthly, $0.25^\circ \times 0.25^\circ$ global precipitation data (Huffman et al., 1997). Another popular dataset, the Climate Prediction Center Merged Analysis of Precipitation (CMAP), included global monthly precipitation with a $2.5^\circ \times 2.5^\circ$ spatial resolution for a 17-year period by merging gauges, satellites and reanalysis data using the maximum likelihood estimation method (Xie and Arkin, 1997). Since then, several blending approaches have been developed to generate gridded rainfall product with higher quality by merging gauge, radar and satellite observations (e.g., Li et al., 2015; Beck et al., 2017; Xie and Xiong, 2011; Yang et al., 2017; Baez-Villanueva et al., 2020). Overall, those fusion methods follow a general concept by eliminating biases in satellite/radar-based data and then merging the bias-corrected satellite/radar estimates with point-wise gauge observations. However, these efforts might be insufficient for quantifying the predicted data uncertainty. Some blended estimates are also partially polluted by the poorly performed individuals (Tang et al., 2018).

~~Thus,~~ This paper develops a new blending approach that enhances the quantitative modelling of individual error structures, prevents potential negative impacts from lower-quality members, and enables an explicit description of the model's predictive uncertainty. In addition, a Bayesian concept for accurate rainfall estimate~~ion~~ is proposed based on these ~~conditions~~assumptions. The Bayesian analysis has the advantage of a statistically post-processing idea that could yield a predictive distribution with quantitative uncertainty (Renard, 2011). For instance, a Bayesian kriging approach, which assumes a Gaussian process ~~for~~of precipitation at any location and considers the elevation a covariate, is developed for merging monthly satellite and gauge precipitation data (Verdin et al., 2015). A dynamic Bayesian model averaging (BMA) method is applied for satellite precipitation data merging across the TP (Ma et al., 2018~~a~~). Given the flexible distribution of multiple sources of precipitation

biases in regions with complex terrain (Derin et al., 2019), continuous efforts ~~should be taken~~ are required to exert the potential merit of Bayesian approach on this critical issue.

In this paper, a two-stage blending (TSB) approach is described for ~~blending-combining multiple S~~satellite ~~P~~precipitation
65 ~~E~~estimates (~~multi-SPE~~) and point-based rain gauge observations. The ~~initial~~ experiment is performed ~~during-in~~ the warm season of 2014 in the northeastern TP (NETP), where a denser network of rain gauges is available compared to other regions of TP. The proposed ~~two-stage-blending~~ approach is ~~also~~ expected to help with the exploration of multi-source/scale precipitation data ~~merging-fusion~~ in other regions with complex terrain ~~but available SPE~~.

70 The ~~remainder of this~~ paper is organized below: ~~Section 2 gives a description of the experiment including the study region and precipitation data sources. Section 3 details the proposed two-stage-blending TSB approach. Section 2 gives a brief introduction of the study area and precipitation data sources. Section 3 details the proposed two-stage-blending approach.~~ Results and discussions are presented in Sections 4 and 5, respectively. The primary ~~summary and future work findings~~ are ~~provided-summarized~~ in Section 6.

75 2 Study area and dataset

~~The study domain is located in the upper Yellow River basin of NETPnortheastern TP (Fig. 1). As shown in the 90-m digital elevation data, the elevation ranges from 785 m in the northeast to 6252 m in the southeast. The total annual precipitation is around 500 mm and the annual mean temperature is 0.7°C (Cuo et al., 2013). To avoid snowfall contamination on the rain gauge observation in the cold season, the warm period from May 1- to September 30 in 2014 is selected as afor demonstration~~
80 ~~purpose.~~

~~Four mainstream SPE are used, including Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)- Climate Data Records (PERCDR) (Ashouri et al., 2015), Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis 3B42 V-version 7 (3B42V7) (Huffman et al., 2007), Climate Prediction Center (CPC) Morphing technique for the bias-corrected research product (CMORPH)-version 1.0 (CMORPH) (Joyce et al., 2004), and the Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) mission V03 Level 3 (IMERG)-version 05 Level 3-final run product (IMERG) (Huffman et al., 2018). The basic information of SPE is shown in Table 1. As the the-IMERG data has a spatial resolution of 0.10° x 0.10° resolution, and other SPE have a spatial resolution of 0.25° x~~

0.25°. To eliminate the scale difference, the IMERG data are resampled from 0.10° to 0.25° using the nearest neighbour interpolation method to eliminate the scale difference.

A ground network including 34 rain gauges are used in this study (Fig. 1). The rain gauge data are carefully checked to ensure its credibility (Shen and Xiong, 2016). All of them are independent from the Global Precipitation Climatology Center (GPCC) stations, which are used for bias adjustment/correction of the TRMM/GPM-era data (e.g., 3B42V7 and IMERG), and CMORPH (Huffman et al., 2007; Hou et al., 2014; Joyce et al., 2004). The rain gauge data are spatially interpolated with a 0.25° x 0.25° resolution in the study region of interest for 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. In addition, the ground cell/GR network is randomly classified into two parts: the black grids are used for training the model, and the red grid/cells are used for model verification (Fig. 1) (Fig. 1). In order to clarify the TSB method, the selection of calibration/training grid/cells is randomly repeated 10 times for the ground GR network, and the remaining ones are used for model validation. Additionally, Meanwhile, the TSB algorithm/method is applied on a heavy rainfall event that occurred on Sep/September 22, 2014 in the survey area to quantify its performance in extreme rainfall scenario. Local recycling plays as a premier role for the moisture sources of rainfall extremes in the NETP (Ma et al., 2020a). The Sep/September 22 rain case is a typical storm that can explain the local heavy rainfall patterns in the warm season. The performance of the TSB approach is further also compared with the two existing existing DBMA fusion algorithm/methods, i.e., BMA and One-outlier removed (OOR), which is/were previously developed by Ma et al. (2018a) and applied for long-term SPE data fusion in the TP (Ma et al., 2018; Shen et al., 2014b).

3 The TSB algorithm

3.1 Overview

This algorithm aims at developing a multi-source data merging framework in the region of interest to provide the best-available gridded precipitation product with rain gauge observations/GR and SPE in the region of interest. Let $R(s, t)$ denote near-surface precipitation at the gauge/GR cell s and the t^{th} day. The original SPE and bias-corrected SPE at the same grid and time are defined as $(Y_1(s, t), Y_2(s, t), \dots, Y_p(s, t))$ and $(Y'_1(s, t), Y'_2(s, t), \dots, Y'_p(s, t))$ at the same grid and time respectively. For simplicity, they are respectively replaced by $R, (Y_1, Y_2, \dots, Y_p)$, and $(Y'_1, Y'_2, \dots, Y'_p)$ respectively. The subscript p implies the number of SPE in terms of its value at 4 in this application, where PERCDR, 3B42V7, CMORPH and IMERG refer to Y_1, Y_2, Y_3, Y_4 , respectively.

Formatted: Font: Italic

The diagram of the TSB ~~algorithm~~method is shown in Figure 2. ~~Step~~page 1 is designed to mitigate the bias of the original SPE based on the GR at the training sites with a Bayesian correction (BC) procedure, where the assumption of probabilistic distribution for GR conditional on each SPE is not limited to Gaussian. Given complex terrain and 0.25° grid resolution, the topography is ~~considered~~added as a covariate in the BC process. In the ~~2nd~~second ~~step~~page, a Bayesian weight (BW) model is ~~developed~~used to merge the bias-corrected SPE. The BW model ~~has the ability to~~can exert benefits from bias-~~adjust~~corrected SPE with high performance and reduce poor impacts from the ones with lower quality. It also produces blended SPE with predictive uncertainty. The details of the TSB algorithm are described in Sections 23.2 and 23.3, respectively.

Formatted: Superscript

3.2 Step 1: Bias correction

In this ~~step~~page, we ~~focus~~perform on conditional modelling of GR on each ~~of the original~~SPE, i.e., on the probabilistic distribution $f(R)$ at the training sets to improve the ~~accuracy of the original~~SPE-accuracy. ~~With regard to the conditional distribution of GR,~~a flexible assumption (e.g., Lognormal, Gaussian, or Student's t distribution) for bias characteristics between GR and SPE is proposed. Given various SPE at different training sites, the specific probabilistic function is not limited to a certain distribution. ~~For demonstration purposes,~~we apply the Student's t distribution, with its mean parameter expressed as a linear regression of the original SPE. Also, ~~the~~goodness-of-fit of the Student's t distribution for the bias between GR and SPE ~~at the training grid cells~~is examined graphically by using a quantile-quantile plot at the training sets (Fig. 3). It is found that ~~where~~they are close to the diagonal red line. A Student's t distribution is thus adopted with its mean parameter expressed as a linear regression of SPE. It is

We ~~parameterized~~parameterized the Student's t distribution ~~as~~as follows:

$$R \sim \text{Student}(v_i, \mu_i, \sigma_i) \quad (1)$$

$$\mu_i = \alpha_i + \beta_i * Y_i + \gamma_i * Z \quad (2)$$

where v_i is known as degree of freedom, μ_i and σ_i stand for sample mean and variance, respectively; the parameter μ_i is correlated with the intensity value of the i^{th} SPE (Y_i) and associated terrain feature (e.g., elevation) (Z). To ~~avoid~~ignore the scale factor ~~data anomaly~~, the normalized elevation feature ranging from 0 to 1 is used as the terrain feature in the regression model of Eq. (2); ~~is~~normalized and its value ranges from 0 to 1 after the normalization. Also, $\theta = \{v_i, \alpha_i, \beta_i, \gamma_i, \sigma_i\}$ ~~is~~is summarized as a ~~the~~parameter set, ~~s~~which enables to write the likelihood function or probability density function (PDF) from Eqs. (1) and (2) conditional on θ and Y_i as:

Formatted: Font: Italic

$$f(R|\theta, Y_i) = \frac{\Gamma(v_i+1/2)}{\Gamma(v_i/2)} \frac{1}{\sqrt{v_i\pi}\sigma_i} \left(1 + \frac{1}{v_i} \left(\frac{R - (\alpha_i + \beta_i Y_i + \gamma_i Z)}{\sigma_i}\right)^2\right)^{-(v_i+1)/2} \quad (3)$$

According to the Bayes's theorem (Gelman et al., 2013), the posterior distribution of parameter set θ given GR and SPE data, and the prior distribution of parameters $f(\theta)$ can be expressed as:

$$f(\theta|R, Y_i) \propto f(R|\theta, Y_i)f(\theta) \quad (4)$$

150 The estimation of the posterior distribution $f(\theta|R, Y_i)$ in Eq. (4) is challenging as its dimension grows with the number of parameters (Renard, 2011). Here, ~~However, the~~ Markov Chain Monte Carlo (MCMC) technique ~~employed in the Stan programming language can be~~ used to address this issue (Gelman et al., 2013). ~~Consider~~ Given that the assumption of the weakly informative priors ensures the Bayesian inference in an appropriate range (Ma et al., 2020b), the priors of $f(\theta)$ are initialized as uniform distribution with $\alpha_i, \beta_i, \gamma_i$ at real numbers in Eq. (5), and with v_i, σ_i at a lower-bound zero of real numbers in Eq. (6).

$$\alpha_i, \beta_i, \gamma_i \sim Uniform(-\infty, +\infty) \quad (5)$$

$$v_i, \sigma_i \sim Uniform(0, +\infty) \quad (6)$$

Based on the estimated parameter set θ above, the next step is to calculate ~~each of~~ the bias-corrected SPE R^* at any new site of the domain at the same period, which ~~It can~~ can be quantitatively simulated from its posterior predictive distribution in Eq.

