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

Yingzhao Ma[1], Xun Sun[2,3], Haonan Chen[1,4], Yang Hong[5], Yinsheng Zhang[6,7]

[1]Colorado State University, Fort Collins, CO 80523, USA
[2]Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China
[3]Columbia Water Center, Earth Institute, Columbia University, New York, NY 10027, USA
[4]NOAA/Physical Sciences Laboratory, Boulder, CO 80305, USA
[5]School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK 73019, USA
[6]Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, 100101, China
[7]CAS 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 precipitation 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 gauge observations would be beneficial to improve the gridded precipitation estimates. In this study, a two-stage blending (TSB) approach is proposed, which firstly reduces the systematic errors of each SPE based on a Bayesian correction model, and then merges the bias-corrected SPE with a Bayesian weight model. In the first stage, the gauge-based observations are assumed as a generalized regression function of SPE and terrain feature. In the second stage, the SPE weights are calculated based on the associated performances relative to ground references. The proposed TSB method has the ability to exert benefits from 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 model parameters, which enables to produce the posterior ensembles associated with their predictive uncertainties. The performance of the TSB method is evaluated with independent validation grids in the warm season of 2014 in the northeastern Tibetan Plateau. Results show that the blended SPE is significantly improved compared to the original SPE, especially 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 any regions of interest.

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 provide precipitation estimates at a large scale (Hou et al., 2014),

but performances of available satellite products vary among different retrieval methods and climate 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 full consideration of the strength of individual members and associated uncertainty.

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° 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° x 2.5° 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).

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 estimation is proposed based on these 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 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). Given the flexible distribution of multiple sources of precipitation biases in regions with complex terrain (Derin et al., 2019), continuous efforts 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 combining satellite precipitation estimates (SPE) and point-based rain gauge observations. The experiment is performed 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 approach is expected to help with the exploration of multi-source/scale precipitation data fusion in other regions with complex terrain.

The remainder of this paper is organized below: Section 2 describes the experiment including the study region and precipitation data sets. Section 3 details the proposed TSB approach. Results and discussions are presented in Sections 4 and 5, respectively. The primary findings are summarized in Section 6.

2 Study area and dataset

The study domain is located in the upper Yellow River basin of NETP (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 for demonstration purpose.

Four mainstream SPE are used, including Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks - 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 research product version 1.0 (CMORPH) (Joyce et al., 2004), and the Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM) mission V03 Level 3 final run product (IMERG) (Huffman et al., 2018). The basic information of SPE is shown in Table 1. As the IMERG has a 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 is resampled from 0.10° to 0.25° using the nearest neighbour interpolation method.

A ground network including 34 rain gauges are used in this study. 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 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 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 GR are randomly classified into two parts: the black grids are used for training the model, and the red ones are used for model verification (Fig. 1). In order to clarify the TSB method, the selection of training cells is randomly repeated 10 times for the GR, and the remaining ones are used for model validation.

Meanwhile, the TSB method is applied on a heavy rainfall event that occurred on September 22, 2014 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 September 22 rain case is a typical storm that can explain the local heavy rainfall patterns in the warm season. The TSB approach is also compared with two existing fusion methods, i.e., BMA and One-outlier removed (OOR), which were previously applied for SPE data fusion in the TP (Ma et al., 2018; Shen et al., 2014).

3 The TSB algorithm

3.1 Overview

This algorithm aims at developing a multi-source data merging framework to provide the best-available gridded precipitation product with GR and SPE in the region of interest. Let $R(s,t)$ denote near-surface precipitation at the GR cell s and the t^{th} day. The original SPE and bias-corrected SPE 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. For simplicity, they are respectively replaced by R, $(Y_1, Y_2, \dots, Y_p)$, and $(Y_1', Y_2', \dots, Y_p')$. The subscript p implies the number of SPE, and PERCDR, 3B42V7, CMORPH and IMERG refer to Y_1, Y_2, Y_3, Y_4, respectively.

The diagram of the TSB method is shown in Figure 2. Stage 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 added as a covariate in the BC process. In the second stage, a Bayesian weight (BW) model is used to merge the bias-corrected SPE. The BW model can exert benefits from bias-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 3.2 and 3.3, respectively.