160 (7) using the ~~associated~~ original SPE Y_i^* , and training data R, Y_i :

$$f(R^*|Y_i^*, R, Y_i) = \int f(R^*, \theta|Y_i^*, R, Y_i) d\theta \quad (7)$$

Following the rule of joint probabilistic distributions, the right term inside the integral of Eq. (7) is written as:

$$f(R^*, \theta|Y_i^*, R, Y_i) = f(R^*|Y_i^*, R, Y_i, \theta)f(\theta|Y_i^*, R, Y_i) \quad (8)$$

Given that Y_i^* is independent with R and Y_i , the first term of the right side in Eq. (8) is transformed as:

$$165 \quad f(R^*|Y_i^*, R, Y_i, \theta) = f(R^*|Y_i^*, \theta) \quad (9)$$

Since the parameters θ are dependent upon the training data R, Y_i , the second term of the right side in Eq. (8) is expressed as:

$$f(\theta|Y_i^*, R, Y_i) = f(\theta|R, Y_i) \quad (10)$$

Therefore, the posterior predictive distribution of R^* in Eq. (7) is written below:

$$f(R^*|Y_i^*, R, Y_i) = \int f(R^*|Y_i^*, \theta) f(\theta|R, Y_i) d\theta \quad (11)$$

170 Since there is no general way to calculate the associated integral in Eq. (11), it is performed again using the MCMC iterations. A numerical algorithm is suggested below: n_{sim} is assumed as the replicate of the post-convergence MCMC samples, and the predicted samples for R^* in Eq. (11) are iterated ($j = 1, \dots, n_{sim}$) as follows:

Formatted: Font: Italic

1) Calculate the model parameters θ from Eqs. (1) to (6);

2) Compute the mean parameter μ_i^* from the regression model of Eq. (2), i.e., $\mu_i^* = \alpha_i + \beta_i * Y_i^* + \gamma_i * Z_i^*$;

175 3) Generate the derived quantity from the posterior distribution of R^* in Eq. (11).

3.3 Stepage 2: Data merging

Formatted: Heading 2

180 On the basis of In Stepage 1 in Section 3.2, the median value of the posterior samples is adapted as the bias-corrected SPE. Here, we redefine the bias-corrected SPE as Y_i^l ($l = 1, 2, \dots, p$), respectively. In this part, it aims to merge the bias-adjusted SPE at each grid-cell in the domain. The formulas of blending the bias-corrected SPE are shown below:

$$B = \sum_{l=1}^p Y_i^l * w_l + \varepsilon \quad (12)$$

$$\sum_{l=1}^p w_l = 1 \quad (13)$$

$$\varepsilon \sim Normal(0, \sigma) \quad (14)$$

$$w_l \sim Uniform(0,1), l = 1, \dots, p \quad (15)$$

185
$$\sigma \sim Uniform(0, +\infty) \quad (16)$$

where B is the blended SPE; w_l ($l = 1, 2, \dots, p$) stands for the relative weight of the l^{th} bias-corrected SPE; ε is the residual error. Ideally, the blended SPE at the training site s and time t should beshould-be close to GR, i.e., $R(s, t)$. Thereby, the model parameters δ , including w_l ($l = 1, 2, \dots, p$) and σ are able to will be estimated based on the GR and bias-corrected SPE at the training sites under a Bayesian analysis. With regard to the conditional distribution of blended SPE on the bias-corrected SPE, 190 we propose a Gaussian distribution for the residual error modelling. The corresponding PDF is written as follows:

$$f(B|\delta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{B - \sum_{l=1}^p Y_i^l * w_l}{\sigma}\right)^2\right) \quad (17)$$

Formatted: Right

The calculation process of δ is similar with to the parameter estimation described in Stepage 1. After the parameters δ are estimated, similar to Eqs. (7) to (11), the blended SPE at any site and time t can be derived with the bias-corrected SPE and corresponding weights using the MCMC iterations. Finally, we can obtain spatial patterns of blended SPE and rain-gauge networks in terms of the median, standard deviation (SD) and associated credible intervals (e.g., 5% and 95% quantiles) in the regions of interest.

2 Study area and dataset

The selected study domain is located in the upper Yellow River basin of northeastern TP (Fig. 1). As shown in the 90-m digital elevation data, the elevation ranges from 785 m in the northeast to 6252 m in the southeast. The total annual precipitation is around 500 mm and the mean annual air temperature is 0.7°C (Cuo et al., 2013). To avoid snowfall contamination in the cold season, the warm period from May 1st to September 30th in 2014 is selected in this demonstration study.

Four popular SPE are used, including Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)—Climate Data Records (PERCDR) (Ashouri et al., 2015), Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis 3B42 Version 7 (3B42V7) (Huffman et al., 2007), Climate Prediction Center (CPC) Morphing technique for the bias-corrected product (CMORPH) (Joyce et al., 2004), and the Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) mission (IMERG) (Huffman et al., 2018) (Table 1). Since IMERG data has a spatial resolution of 0.10° × 0.10°, and other SPE (i.e., PERCDR, 3B42V7 and CMORPH) have a spatial resolution of 0.25° × 0.25°. The IMERG data are resampled from 0.10° to 0.25° so as to match the other individuals before performing the two-stage blending.

A ground rain-gauge network including 34 stations are used in this study (Fig. 1). The gauge data are carefully checked to ensure its creditability (Shen and Xiong, 2016). All of them are independent from the Global Precipitation Climatology Center (GPCC) stations, which are used for bias adjustment of the TRMM/GPM-era data, such as 3B42V7 and IMERG (Huffman et al., 2007; Hou et al., 2014). In addition, the gauge network is randomly classified into two parts: the black dots are used for training the model, and the remaining ones are for model verification. In order to demonstrate the reliability of the proposed two-stage blending approach, the selection of training sites is randomly repeated for 10 times to further examine the blending performance. In addition, the proposed blending algorithm is applied on a heavy rainfall case of Sep 22, 2014 in the NETP, to quantify the performance during heavy rainfall scenario. Local recycling performs as a premier role for the moisture sources

Formatted: Superscript

Formatted: Superscript

of rainfall extremes in the NETP (Ma et al., 2020). This case is a typical storm that could stand for the local heavy rainfall patterns to a large extent during the warm season.

3 A two-stage blending algorithm

3.1 Overview

225 This algorithm aims at developing a multi-source data merging framework so as to provide the best available precipitation product in any region of interest. Let $R(s, t)$ denote near surface precipitation at gauge site s and the t^{th} day in a year. The original and bias adjusted multi-SPE at the same location and time are defined as $(Y_1(s, t), Y_2(s, t), \dots, Y_p(s, t))$ and $(Y_1^t(s, t), Y_2^t(s, t), \dots, Y_p^t(s, t))$, respectively. For simplicity, they are accordingly replaced by $R, (Y_1, Y_2, \dots, Y_p)$, and $(Y_1^t, Y_2^t, \dots, Y_p^t)$. Noted that the value of p equals to 4 in this study, where PERCDR, 3B42V7, CMORPH and IMERG refers to Y_1, Y_2, Y_3, Y_4 , respectively.

The diagram of the proposed two-stage blending approach is shown in Figure 2. Stage 1 is designed to correct the systematic errors of individual SPE using point-based rain-gauge observations (training sites) as ground references, where the assumptions of various probabilistic distribution for gauge references conditional on each SPE are not limited to Gaussian prototype. The impact of topography is also considered. In the 2nd step, a Bayesian weight model is applied to blend the improved multi-SPE. It has the ability to exert benefits from multi-SPE of higher performance and mitigate negative impacts from the ones with lower quality. It is expected to produce posterior blended results associated with their predictive uncertainties in the survey region.

240 The details of the two-stage blending algorithm are described in Sections 3.2 and 3.3, respectively.

3.2 Stage 1: Bias adjustment

A generalized regression function between gauge references, individual SPE, and terrain features is proposed in the 1st stage. Because the bias of SPE generally follows a skew Normal distribution, it is important to fit an appropriate function. In this paper, a Student's t distribution is assumed for modelling of gauge observations conditional on the individual SPE. It is written as:

$$R|Y_i \sim \text{Student}(v_i, \alpha_i + \beta_i * Y_i + \gamma_i * Z, \sigma_i), \alpha_i, \beta_i, \gamma_i \in R, v_i, \sigma_i \in R^+ \quad (1)$$

where $\theta = \{\nu_i, \alpha_i, \beta_i, \gamma_i, \sigma_i\}$ are model parameter sets in order to adjust the i^{th} SPE. $(\alpha_i + \beta_i * Y_i + \gamma_i * Z)$ represents the sample mean and Z is the associated collection of covariates (e.g., topography). More specifically, the normalized elevation is used as a covariate in this experiment. These parameters are real numbers and ν_i, σ_i are positive. It should be noted that some other distributions (e.g., Lognormal, Normal) are also examined but there are no obvious improvements in terms of the bias-adjusted result compared to Student's t distribution for this test.

~~It should be noted that some other distributions (e.g., Lognormal, Normal) are also examined but there are no obvious improvements in terms of the bias-adjusted result compared to Student's t distribution for this test.~~

Based on the gauge observations and multi-SPE at the training sites, model parameters for each SPE could be estimated within a Bayesian analysis using the Markov Chain Monte Carlo (MCMC) technique (Gelman et al., 2013). Next, it is to calculate each of the bias-adjusted SPE at any new site (s^t) of the domain at the same period. The conditional distribution of bias-adjusted SPE at any new site is mathematically defined as:

$$f(R_{s^t}|Y_t) = \int f(R_{s^t}, \theta | Y_t) d\theta \quad (2a)$$

$$= \int f(R_{s^t} | \theta) f(\theta | Y_t) d\theta \quad (2b)$$

where the posterior distribution of $R_{s^t}|Y_t$ from Eq. 2 can be simulated numerically based on the calculated parameter samples θ at the training sites using the MCMC samplings. We further assume N as the size of post-convergence MCMC samples. The above process is repeated N times and produces a predictive posterior distribution at any new site s^t and time t .

3.3 Stage 2: Data merging

On the basis of Stage 1, this part is designed to blend the updated multi-SPE at each grid cell of the domain. With regard to the individual SPE, the median value of the posterior samples from Stage 1 is assumed as the new SPE. Here, we redefine the bias-adjusted multi-SPE as $Y_i^t (i = 1, 2, \dots, p)$, respectively.

The formulas of blending the bias-adjusted multi-SPE are shown below:

$$B = \sum_{i=1}^p Y_i^t * w_i + \varepsilon, w_i \in R(0,1), \varepsilon \in R^+ \quad (3)$$

$$\sum_{i=1}^p w_i = 1 \quad (4)$$

Formatted: Justified

Formatted: Font: Not Bold

Formatted: Right, Line spacing: Double

Formatted: Line spacing: Double

where B means the blended result; w_i ($i=1,2,\dots,p$) stands for the relative weight of the i^{th} -SPE, respectively, and their values range from 0 to 1; ϵ is the residual error, whose value is positive real number. Ideally, the blended multi-SPE, i.e., B , at the training site s and time t should be close to gauge references $R(s, t)$. Thereby, the weight parameters, including w_i ($i=1,2,\dots,p$) and ϵ are able to be estimated at the training sites based on gauge observations and new multi-SPE within a Bayesian analysis using the MCMC technique.

As the weight parameters are successfully derived above, similar to Eq. 4, the blended result at any new sites at time t are calculated based on the retrieved new multi-SPE and corresponding optimal weights. Finally, we can obtain spatial patterns of blended multi-SPE and point-based rain gauge observations in terms of the median, standard deviation (SD) and associated confidence intervals (e.g., 5% and 95% quantiles) in regions of interest.

4 Results

To assess the performance of the proposed **two-stage blending (TSB)** method, several statistical error indices including **Root Mean Square Errors (RMSE)**, **Normalized Mean Absolute Errors (NMAE)**, and the Pearson's **Correlation Coefficients (CC)** are used in this study. The specific formulas of these metrics can be found in [the literature \(e.g., Chen et al., 2019 among others\)](#).

4.1 Evaluation of **the original, bias-corrected, and blended adjustment of multi-SPE at the validated grids**

Compared to the **the gauge references (GR)**, the original multi-SPE including (i.e., **PERCDR, 3B42V7, CMORPH and IMERG**) show **significant large** biases at the **independent-validation grid sites/cells** over in the NETP during the warm season of 2014 (Table 2). Their statistical error metrics **including RMSE, NMAE, and CC** range from 6.59-8.07 mm/d, 63.2-83.5%, and 0.403-0.5768, **in terms of RMSE, NMAE, and CC**, respectively. **3B42V7 performs has** the worst **skill** with the highest RMSE of 8.07 mm/d and the highest NMAE at of 83.5%, and the lowest CC of 0.403. **IMERG shows the best performance** in terms of the lowest NMAE at of 63.2% and highest CC at 0.5768 **among the four-SPEs, which presents its superiority compared with the other SPE in the survey area**. **It seems that the satellite retrievals need to be further clarified with regard to the mainstream SPE in the NETP.**

After the bias-Based on the BC model adjustment of each SPE, the updated multi-bias-corrected SPE (i.e., BC-PER, BC-V7, BC-CMO and BC-IME) show great improvement-have better agreements with GR in data quality at the validation grids in the

300 experiment. during this experiment (Fig. 3). These changes result in better agreement of SPE with rain-gauge measurements at
the validated sites in the NETP. Both their RMSE scores range from 4.56 to 5.06 mm/d, and
correspondingly decrease by 27~37.3%, and their NMAE scores vary from 50.9 to 58.7%, and decline by 19.1 to 31.1%,
respectively. As compared with the original multi-SPE, the updated ones decrease by 27~37.3% and 19.1~31.1% with respect
to RMSE and NMAE, respectively. Moreover, meanwhile, their CC index values of four SPE vary range from 0.4210 to 0.5768
305 after bias adjustment (Table 2), which slightly increases as compared to the original SPE. Considering that the linear
assumption of mean parameter in the Student's t distribution at Stage 1 might fail to expect significant difference in the
correlation, the CC value does not improve effectively for the bias-corrected SPE. After Stage 2, the blended SPE is closer to
the GR in terms of RMSE, NMAE and CC at 4.34 mm/h, 49.2%, and 0.606, respectively, compared with both the original and
bias-corrected SPE at the validation grid cells (Fig. 4). It is also found that 3B42V7 improves the most in terms of its RMSE
310 decreasing from 8.07 mm to 5.06 mm using the step 1's method. Basically, it implies that the proposed bias-adjusted algorithm
occurred in phase 1 is very effective for reducing systematic errors of four involved SPE in the warm season of 2014 in the
NETP.

4.2 Blending multi-SPE and independent validation

To test the performance of the proposed two-step blending approach, the blended multi-SPE at the validation sites are further
315 examined. As shown in Figure 3, the fusion result is closer to the ground reference in terms of RMSE, NMAE and CC,
compared to the individual SPE. The RMSE and NMAE indices values of the merging blended data SPE decrease by
34.1~65.4% and 27.1~41.1%, respectively, compared to the individual SPE, while and the CC index value increases by
6.7~50.4%, accordingly, compared with the original SPE (Table 2). As compared to with the bias-corrected multi-SPE, the
performance of the blended data the blended SPE increases by 5.1~14.2%, 3.3~16.2%, and 5.9~47.8% in terms of RMSE,
320 NMAE and CC, respectively. That is to say, It is found that the merged blended precipitation SPE in the warm season of 2014
at the validated sites of NETP exhibits higher quality at the validation grids after Stage 2, due to the ensemble contribution
of the bias-corrected SPE with their relative weights at 0.019, 0.052, 0.289, and 0.640, respectively. The BC-IME and BC-
PER have the highest and lowest weights, respectively, and the contributions of BC-V7 and BC-CMO on the blended SPE
rank between BC-IME and BC-PER (Fig. 5a). Based on the TSB approach, merging the bias-corrected multi-SPE using the
325 optimal relative weights shown in Figure 4a. The blended data SPE has been effectively dropped towards the gauge
reference GR at the validation grids (Fig. 5b), especially for the rain intensity values less than 15 mm/d (Fig. 5c), which
is evidenced from the scatterplots in terms of red dots in Figure 4b. Also, there is an overestimation in the original SPE but an
underestimation in 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. Overall, this TSB method has its ability to exert benefits from SPE in terms of higher
330 performances and mitigate poor impacts from the ones with lower quality. Meanwhile, BC-PER seems to be clearly very

different from the others (Fig. 5a), and to this point in the study has shown little value to be kept in consideration in the merging process. However, it is worth noting that PERCDR can in fact be informative and on a case by case basis.

335 The results presented in Figures 4 and 5 are an average assessment of the TSB algorithm at all the validation grids, which can possibly homogenize some individual feature. The time series plot of daily rainfall estimates and rainfall accumulations of the GR, original and blended SPE at a validation grid with a rain gauge labeled as ID 56173 is shown in Figure 6 as a demonstration example. This rain gauge, which is located at (32.8° N, 102.55°E, 3484 m), has the maximum rainfall record in the warm season of 2014 in the NETP. Visual analysis of Figure 6 shows that the blended SPE provides reasonable rainfall compared to the original SPE. Also, the blended SPE has a better skill in terms of RMSE at 4.95 mm/d compared with the original SPE
340 including PERCDR (10.71), 3B42V7 (9.76), CMORPH (8.0), and IMERG (10.49), respectively.

These improvements prove the significant superiority of the two-step blending method on reducing the systematic errors of the original multi-SPE and supplying higher daily precipitation in the warm season of 2014 over the NETP.

345 The best-performed merging result is due to the ensemble contributions of the bias-corrected multi-SPE. The optimally relative weights are 0.019, 0.052, 0.289, and 0.640 with respect to PERCDR, 3B42V7, CMORPH and IMERG, respectively. It shows that the bias-adjusted IMERG and PERCDR contributes the highest and lowest weights, respectively, in this blending process, and the contributions of the other SPE rank between IMERG and PERCDR accordingly. As the bias-adjusted IMERG shows the best performance among all the individuals, it proves that higher informative SPE shows more positive impact on the blended result under this two-step fusion approach. It is further concluded that this blending method has its ability to exert
350 benefits from multi-SPE in terms of higher performance and mitigate poor impacts from the ones with lower quality.

4.3 Time series of original, bias-corrected, and blended SPE

4.3.42 Model clarification with random validation sitegrids

355 Figures 57 and 8 shows the statistics of evaluation scores of RMSE, NMAE, and CC for the original multi-SPE and blended estimates at the validation gridsites with 10 random split of the gauge stationslocations. For each test, 7 gauge-grid sites are randomly selected from the 34 grid sitecells and used for model verification, and the remaining 27 gauge-sitesgrid sites are used for training the model fitting.

As for the blended resultSPE, it performs similar skill scores at the the independent validation grids foramong the 10 random tests, but It also shows better skillsperformance in terms of RMSE, NMAE, and CC, at which are 4.34~5.57 mm/h, 49.2~61.7%, and 0.492~0.6765, respectively, compared to with the raw multi-original SPE at each random experimentest (Fig. 7). Statistically, the averaged-mean values of RMSE, NMAE and CC for the blended data-SPE are 4.98 mm/h, 54.9% and 0.60597, respectively while four (Table 3). The averaged improvement ratios of RMSE for the blended dataSPE are 35.1%, 33.7%, 19.6% and 32.1% compared to the PERCDR, 3B42V7, CMORPH and IMERG, respectively (Table 4), and S similar performance is seen from the NMAE scores, where with their the mean average improvement ratios are of 29.8%, 30.1%, 17.0% and 21.3%, respectively (Table 4) for the four SPE. The 10 random tests clarify that the blended SPE has a higher accuracy of gridded precipitation which receives different credits from various SPE on an event basis.

As seen in Figure 6, the blended result shows a significant improvement over the original multi-SPE in the survey area, especially for PERCDR and 3B42V7. It is concluded that the biases of multi-SPE could be significantly reduced as the impacts of bias functions are well considered in the proposed two-step blending algorithm.

4.5 Model comparison with the existing fusion algorithm

4.4.3 Model application in spatial domain

It is important to explore the Bayesian ensembles at any unknown site in the study domain. Each SPE can capture the spatial pattern of daily mean precipitation in the warm season, but might fail in the representation of precipitation amount in the NETP (Fig. 9), partly because of the satellite retrieval bias in complex terrain and limited GR network. ThereforeHere, the two-step blendingTSB approach is applied in four spatially distributed SPE (i.e., PERCDR, 3B42V7, CMORPH, IMERG) from Figures 7a to 7d to obtain the blended estimates-SPE in terms of daily mean precipitation in the warm season of 2014 for over the whole study domaindomain (not only at the gauge stations).

There is an overestimation for most of the original SPE, and the bias of the blended SPE is reduced based on the TSB approach. Spatial maps of the merging predictions and the associated predictive uncertainties including SD, 5% and 95% quantiles are shown along with the gauge observations (Figure 8).

Formatted: Normal

All of the multi-SPE have the ability in capturing the spatial patterns of daily mean precipitation during the warm season, but might fail in the representation of precipitation amounts in the NETP (Figure 7), likely because of the satellite retrieval biases in complex terrain and limited ground validation network. The spatial patterns of blended multi-SPE are shown in Figure 8a, which It is found that the blended SPE shows a better performance in terms of magnitude and distribution in the study area (Fig. 10a). Higher values disappear from the map except in southwest corner. The possible reason is that daily mean rainfall is the highest in southwest corner for each SPE, and higher value still exists after the TSB process. Meanwhile, the predictive uncertainties including SD, 5% and 95% quantiles are displayed from Figures 10b to 10d in order to illustrate the fusion variance.

has a similar spatial pattern with a higher precipitation amounts in the southwest compared with the individual SPE. Based on the proposed blending approach, the fusion estimate performs a higher adjustment compared to the original SPE. It is expected to show a better performance in terms of magnitude and distribution in the study area. Moreover, the predictive uncertainties are displayed from Figures 8b to 8d so as to illustrate the blending variance. In total, this study confirms the priority of exploring daily precipitation in spatial at higher accuracy and quantifying the associated uncertainty in the study domain.

4.5.4 Model performance during a heavy rainfall case

Accurate precipitation on extreme weather is very important for flood hazard mitigation. Here, we investigate the utility of this two-step blending-TSB approach on a heavy rainfall case of SepSeptember 22, 2014 in-over the NETP (Fig. 9a11a). The relative weights of BC-PER, BC-V7, BC-CMO, and BC-IME for the blended data are 0.464, 0.123, 0.112 and 0.301, respectively, during-on this particular heavy rainfall event (Fig. 911b).

Table 54 reports the evaluation statistics reflecting the blended model performance during this heavy rainfall case, where the RMSE, NMAE and CC values/indices of the individual-original SPE range from 6.28~10.48 mm/d, 40.6~59.5%, and 0.6986~0.820, respectively. Overall, As compared to the individual/original SPEs, the merged product has lower RMSE of 4.13 mm/d, and lower NMAE of 27.4%, and higher CC of 0.850, respectively. In other words, the RMSE and NMAE values of the blended result-SPE decrease by 34.2~60.6% and 32.5~53.9%, respectively, while-and the CC index correspondingly increases by 3.4~23.9% on this heavy rainfall case compared to the original SPE. The two-step blending approach has a great influence on the performance of SPE in terms of rainfall extremes in the warm season of the NETP.

415 The blended model performance is further explored at three gauge [sites-cells](#) (i.e., IDs 56171, 56152, 56182) with the top three
420 [daily rainfall](#) records ~~in terms of daily rainfall amounts~~ on ~~Sep~~September 22, 2014 (Fig. [9-11a](#)). Figure [12-9](#) shows the
[Probabilistic-Density-Function \(PDF\)](#) curves of blended samples ~~of at~~ the above three grid sites ~~during in~~ this [rainfall case](#) event.
It ~~aims to~~ demonstrates the blended performance on quantifying the predictive uncertainty on rainfall extremes ~~in the survey~~
[region at each grid](#). At ID 56171, the estimated rainfall derived from [PERCDR](#), [3B42V7](#), [CMORPH](#) and [IMERG](#)~~the original~~
425 [SPE](#) are 19.8 mm ([PERCDR](#)), 35.3 mm ([3B42V7](#)), 26 mm ([CMORPH](#)), and 40.2 mm ([IMERG](#)), respectively. 3B42V7 and
IMERG shows an overestimation, while PERCDR and CMORPH underperform the daily rainfall at the corresponding pixel
(Fig. [10a-12a](#)). Based on the [two-step-blending-TSB](#) methods, the median and SD values of the merging estimates are 24.1
mm/d, and 4.4 mm/d, respectively. At IDs 56152 and 56182, the median/SD values of blended [multi-SPE](#) are 24.3/5.0 mm/d
and 21.9/4.5 mm/d, respectively. ~~and they~~ ~~As learned from Figures 10b and 10e,~~ ~~the medians of the blended result at IDs~~
430 ~~56152 and 56182~~ are very close to the [gauge-observations](#) [GR](#) with the daily amounts of ~~in terms of~~ 24.6 mm, and 23.1 mm,
respectively (Figs. [12b](#) and [12c](#)). ~~It shows that~~ These analyses reveal that ~~this the~~ proposed [two-step fusion-TSB algorithm](#)
[method](#) can ~~not only~~ quantify its predictive uncertainty, but also improve the daily rainfall amount even on [rainfall](#)
[extremes](#) provides a posterior inference and quantifies its predictive uncertainties on the heavy rainfall events. It is confirmed
435 that the proposed two-step blending method is able to improve the daily precipitation amounts even during rainfall extremes
in the NETP.