3.2 Stage 1: Bias correction

In this stage, we perform on conditional modelling of GR on each SPE, i.e., on the probabilistic distribution $f(R)$ at the training sets to improve the accuracy of the original SPE. 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. 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 sets (Fig. 3). It is found that 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 parameterized as follows:

$$R \sim Student(\nu_i, \mu_i, \sigma_i) \tag{1}$$

$$\mu_i = \alpha_i + \beta_i * Y_i + \gamma_i * Z \tag{2}$$

where ν_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 (Z). To ignore the scale factor, the elevation feature in Eq. (2) is normalized and its value ranges from 0 to 1 after the normalization. $\boldsymbol{\theta} = \{\nu_i, \alpha_i, \beta_i, \gamma_i, \sigma_i\}$ is summarized as a parameter set, which enables to write the likelihood function or probability density function (PDF) from Eqs. (1) and (2) conditional on $\boldsymbol{\theta}$ and Y_i as:

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

According to the Bayes's theorem, the posterior distribution of parameter set $\boldsymbol{\theta}$ given GR and SPE data, and the prior distribution of parameters $f(\boldsymbol{\theta})$ can be expressed as:

$$f(\boldsymbol{\theta}|\mathrm{R}, Y_i) \propto f(R|\boldsymbol{\theta}, Y_i) f(\boldsymbol{\theta}) \tag{4}$$

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

$$\alpha_i, \beta_i, \gamma_i \sim Uniform(-\infty, +\infty) \tag{5}$$

$$\nu_i, \sigma_i \sim Uniform(0, +\infty) \tag{6}$$

Based on the estimated parameter set $\boldsymbol{\theta}$, the next step is to calculate the bias-corrected SPE R^* at any new site. It can be quantitatively simulated from its posterior distribution in Eq. (7) using the original SPE Y_i^*, and training data R, Y_i:

$$f(R^*|Y_i^*, R, Y_i) = \int f(R^*, \boldsymbol{\theta}|Y_i^*, R, Y_i)\, d\boldsymbol{\theta} \tag{7}$$

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

$$f(R^*, \boldsymbol{\theta}|Y_i^*, R, Y_i) = f(R^*|Y_i^*, R, Y_i, \boldsymbol{\theta}) f(\boldsymbol{\theta}|Y_i^*, R, Y_i) \tag{8}$$

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

$$f(R^*|Y_i^*, R, Y_i, \boldsymbol{\theta}) = f(R^*|Y_i^*, \boldsymbol{\theta}) \tag{9}$$

Since the parameters $\boldsymbol{\theta}$ are dependent upon the training data R, Y_i, the second term of the right side in Eq. (8) is expressed as:

$$f(\boldsymbol{\theta}|Y_i^*, R, Y_i) = f(\boldsymbol{\theta}|R, Y_i) \tag{10}$$

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

$$f(R^*|Y_i^*, R, Y_i) = \int f(R^*|Y_i^*, \boldsymbol{\theta}) f(\boldsymbol{\theta}|R, Y_i)\, d\boldsymbol{\theta} \tag{11}$$

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 (i = 1, …, n_{sim}) as follows:

1) Calculate the model parameters $\boldsymbol{\theta}$ from Eqs. (1) to (6);

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

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

3.3 Stage 2: Data merging

In Stage 1, the median value of the posterior samples is used as the bias-corrected SPE. Here, we redefine the bias-corrected SPE as $Y_i'(i = 1,2, \dots, p)$. The formulas of blending the bias-corrected SPE are shown below:

$$B = \sum_{i=1}^{p} Y_i' * w_i + \varepsilon \tag{12}$$

$$\sum_{i=1}^{p} w_i = 1 \tag{13}$$

$$\varepsilon \sim Normal(0, \sigma) \tag{14}$$

$$w_i \sim Uniform(0,1), i = 1, \dots, p \tag{15}$$

$$\sigma \sim Uniform(0, +\infty) \tag{16}$$

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

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

The calculation process of $\boldsymbol{\delta}$ is similar to the parameter estimation described in Stage 1. After the parameters $\boldsymbol{\delta}$ 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 in terms of the median, standard deviation (SD) and associated credible intervals (e.g., 5% and 95% quantiles) in the regions of interest.