4.5 Model comparison with other fusion methods

To assess the performance of the proposed TSB algorithm, it is beneficial to compare the TSB result with the existing fusion
435 approach. Herein, we compare it with the BMA and OOR methods at the validation grids of NETP (red pixels in Figure 1) in
the warm season of 2014 and their statistical error summary is shown in Table 6. It is found that the TSB method performs
better skill with the RMSE, NMAE and CC values of 4.34 mm/d, 49.2%, and 0.606, respectively, compare with the other two
fusion methods. OOR shows the worst performance in terms of RMSE, NMAE, and CC at 5.63 mm/d, 59.2%, and 0.547,
respectively. As learned from model comparison among the three methods in this case, the TSB method has an advantage for
combining the SPE and reducing the bias of the individuals.

440 5 Discussion

This study proposes a flexible two-step blending algorithm for merging multi-satellite and rain-gauge precipitation data at
daily scale, aiming to provide a more accurate precipitation datasets in regions with complex terrain. In spite of the superior
performance of [the TSB algorithm](#), some issues still need to be considered in the practical applications:

445 Because of limited knowledge on the influences of complex terrain and local climate on the rainfall patterns in the study area,
the elevation feature is ~~merely considered~~ considered in the first stage. [Table 7 quantifies the impact of elevation covariate on
the bias-corrected and blended SPE performances at the validation grids. It is found that the inclusion of elevation feature
provides slightly better skill compared with the results without terrain information in this experiment. Considering that -It is
noted that deep convective systems occurring near the mountainous area have an effect on the precipitation cloud \(Houze,
450 2012\), which should be](#) more attempts are required to improve ~~in futu~~the orographic ~~adjustment of individual~~ precipitation [in
the TP](#) in future.

[The fusion application is based on four mainstream SPE, and it is found that BC-IME and BC-PER show the best and worst
performances among the bias-corrected SPE. It raises a question that why not simply apply the first stage of bias correction
455 and then select the best-performed bias-corrected SPE as the final product. To address this issue, we investigate the statistical
error differences among the BC-IME and blended SPE before and after the removing of BC-PER for 10 random verified tests
in the warm season of 2014 in the NETP \(Figure 13\). It shows that it is beneficial to involve the Stage 2 in the TSB method
because the blended SPE performs better skill than the best-performed bias-corrected SPE \(i.e., BC-IME\) in the Stage 1
process. The primary reason is that the BW model is designed to integrate various types of bias-corrected SPE, which is limited
460 in the BC model. Also, both blended SPE in Figure 13 show similar performances of the RMSE, NMAE, and CC indices. It
implies that the TSB approach has an advantage of not impacted by the poor quality individuals \(e.g., BC-PER\), partly because
the BW model can reallocate the contribution of the bias-corrected SPE based on their corresponding bias characteristics.](#)

In addition, as calculating the blended result at any new sites, the model parameters derived from the training grid sites are
465 assumed to be applicable in the whole domain. Since ~~the domain of this study is not very large and~~ we have a relatively dense
[rain-gaugeGR network in the survey region](#), the current assumption ~~seems to be~~ acceptable according to the performance of
the blended ~~data~~SPE. ~~However, -It is helpful to give some guideline on how many training sites are needed to apply the TSB
approach in a region with complex terrain and limited GR. The sensitivity analysis of the number of training grid cells on the
performance of blended SPE at the validation grids is explored in Figure 14. As the number of training sites is increasing, there
470 is a decreasing trend for the RMSE and NMAE values, but a slight increasing trend for the CC value. Except for an anomaly
with the No. 23, it seems that the performance of the blended SPE becomes similar as the number of training sites increases to
21. We admit that the more information from the ground observations, it would be more beneficial for the blended gridded
product in the region of interest. ~~It~~It is noted that, if extended to the TP or global scale, the extension of model parameters [and](#)~~

475 [training sites](#) should be carefully considered. For instance, there are few gauges installed in the western and central TP (Ma et al., 2015), it might be a potential risk to directly apply this fusion algorithm for these regions.

480 The [goal-aim](#) of this study is not to model rainfall process in a target domain, but to propose an idea to extract valuable information from available [multi-satellite-sourcesSPE](#) and provide more reliable [gridded](#) precipitation in high-cold regions with complex [topographyterrain](#). Considering its spatiotemporal differences and the existence of many zero-value records, rainfall is extremely difficult to observe and predict (Yong et al., 2015; Bartsotas et al., 2018). With regard to the probability of rainfall occurrence, a zero-inflated model, which is coherent with the empirical distribution of rainfall [amount-data](#), is expected to [further-improve](#) the [proposed-two-step-fusionTSB](#) algorithm. [In-additionAlso](#), hourly or even instantaneous precipitation intensity is extremely vital for flood prediction, which [should-should-bebe-specifically-considered-designed-when](#) when extending this [fusion-framework](#) [in-the-next-step](#).

485 6 Summary and prospects

This study proposes a [two-step-blendingTSB](#) algorithm for multi-SPE data fusion. A preliminary experiment is conducted [over](#) in the NETP using four [mainstream-mainstream-SPE](#) (i.e., PERCDR, 3B42V7, CMORPH, and IMERG) to demonstrate the performance of this [fusion-TSB](#) approach, [including-PERCDR, 3B42V7, CMORPH, and IMERG](#). Primary conclusions are summarized below:

490

(1) This [blending-TSB](#) algorithm [has-two-stages-and-involves-the-BC-and-BW-models](#). It is found that [this-blended-method-is-designed-with-high-flexibility-which](#) is capable of involving a group of [multi-SPE-with-their-biases-that-may-following](#) different [probabilistic-distributionsPDF-curves-conditional-on-ground-references](#). [In-additionMeanwhile](#), it provides a convenient way to [compare-quantify](#) the [merging-fusion](#) performance and [further-quantify](#) the associated [fusion-uncertainty](#).

495

(2) The [experimentease-studies](#) shows that the [merged-blended-SPEprecipitation](#) has better skill scores compared [with-to](#) the [individual-original-SPE](#) at the [independent-validation-siteslocations](#). The 10 random [verification](#)-tests also confirms the superiority of the [proposed-two-step-blendingTSB](#) algorithm. The performance of this fusion [algorithm-method](#) is further demonstrated using [for-the-a](#) heavy rainfall event. In addition, [the-TSB-method-outperforms](#) another [two-fusion-methods\(i.e.,BMA-and-OOR\)](#).

500

(3) The ~~experiment-application~~ proves that this algorithm can allocate the contribution of individual SPE on the blended ~~prediction-result~~ because it is capable of ingesting useful information from uneven individuals and alleviating potential negative impacts from the poorly performing members.

Overall, this work provides an opportunity for ~~blending-merging multi-SPE products~~ in high-cold region with complex terrain. It is expected to promote the development of higher quality precipitation product in the remotely high-cold regions with widely available satellite precipitation retrievals. The ~~exploration of model reliability-evaluation analysis of this two-step blending TSB algorithm method for long-term period and extended regions (e.g., TP) at larger scale (e.g., the TP) and in terms of higher temporal resolution (e.g., hourly) will be performed~~ should be pursued in a ~~future-future study~~ study.

Data availability

The gauge data are from China Meteorological Data Service Center (<http://data.cma.cn>). The PERCDR data are obtained from <http://www.ncei.noaa.gov/data/precipitation-persiann/>; the 3B42V7 data are obtained from <https://pmm.nasa.gov/data-access/downloads/trmm>; the CMORPH data are obtained from ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0; the IMERG data are obtained from <https://pmm.nasa.gov/data-access/downloads/gpm>.

Author contributions

YM and XS conceived the idea; XS, YH, and YZ acquired the project and financial supports. YM conducted the detailed analysis; HC, XS and YZ gave comments on the analysis; all the authors contributed to the writing and revisions.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This study is supported by the National Key Research and Development Program of China (No. 2017YFA0603101), Strategic Priority Research Program (A) of CAS (No. XDA2006020102), and the National Natural Science Foundation of China (No. 91437214). We also thank Dr. Ning Ma from the Institute of Tibetan Plateau Research, Chinese Academy of Sciences, for the great comments and suggestions.

References

- Ashouri, H., Hsu, K. L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R., and Prat O. P.:
530 PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate
Studies, *Bull. Amer. Meteor. Soc.*, 96, 69-83, 2015.
- Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Beck, H. E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K.,
Giraldo-Osorio, J. D., and Thinh, N. X.: RF-MEP: A novel Random Forest method for merging gridded precipitation
products and ground-based measurements, *Remote Sens. Environ.*, 239, 111606, 2020.
- Bartsotas, N. S., Anagnostou, E. N., Nikolopoulos, E. I., and Kallos, G.: Investigating satellite precipitation uncertainty over
535 complex terrain, *J. Geophys. Res.-Atmos.*, 123, 5346-5369, 2018.
- Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. J., Martens, B., and de Roo A.: MSWEP: 3-
hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data, *Hydrol. Earth Syst.
Sci.*, 21, 589-615, 2017.
- Chen, H., Cifelli, R., Chandrasekar, V., and Ma, Y.: A flexible Bayesian approach to bias correction of radar-derived
540 precipitation estimates over complex terrain: model design and initial verification, *J. Hydrometeorol.*, 20, 2367-2382, 2019.
- Cuo, L., Zhang, Y., Gao, Y., Hao, Z., and Cairang L.: The impacts of climate change and land cover/use transition on the
hydrology in the upper Yellow River Basin, China, *J. Hydro.*, 502, 37-52, 2013.
- [Derin, Y., Anagnostou, E., Berne, A., Borga, M., Boudevillain, B., Buytaert, W., Chang, C., Chen, H., Deirieu, G., Hsu, Y.,
545 Lavado-Casimiro, W., Manz, B., Moges, S., Nikolopoulos, E., Sahu, D., Salerno, F., Rodriguez-Sanchez, J., Vergara, H.,
and Yilmaz, K.: Evaluation of GPM-era global satellite precipitation products over multiple complex terrain regions, *Remote
Sensing*, 11, 2936, 2019.](#)
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin D. B.: *Bayesian Data Analysis-Third Edition*,
CPC Press. 2013.
- Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi,
550 T.: The Global Precipitation Measurement Mission, *Bull. Amer. Meteor. Soc.*, 95, 701-722, 2014.
- Houze., R. A.: Orographic effects on precipitating clouds, *Rev. Geophys.*, 50, RG1001, 2012.
- Huffman, G., Adler, R., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J. E., McNab, A., Rudolf, B., and
Schneider, U.: The global precipitation climatology project (GPCP) combined precipitation dataset, *Bull. Amer. Meteor.
Soc.*, 78, 5-20, 1997.

- 555 Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker E. F.:
The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates
at fine scales, *J. Hydrometeorol.*, 8, 38-55, 2007.
- Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Tan, J., and Xie,
P.: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG), Algorithm
560 Theoretical Basis Document (ATBD) Version 5.2, NASA/GSFC, Greenbelt, MD, USA, 2018.
- Joyce, R. J., Janowiak, J. E., Arkin, P. A. and Xie, P.: CMORPH: A method that produces global precipitation estimates
from passive microwave and infrared data at high spatial and temporal resolution, *J. Hydrometeorol.*, 5, 487-503, 2004.
- Krajewski, W. F.: Cokriging radar-rainfall and rain gage data, *J. Geophys. Res.*, 92, 9571-9580, 1987.
- Li, H., Hong, Y., Xie, P., Gao, J., Niu, Z., Kirstetter, P. E., and Yong, B.: Variational merged of hourly gauge-satellite
565 precipitation in China: preliminary results, *J. Geophys. Res.-Atmos*, 120, 9897-9915, 2015.
- Ma, Y., Zhang, Y., Yang, D., and Farhan S. B.: Precipitation bias variability versus various gauges under different climatic
conditions over the Third Pole Environment (TPE) region, *Int. J. Climatol.*, 35, 1201-1211, 2015.
- Ma, Y., Tang, G., Long, D., Yong, B., Zhong, L., Wan, W., and Hong, Y.: Similarity and error intercomparison of the GPM
and its predecessor-TRMM Multi-satellite Precipitation Analysis using the best available hourly gauge network over the
570 Tibetan Plateau, *Remote Sensing*, 8, 569, 2016.
- Ma, Y., Hong, Y., Chen, Y., Yang, Y., Tang, G., Yao, Y., Long, D., Li, C., Han, Z., and Liu, R.: Performance of optimally
merged multisatellite precipitation products using the dynamic Bayesian model averaging scheme over the Tibetan Plateau,
J. Geophys. Res.-Atmos., 123, 814-834, 2018.
- Ma, Y., Lu, M., Bracken, C., and Chen, H.: Spatially coherent clusters of summer precipitation extremes in the Tibetan
575 Plateau: Where is the moisture from? *Atmos. Res.*, 237, 104841, 2020a.
- [Ma, Y., Chandrasekar, V., and Biswas, S. K.: A Bayesian correction approach for improving dual-frequency precipitation
radar rainfall rate estimates, *J. Meteor. Soc. Japan*, 98, 2020b.](#)
- Prat, O. P., and Nelson, B. R.: Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain
580 gauge data sets at daily to annual scales (2002–2012), *Hydrol. Earth Syst. Sci.*, 19, 2037-2056, 2015.
- Renard, B.: A Bayesian hierarchical approach to regional frequency analysis, *Water Resour. Res.*, 47, W11513, 2011.

Sanso, B., and Guenni, L.: Venezuelan rainfall data analysed by using a Bayesian space time model, *J. R. Stat. Soc. C - Appl.*, 48, 345-362, 1999.

Shen, Y., and Xiong, A.: Validation and comparison of a new gauge-based precipitation analysis over mainland China, *Int. J. Climatol.*, 36, 252-265, 2016.

Shen, Y., Xiong, A., Hong, Y., Yu, J., Pan, Y., Chen, Z., and Saharia, M.: [Uncertainty analysis of five satellite-based precipitation products and evaluation of three optimally merged multi-algorithm products over the Tibetan Plateau. *Int. J. Remote Sens.*, 35, 6843-6858, 2014.](#)

Tang, Y., Yang, X., Zhang, W., and Zhang, G.: Radar and Rain Gauge Merging-Based Precipitation Estimation via Geographical-Temporal Attention Continuous Conditional Random Field, *IEEE Trans Geosci Remote Sens.*, 56, 1-14, 2018.

Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., and Wu, H.: Modeling errors in daily precipitation measurements: additive or multiplicative, *Geophys. Res. Lett.*, 40, 2060-2065, 2013.

Verdin, A., Rajagopalan, B., Kleiber, W., and Funk, C.: A Bayesian kriging approach for blending satellite and ground precipitation observations, *Water Resour. Res.*, 51, 908-921, 2015.

Xie, P., and Arkin, P.: Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, *Bull. Amer. Meteor. Soc.*, 78, 2539-2558, 1997.

Xie, P., Xiong, A.-Y.: A conceptual model for constructing high-resolution gaugesatellite merged precipitation analyses, *J. Geophys. Res.-Atmos.* 116, D21106, 2011.

Yang, Z., Hsu, K., Sorooshian, S., Xu, X., Braithwaite, D., Zhang, Y., and Verbist, K.M.J.: Merging high-resolution satellite-based precipitation fields and point-scale rain gauge measurements - a case study in Chile, *J. Geophys. Res.-Atmos.*, 122, 5267-5284, 2017.

Yong, B., Liu, D., Gourley, J. J., Tian, Y., Huffman, G. J., Ren, L., and Hong, Y.: Global View Of Real-Time Trmm Multisatellite Precipitation Analysis: Implications For Its Successor Global Precipitation Measurement Mission, *Bull. Amer. Meteor. Soc.*, 96, 283-296, 2015.

Figure and Table Captions

Table 1: Basic information of [the original multi-SPE](#) used in this study.

Table 2: Summary of statistical error indices (i.e., RMSE, NMAE, and CC) of the original, bias-~~adjust~~corrected, and blended ~~multi-SPE (i.e., PERCDR, 3B42V7, CMORPH, and IMERG)~~ at the validation gridsites of NETP in the warm season of 2014.

610 **Table 3:** Summary of the mean values of RMSE, NMAE and CC for the original and blended ~~multi-SPE (i.e., PERCDR, 3B42V7, CMORPH, and IMERG)~~ at 10 random verified tests in the warm season of 2014 over the NETP.

Table 4: ~~Summary of the m~~Mean improvement ratios of statistical error indices of the blended ~~multi-SPE~~, in terms of RMSE, NMAE and CC ~~as compared to~~with the original ~~SPEPERCDR, 3B42V7, CMORPH, and IMERG~~ at 10 random verified tests in the warm season of 2014 over the NETP.

615 **Table 5:** Summary of statistical error indices (i.e., RMSE, NMAE, and CC) for the original and blended ~~multi-SPE (i.e., PERCDR, 3B42V7, CMORPH, and IMERG)~~ during a heavy rainfall event over the NETP on ~~Sep~~September 22, 2014.

Table 6: 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.

620 **Table 7:** Summary of statistical error indices (i.e., RMSE, NMAE, and CC) for bias-corrected and blended SPE with and without consideration of terrain feature as a covariate in the TSB method at the validation grids of NETP in the warm season of 2014.

Figure 1: Spatial map of the topography and ~~ground-references~~GR network used in the study, where 27 black cells are used for model calibration and 7 red cells are for model verification.

625 **Figure 2:** The diagram of the proposed TSB algorithm.

Figure 3: Quantile-quantile plots at training sets for the bias between GR and SPE, where (a) to (d) shows PERCDR, 3B42V7, CMORPH, and IMERG, respectively.

Formatted: Justified

630 **Figure 1:** Overview of the topography and gauge observation network used in the study, where 27 gauges (black dots) are used for training and 7 (red dots) are used for independent verification.

Figure 2: The diagram of the proposed two-step blending algorithm.

Figure 34: Intercomparisons of statistical error indices for the original, bias-adjusted, corrected, and blended multi-SPE at the validation grids during-in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

635 **Figure 45:** (a) The Box-Whisker plots of relative weights of the bias-corrected adjusted multi-SPE in Stage 2; (i.e., PERCDR, 3B42V7, CMORPH and IMERG) in the stage-2 process; (b) Scatter plots between GR and various SPE; (c) The PDF of daily rainfall in terms of the GR, original and blended SPE with various intensities at the validation grids in the warm season of 2014.

Figure 6: Time series of daily rainfall estimates and rainfall accumulations at a selected validation grid with the maximum rainfall record in the warm season of 2014: (a) daily rainfall estimates, and (b) rainfall accumulations.

640 -
Figure 57: Statistical error indices of the original and blended multi-SPE (i.e., PERCDR, 3B42V7, CMORPH, and IMERG) for 10 random tests during-in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

645 **Figure 68:** The Box-Whisker plots of improvement ratios of statistics for the blended multi-SPE compared with the original SPE, including PERCDR, 3B42V7, CMORPH, and IMERG for 10 random tests during-in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

Figure 79: Spatial patterns of the daily mean precipitation derived from in terms of the original multi-SPE during-in the warm season of 2014: (a) PERCDR, (b) 3B42V7, (c) CMORPH, and (d) IMERG.

Figure 810: Spatial patterns of the blended multi-SPE in terms of (a) median, (b) standard deviation SD, (c) 5% and (d) 95% quantiles of daily mean precipitation during-in the warm season of 2014.

650 **Figure 911:** (a) Spatial pattern of gauge-based measurements during a heavy rainfall case over the NETP on SepSeptember 22, 2014 over the NETP, where the site IDs 56171, 56152 and 56182 report the top three daily rainfall amounts of 30.4 mm, 24.6 mm and 23.1 mm, respectively; (b) the corresponding Box-Whisker plots of relative weights of the bias-adjusted multi-individual-SPE (i.e., PERCDR, 3B42V7, CMORPH and IMERG) in the stage-Stage 2 process.

655 **Figure 102:** The PDF curves of blended samples-SPE and the corresponding median value at three gauge sites during a heavy rainfall case on SepSeptember 22, 2014: (a) ID 56171, (b) ID 56152, and (c) ID 56182. The individual-original SPE and GR

Formatted: Justified

Formatted: Justified

including PERCDR, 3B42V7, CMORPH, and IMERG as well as gauge-based measurement at each pixel are also indicated in each subfigure.

Figure 13. Statistical error indices (i.e., RMSE, NMAE, and CC) of the best-performed bias-corrected SPE (i.e., BC-IME, black) and blended SPE before (red) and after (blue) removing the worst-performed BC-PER for 10 random verified tests in the warm season of 2014 in the NETP.

Figure 14: Statistical error indices (i.e., RMSE, NMAE, and CC) of the blended SPE at the validation grid locations in terms of different number of training sites in the warm season of 2014 in the NETP.

660

665 **Table 1:** Basic information of [multi](#)~~the original~~-SPE used in this study.

Short name	Full name and details	Temporal resolution	Spatial resolution	Input data	Retrieval algorithm	References
PERCDR	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) Climate Data Record (CDR)	Daily	0.25° x 0.25°	2014.5-2014.9	Adaptive artificial neural network	<i>Ashouri et al., 2015</i>
3B42V7	TRMM Multi-satellite Precipitation Analysis (TMPA) 3B42 Version 7	3 hourly	0.25° x 0.25°	2014.5-2014.9	GPCC monthly gauge observation to correct this bias of 3B42RT	<i>Huffman et al., 2007</i>
CMORPH	Climate Prediction Center (CPC) MORPHing technique for bias-corrected product version 1.0	3 hourly	0.25° x 0.25°	2014.5-2014.9	Morphing technique	<i>Joyce et al., 2004</i>
IMERG	Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) mission V03 Level 3 final run product	0.5 hourly	0.10° x 0.10°	2014.5-2014.9	2014 version of the Goddard profiling algorithm	<i>Huffman et al., 2018</i>

Formatted Table

Formatted: Justified

670 **Table 2:** Summary of statistical error indices (i.e., RMSE, NMAE, and CC) of the original, bias-corrected, and blended multi-SPE (i.e., PERCDR, 3B42V7, CMORPH, and IMERG) at the validation gridsites of NETP in the warm season of 2014.

Type/Product	RMSE (mm/d)	NMAE (%)	CC
PERCDR Original	7.36	74.6	0.4216
3B42V7 Original	8.07	83.5	0.403
CMORPH Original	6.59	67.5	0.493
IMERG Original	7.18	63.2	0.5768
Adjusted BC-PER	5.02	58.7	0.4218
Adjusted BC-V7	5.06	57.5	0.410
Adjusted BC-CMO	4.81	54.6	0.50497
Adjusted BC-IME	4.56	50.9	0.572
Blended SPE	4.34	49.2	0.606
Blended multi-SPE	4.34	49.2	0.61

Formatted: Justified, Line spacing: 1.5 lines

Formatted: Left, Line spacing: single

Formatted: Line spacing: single

Formatted Table

Formatted: Left, Line spacing: single

Formatted: Line spacing: single

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Left, Line spacing: single

Formatted: Line spacing: single

Formatted: Left, Line spacing: single

Formatted: Line spacing: single

Formatted: Font: Not Bold

Formatted: Left, Line spacing: single

Formatted: Font: Not Bold

Formatted: Line spacing: single

Formatted: Left, Line spacing: single

Formatted: Line spacing: single

Formatted: Font: Not Bold

Formatted: Left, Line spacing: single

Formatted: Font: Not Bold

Formatted: Line spacing: single

Formatted: Left, Line spacing: single

Formatted: Line spacing: single

Formatted: Font: Not Bold

Formatted: Left, Line spacing: single

Formatted: Line spacing: single

Formatted: Font: Not Bold

Formatted: Left, Line spacing: single

Formatted: Font: 10 pt, Not Bold

Formatted: Line spacing: single

Formatted: Font: Not Bold

Formatted: Line spacing: single

Formatted Table

Table 3: Summary of the mean values of RMSE, NMAE and CC for the original and blended [multi-SPE](#) (i.e., [PERCDR](#), [3B42V7](#), [CMORPH](#), and [IMERG](#)) at 10 random verified tests in the warm season of 2014 [overin](#) the NETP.

Product	RMSE (mm/d)	NMAE (%)	CC
PERCDR	7.72	78.5	0.378
3B42V7	7.57	78.9	0.433
CMORPH	6.21	66.3	0.513
IMERG	7.37	70.0	0.572
Blended SPE	4.98	54.9	0.597

675

Table 4: Summary of the mean improvement ratios of statistical error indices of the blended multi-SPE, in terms of RMSE, NMAE, and CC including RMSE, NMAE and CC as compared to with the original SPEPERCDR, 3B42V7, CMORPH, and IMERG at 10 random verified tests in the warm season of 2014 over the NETP.