4 Results

To assess the performance of the proposed 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 SPE at the validation grids

Compared to the GR, the original SPE show large biases at the validation grids 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 0.403-0.568, respectively. 3B42V7 has the worst skill with the highest RMSE of 8.07 mm/d and the highest NMAE of 83.5%, and the lowest CC of 0.403. IMERG shows the best performance in terms of the lowest NMAE of 63.2% and highest CC at 0.568, which presents its superiority compared with the other SPE in the survey area.

Based on the BC model, the bias-corrected SPE (i.e., BC-PER, BC-V7, BC-CMO and BC-IME) have better agreements with GR at the validation grids in the experiment. Their RMSE scores range from 4.56 to 5.06 mm/d, and 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, compared with the original SPE. Meanwhile, their CC values range from 0.410 to 0.568 after bias adjustment (Table 2). 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). The RMSE and NMAE values of the blended SPE decrease by 34.1~65.4% and 27.1~41.1%, respectively, and the CC value increases by 6.7~50.4%, accordingly, compared with the original SPE (Table 2). As compared with the bias-corrected SPE, the blended SPE increases by 5.1~14.2%, 3.3~16.2%, and 5.9~47.8% in terms of RMSE, NMAE and CC, respectively. It is found that the blended SPE 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, the blended SPE has been effectively dropped towards the GR at the validation grids (Fig. 5b), especially for the rain intensity values less than 15 mm/d (Fig. 5c). 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 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.

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 including PERCDR (10.71), 3B42V7 (9.76), CMORPH (8.0), and IMERG (10.49), respectively.

4.2 Model clarification with random validation grids

Figures 7 and 8 show the statistics of evaluation scores of RMSE, NMAE, and CC for the original SPE and blended estimates at the validation grids with 10 random split of the gauge locations. For each test, 7 grid sites are randomly selected from the 34 grid cells and used for model verification, and the remaining 27 grid sites are used for training the model.

As for the blended SPE, it performs similar scores at the validation grids among the 10 random tests, but better skills in terms of RMSE, NMAE, and CC at 4.34~5.57 mm/h, 49.2~61.7%, and 0.492~0.665, respectively, compared with the original SPE at each test (Fig. 7). Statistically, the mean values of RMSE, NMAE and CC for the blended SPE are 4.98 mm/h, 54.9% and 0.597, respectively (Table 3). The averaged improvement ratios of RMSE for the blended SPE are 35.1%, 33.7%, 19.6% and 32.1% compared to the PERCDR, 3B42V7, CMORPH and IMERG, respectively, and similar performance is seen from NMAE with the average improvement ratios of 29.8%, 30.1%, 17.0% and 21.3%, respectively (Table 4). 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.

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. Here, the TSB approach is applied in spatial to obtain the blended SPE in terms of daily mean precipitation in the warm season of 2014 over the whole domain.

There is an overestimation for most of the original SPE, and the bias of the blended SPE is reduced based on the TSB approach. 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.

4.4 Model performance during a heavy rainfall case

Accurate precipitation on extreme weather is very important for flood hazard mitigation. We investigate the utility of this TSB approach on a heavy rainfall case of September 22, 2014 over the NETP (Fig. 11a). 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, on this particular heavy rainfall event (Fig. 11b).

Table 5 reports the evaluation statistics reflecting the blended model performance during this heavy rainfall case, where the RMSE, NMAE and CC values of the original SPE range from 6.28~10.48 mm/d, 40.6~59.5%, and 0.686~0.820, respectively. As compared to the original SPE, 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 SPE decrease by 34.2~60.6% and 32.5~53.9%, respectively, and the CC index correspondingly increases by 3.4~23.9% on this heavy rainfall case compared to the original SPE.

The blended model performance is further explored at three gauge cells (i.e., IDs 56171, 56152, 56182) with the top three daily rainfall records on September 22, 2014 (Fig. 11a). Figure 12 shows the PDF curves of blended samples at the above three grid sites in this rainfall case. It demonstrates the blended performance on quantifying the predictive uncertainty on rainfall extremes at each grid. At ID 56171, the estimated rainfall derived from the original 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. 12a). Based on the TSB method, 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 SPE are 24.3/5.0 mm/d and 21.9/4.5 mm/d, respectively, and they are very close to the GR with the daily amounts of 24.6 mm and 23.1 mm, respectively (Figs. 12b and 12c). These analyses reveal that the proposed TSB method can not only quantify its predictive uncertainty, but also improve the daily rainfall amount even on rainfall extremes.