	Index	PERCDR	3B42V7	CMORPH	IMERG
Improvement	RMSE (mm/d)	35.1	33.7	19.6	32.1
	NMAE (%)	29.8	30.1	17.0	21.3
Ratio (%)	CC	61.3	38.2	17.5	4.3

Table 5: Summary of statistical error indices (i.e., RMSE, NMAE, and CC) for the original and blended multi-SPE (i.e., PERCDR, 3B42V7, CMORPH, and IMERG) during a heavy rainfall event over the NETP on SepSeptember 22, 2014.

Product	RMSE (mm/d)	NMAE (%)	CC
PERCDR	6.28	40.6	0.822
3B42V7	10.12	59.5	0.686
CMORPH	6.80	45.6	0.734
IMERG	10.48	53.3	0.805
Blended SPE	4.13	27.4	0.850

Table 6: Summary of statistical error indices (i.e., RMSE, NMAE, and CC) for three fusion methods (i.e., OOR, BMA, and TSB) at the validation grids of NETP in the warm season of 2014.

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

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold

690 **Table 7** Summary of statistical error indices (i.e., RMSE, NMAE, and CC) for bias-corrected and blended SPE with and without consideration of terrain feature as a covariate in the TSB method at the validation grids 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

Formatted: Line spacing: single

Formatted: Font: Not Bold

Formatted: Left, Line spacing: single

Formatted: Line spacing: single

Formatted: Left, Line spacing: single

Formatted: Font: Not Bold

Formatted: Line spacing: single

Formatted: Font: Not Bold

Formatted: Left, Line spacing: single

Formatted: Font: 10 pt, Not Bold

Formatted: Line spacing: single

Formatted: Left, Line spacing: single

Formatted: Font: Not Bold

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Not Bold

Multi- SPMethod	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

Formatted: Line spacing: Multiple 1.08 li

Formatted: Line spacing: single

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold

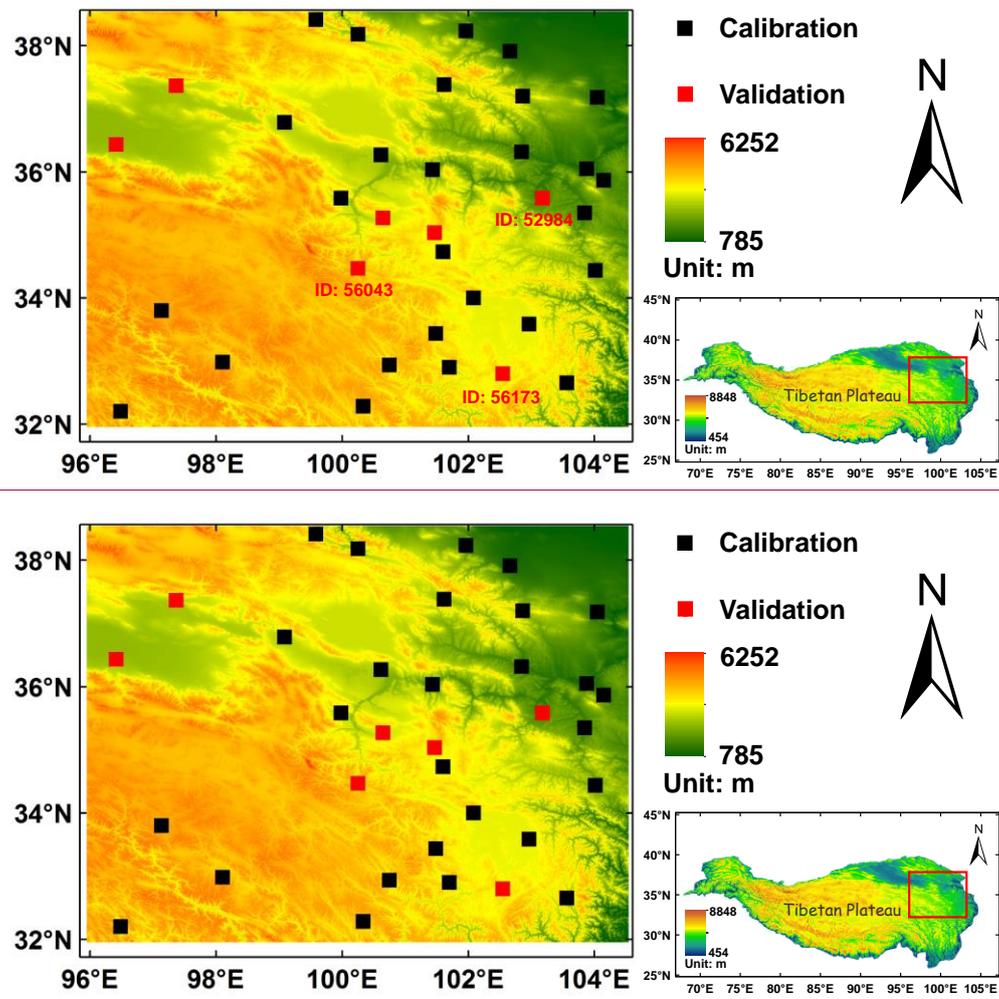
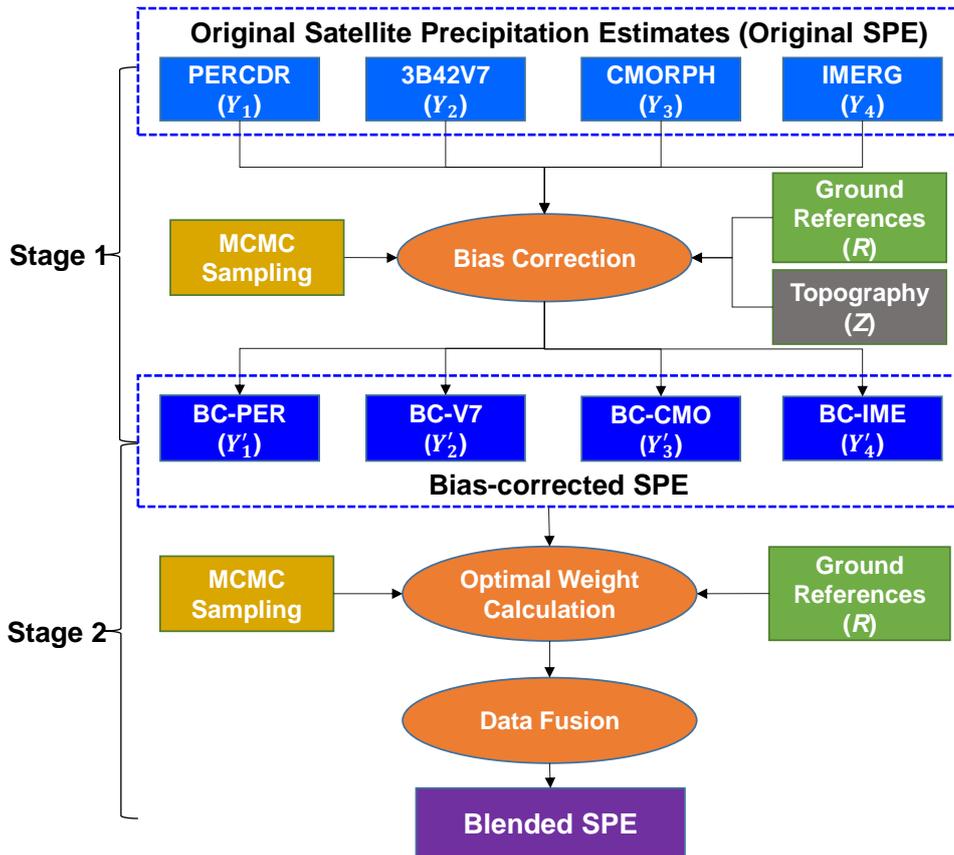


Figure 1: Spatial map of the topography and [ground-references](#) GR network used in the study, where 27 black cells are used for model calibration and 7 red cells are for model verification.



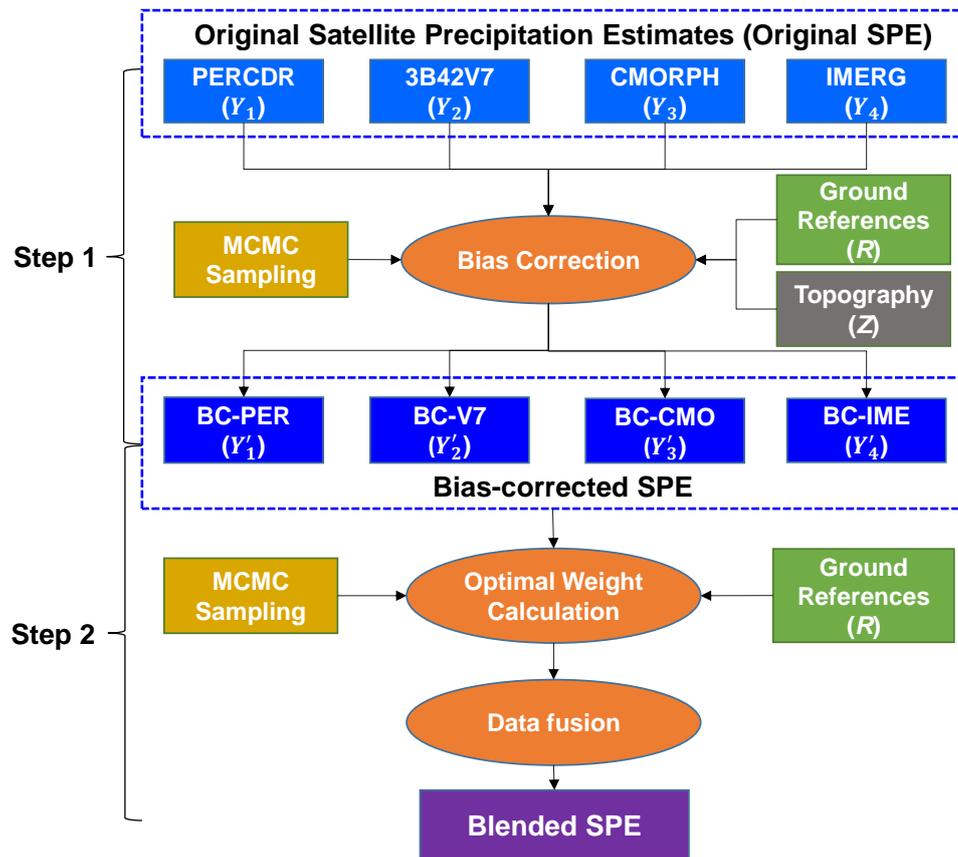
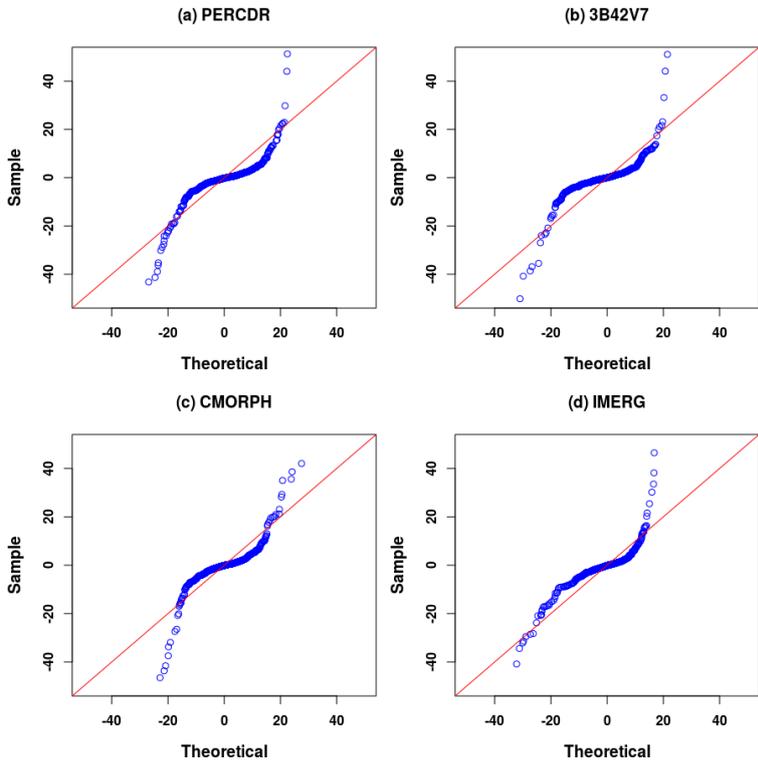
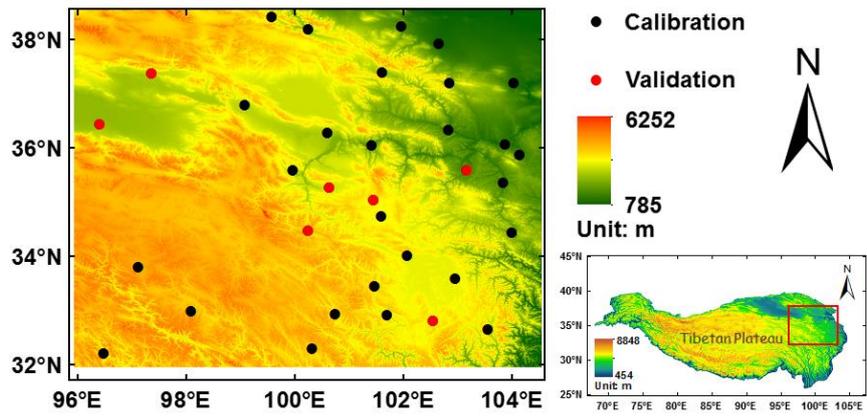


Figure 2: The diagram of the proposed TSB algorithm.