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 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.

5 Discussion

In spite of the superior performance of the TSB algorithm, some issues still need to be considered in the practical applications:

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 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 deep convective systems occurring near the mountainous area have an effect on the precipitation cloud (Houze, 2012), more attempts are required to improve the orographic 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 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 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 assumed to be applicable in the whole domain. Since we have a relatively dense GR network in the survey region, the current assumption is acceptable according to the performance of the blended SPE. 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 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 is noted that, if extended to the TP or global scale, the extension of model parameters and 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.

The 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 SPE and provide more reliable gridded precipitation in high-cold region with complex terrain. 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, is expected to improve the proposed TSB algorithm. Also, hourly or even instantaneous precipitation intensity is extremely vital for flood prediction, which should be specifically designed when extending this fusion framework in the next step.

6 Summary and prospects

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

(1) This TSB algorithm has two stages and involves the BC and BW models. It is found that this blended method is capable of involving a group of SPE with their biases following different PDF curves. Meanwhile, it provides a convenient way to quantify the fusion performance and the associated uncertainty.

(2) The experiment shows that the blended SPE has better skill scores compared to the original SPE at the validation locations. The 10 random tests also confirms the superiority of the TSB algorithm. The performance of this fusion method is further demonstrated using a heavy rainfall event. In addition, the TSB method outperforms another two fusion methods (i.e., BMA and OOR).

(3) The application proves that this algorithm can allocate the contribution of individual SPE on the blended 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 merging SPE in high-cold region with complex terrain. The evaluation analysis of this TSB method for long-term period and extended regions (e.g., TP) in terms of higher temporal resolution (e.g., hourly) will be performed in a future 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.: 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 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 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., Lavado-Casimiro, W., Manz, B., Moges, S., Nikolopoulos, E., Sahlu, 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, 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.

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 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 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 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 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 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, Geophy. 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 SPE used in this study.

Table 2: Summary of statistical error indices (i.e., RMSE, NMAE, and CC) of the original, bias-corrected, and blended SPE at the validation grids of NETP in the warm season of 2014.

Table 3: Summary of the mean values of RMSE, NMAE and CC for the original and blended SPE at 10 random verified tests in the warm season of 2014 over the NETP.

Table 4: Mean improvement ratios of statistical error indices of the blended SPE, in terms of RMSE, NMAE and CC compared with the original SPE at 10 random verified tests in the warm season of 2014 over the NETP.

Table 5: Summary of statistical error indices (i.e., RMSE, NMAE, and CC) for the original and blended SPE during a heavy rainfall event over the NETP on 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.

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 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.

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.

Figure 4: Intercomparisons of statistical error indices for the original, bias-corrected, and blended SPE at the validation grids in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

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.

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 7: Statistical error indices of the original and blended SPE for 10 random tests in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

Figure 8: The Box-Whisker plots of improvement ratios of statistics for the blended SPE compared with the original SPE, including PERCDR, 3B42V7, CMORPH, and IMERG for 10 random tests in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

Figure 9: Spatial patterns of the daily mean precipitation in terms of the original SPE in the warm season of 2014: (a) PERCDR, (b) 3B42V7, (c) CMORPH, and (d) IMERG.

Figure 10: Spatial patterns of the blended SPE in terms of (a) median, (b) SD, (c) 5% and (d) 95% quantiles of daily mean precipitation in the warm season of 2014.

Figure 11: (a) Spatial pattern of gauge-based measurements during a heavy rainfall case over the NETP on September 22, 2014, 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 in the Stage 2 process.

Figure 12: The PDF curves of blended SPE and the corresponding median value at three gauge sites during a heavy rainfall case on September 22, 2014: (a) ID 56171, (b) ID 56152, and (c) ID 56182. The original SPE and GR 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.

Table 1: Basic information of 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	*Ashouri et al.*, 2015
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	*Huffman et al.*, 2007
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	*Joyce et al.*, 2004
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	*Huffman et al., 2018*

Table 2: Summary of statistical error indices (i.e., RMSE, NMAE, and CC) of the original, bias-corrected, and blended SPE at the validation grids of NETP in the warm season of 2014.

Product	RMSE (mm/d)	NMAE (%)	CC
PERCDR	7.36	74.6	0.416
3B42V7	8.07	83.5	0.403
CMORPH	6.59	67.5	0.493
IMERG	7.18	63.2	0.568
BC-PER	5.02	58.7	0.418
BC-V7	5.06	57.5	0.410
BC-CMO	4.81	54.6	0.497
BC-IME	4.56	50.9	0.572
Blended SPE	4.34	49.2	0.606

Table 3: Summary of the mean values of RMSE, NMAE and CC for the original and blended SPE at 10 random verified tests in the warm season of 2014 in 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

Table 4: Mean improvement ratios of statistical error indices of the blended SPE, in terms of RMSE, NMAE, and CC compared with the original SPE at 10 random verified tests in the warm season of 2014 over the NETP.

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

490

Table 5: Summary of statistical error indices (i.e., RMSE, NMAE, and CC) for the original and blended SPE during a heavy rainfall event over the NETP on September 22, 2014.

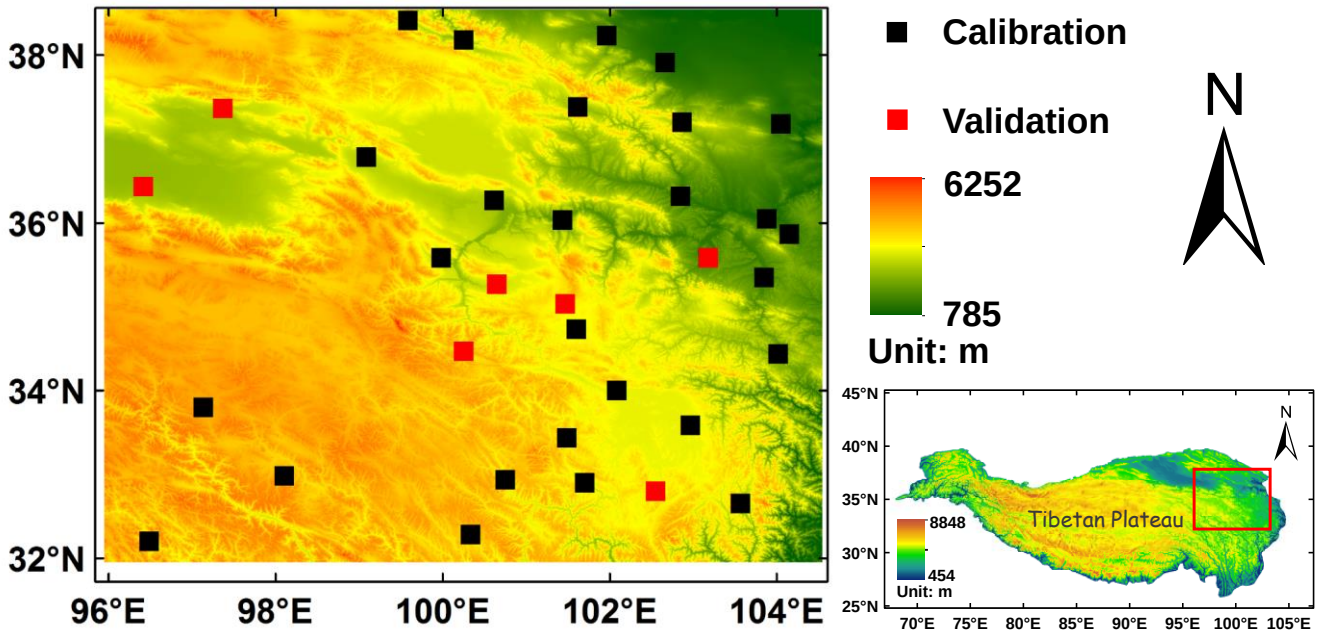
Product	RMSE (mm/d)	NMAE (%)	CC
PERCDR	6.28	40.6	0.822
3B42V7	10.12	59.5	0.686
CMORPH	6.80	45.6	0.734
IMERG	10.48	53.3	0.805
Blended SPE	4.13	27.4	0.850

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.