Formatted: Centered

Figure 3: Quantile-quantile plots at the training sets for the bias between GR and SPE, where (a) to (d) shows PERCDR, 3B42V7, CMORPH, and IMERG, respectively.



715 **Figure 1:** Overview of the topography and gauge observation network used in the study, where 27 gauges (black dots) are used for training and 7 (red dots) are used for independent verification.

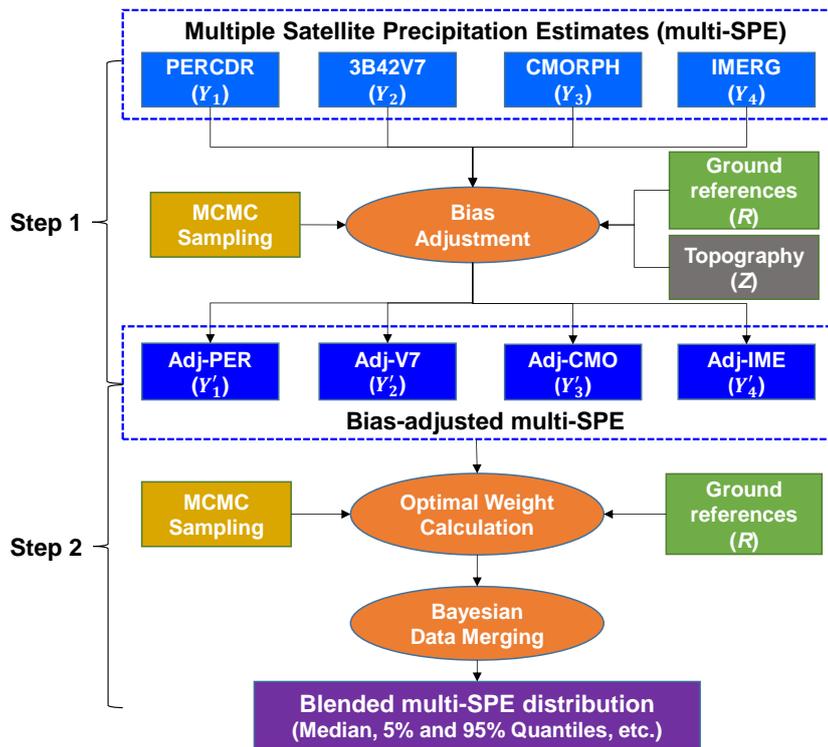


Figure 2: The diagram of the proposed two-step SPE-blending algorithm.

Formatted: Justified

Formatted: Left, Line spacing: single

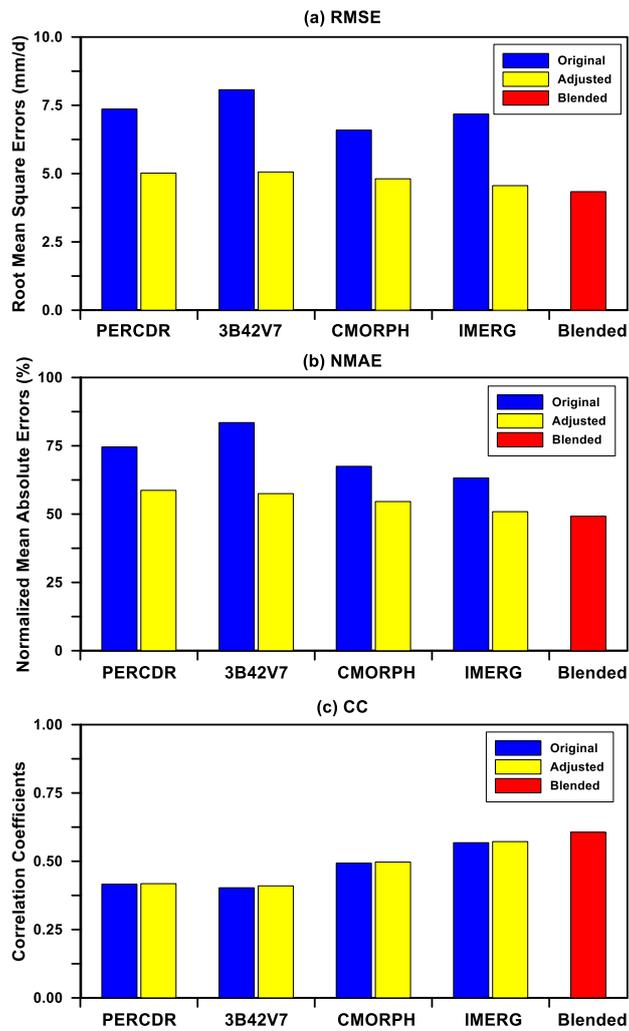
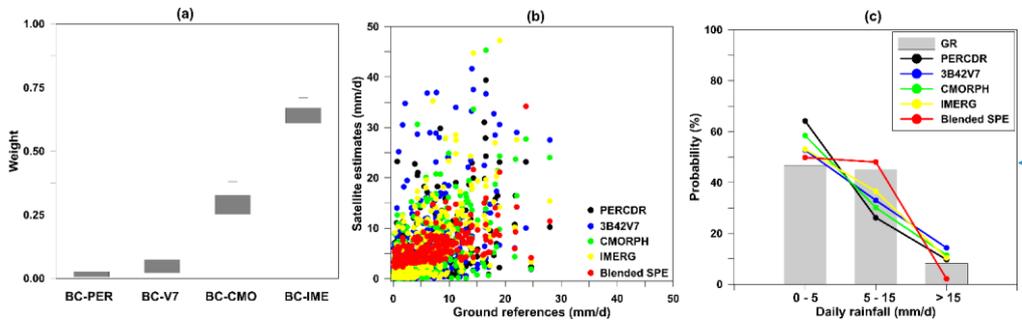


Figure 34: Intercomparison of statistical error indices for the original, bias-adjusted-corrected, and blended multi-SPE at the validation grids during-in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

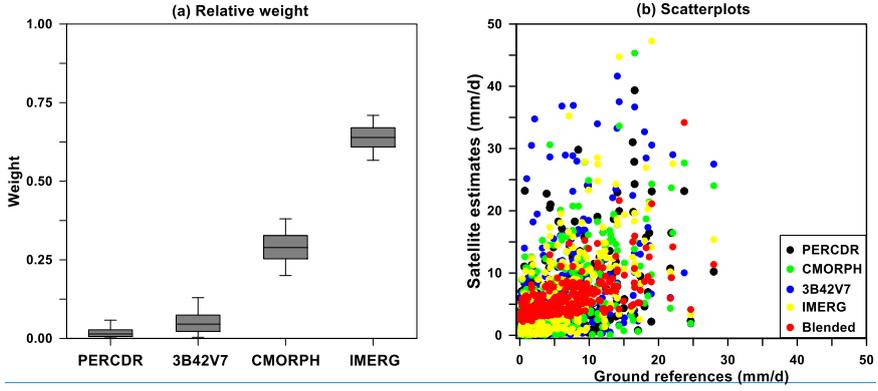
Formatted: Line spacing: single

Formatted: Left



Formatted: Centered

725



Formatted: Line spacing: 1.5 lines

Figure 45: (a) The Box-Whisker plots of relative weights of the bias-adjusted multi-SPE (i.e., PERCDR, 3B42V7, CMORPH and IMERG) in the Stage 2 process; (b) intercomparison Scatter plots between GR of the various original and blended multi-SPE (original and blended) at the validation grids during in the warm season of 2014; (c) The PDF of daily rainfall in terms of the GR, original and blended SPE with various intensities at the validation grids in the warm season of 2014.

Formatted: Justified, Line spacing: 1.5 lines

730

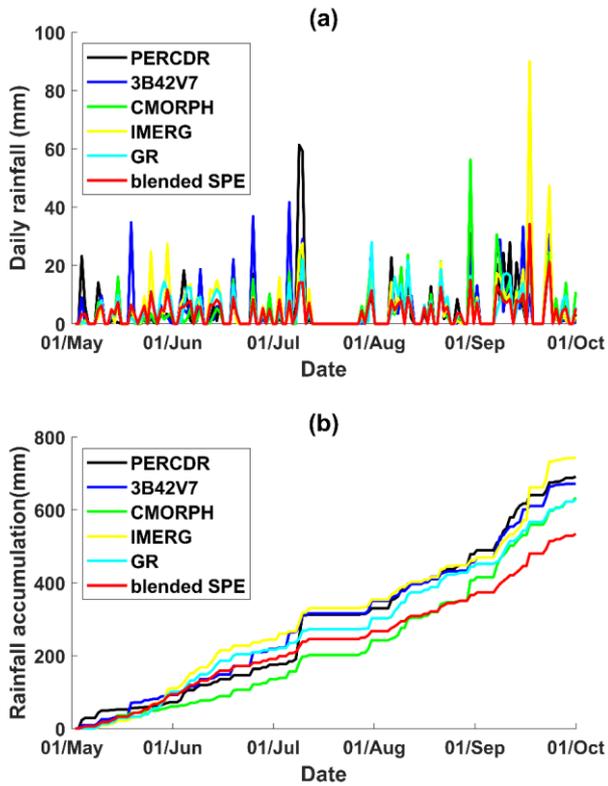


Figure 6: Time series of daily rainfall estimates and rainfall accumulations at a selected validation grid with the maximum rainfall record in the warm season of 2014: (a) daily rainfall estimates, and (b) rainfall accumulations.

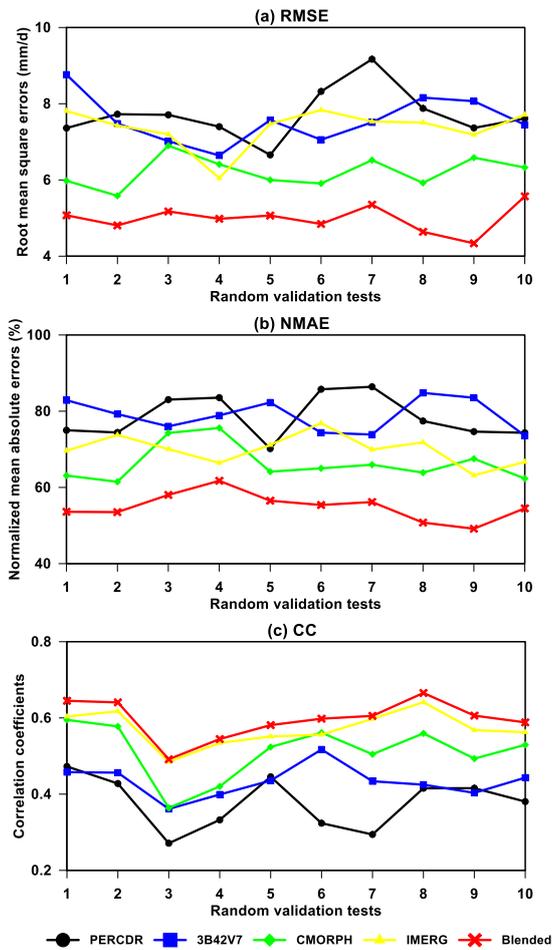


Figure 57: Statistical error indices of the original and blended multi-SPE (i.e., PERCDR, 3B42V7, CMORPH, and IMERG) for 10 random tests during the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

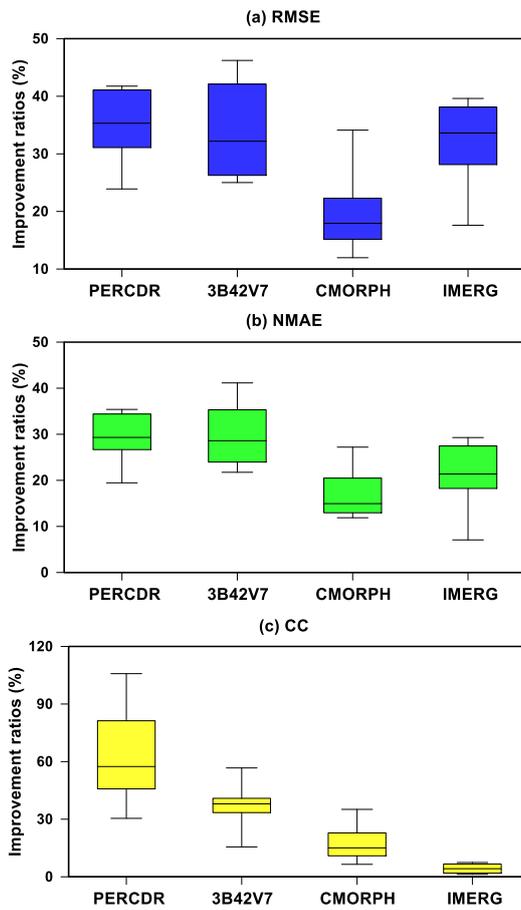


Figure 68: The Box-Whisker plots of improvement ratios of statistics for the blended multi-SPE compared to with the original SPE, including PERCDR, 3B42V7, CMORPH, and IMERG for 10 random tests during in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

Formatted: Justified, Line spacing: 1.5 lines

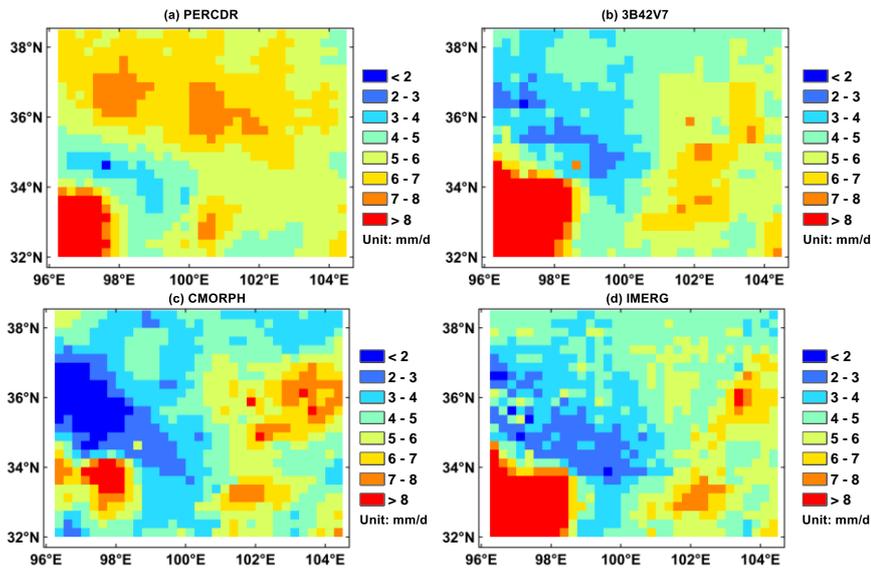
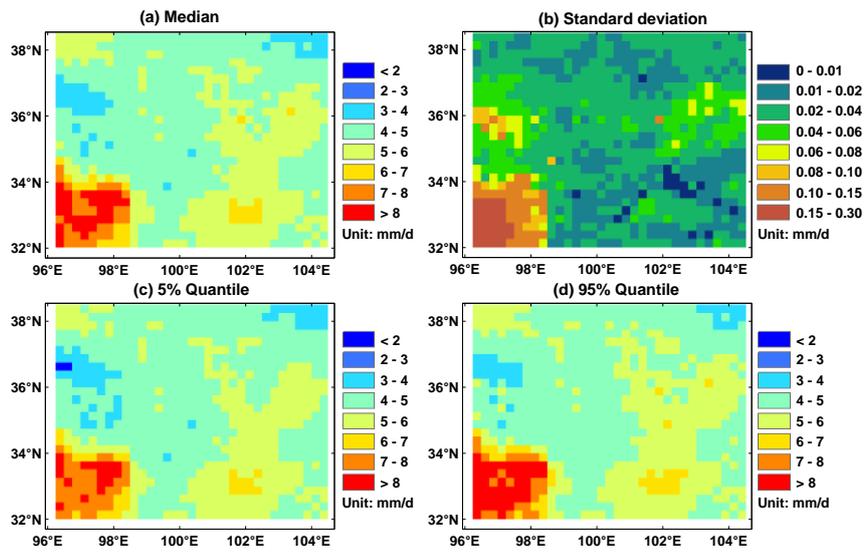


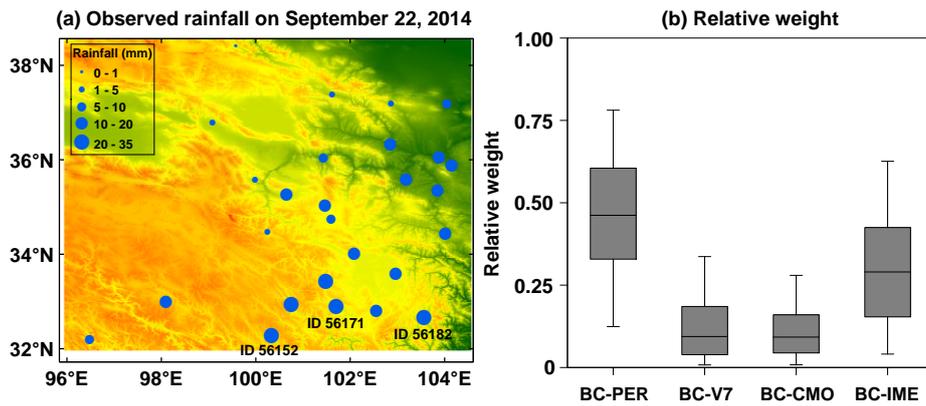
Figure 79: Spatial patterns of the daily mean precipitation [derived from in terms of](#) the original [multi-SPE during in](#) the warm season of 2014: (a) PERCDR, (b) 3B42V7, (c) CMORPH, and (d) IMERG.

Formatted: Justified, Line spacing: 1.5 lines

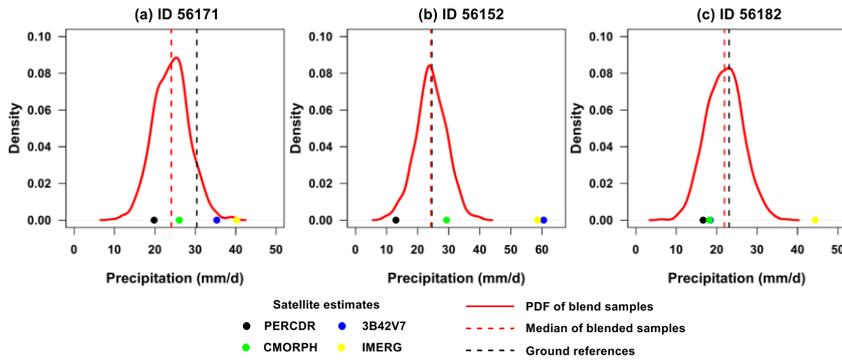
745



750 **Figure 108:** Spatial patterns of the blended [multi-SPE](#) in terms of (a) median, (b) [standard deviationSD](#), (c) 5% and (d) 95% quantiles of daily mean precipitation [duringin](#) the warm season of 2014.



755 **Figure 911:** (a) Spatial pattern of gauge-based measurements during a heavy rainfall case over the NETP on SepSeptember 22, 2014-over the NETP, where the site IDs 56171, 56152 and 56182 report the top three daily rainfall amounts of 30.4 mm, 24.6 mm and 23.1 mm, respectively; (b) the corresponding Box-Whisker plots of relative weights of the individual SPE of the bias-adjusted multi-SPE (i.e., PERCDR, 3B42V7, CMORPH and IMERG) in the stage-Stage 2 process.



760 **Figure 4012:** The PDF curves of blended samples-SPE and the corresponding median value at three gauge sites during a heavy rainfall case on SepSeptember 22, 2014: (a) ID 56171, (b) ID 56152, and (c) ID 56182. The individual-original SPE including PERCDR, 3B42V7, CMORPH, and IMERG as well as and gauge based measurementGR at each pixel are also indicated in the each subfigure.

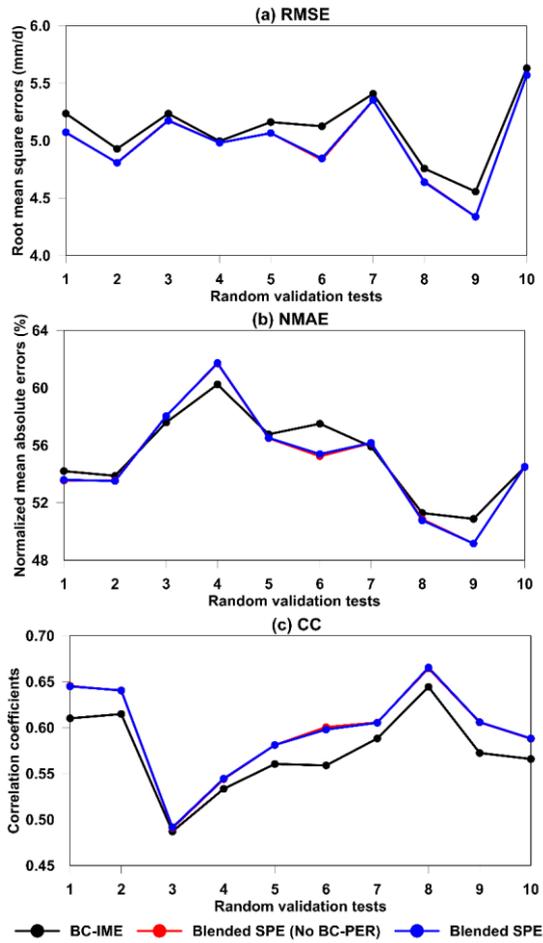
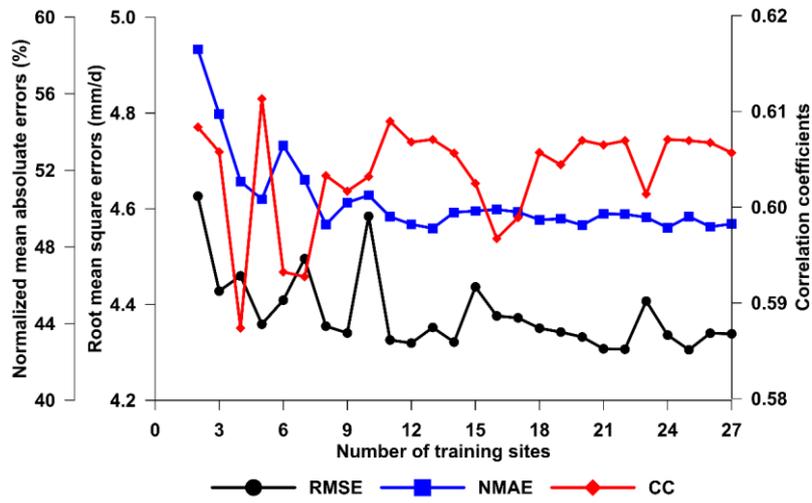


Figure 13: Statistical error indices (i.e., RMSE, NMAE, and CC) of the best-performed bias-corrected SPE (i.e., BC-IME, black) and blended SPE before (red) and after (blue) removing the worst-performed BC-PER for 10 random verified tests in the warm season of 2014 in the NETP.



770

[Figure 14: Statistical error indices \(i.e., RMSE, NMAE, and CC\) of the blended SPE at the validation grid locations in terms of different number of training sites in the warm season of 2014 in the NETP.](#)