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

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



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

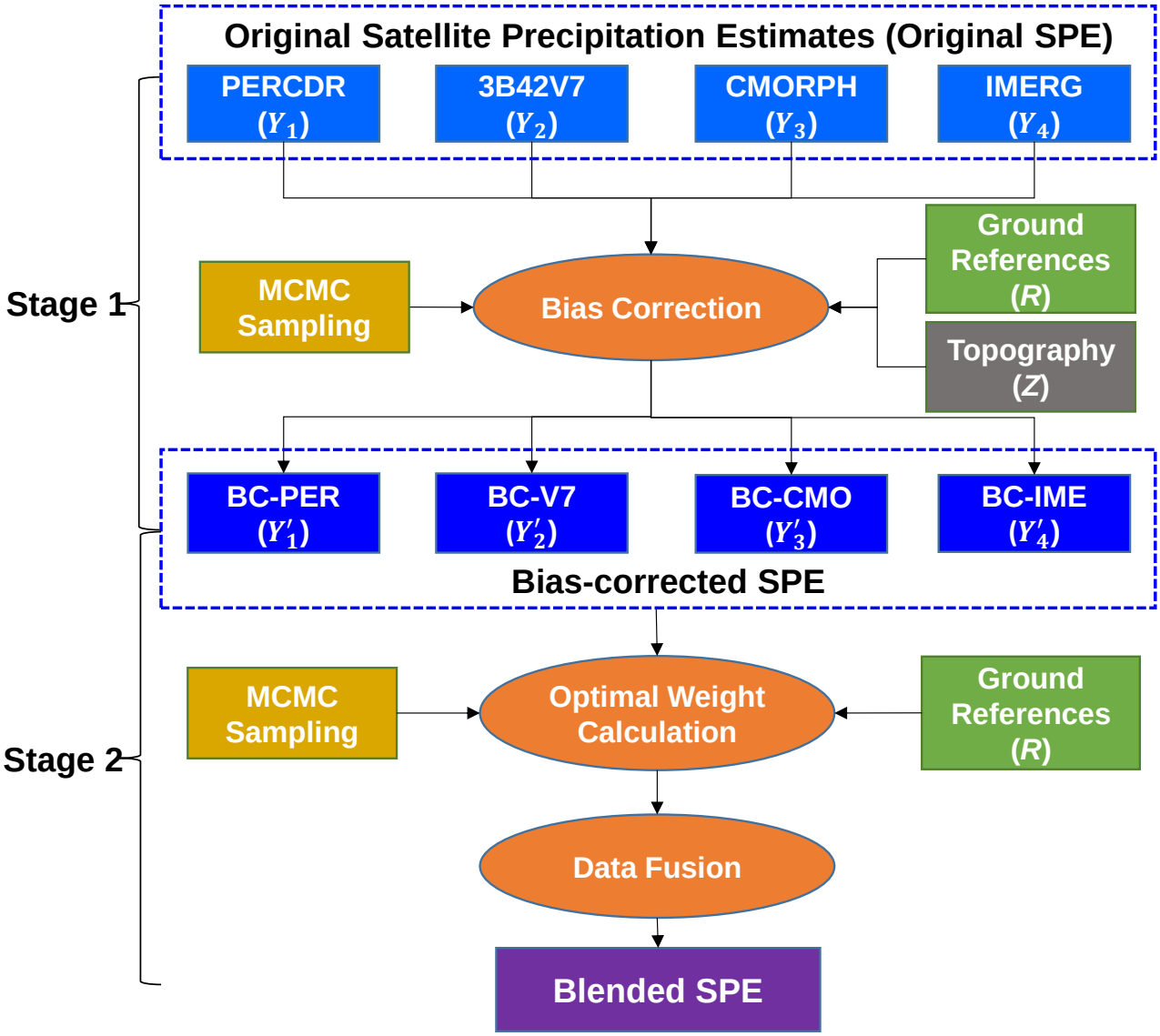


Figure 2: The diagram of the proposed TSB algorithm.

510

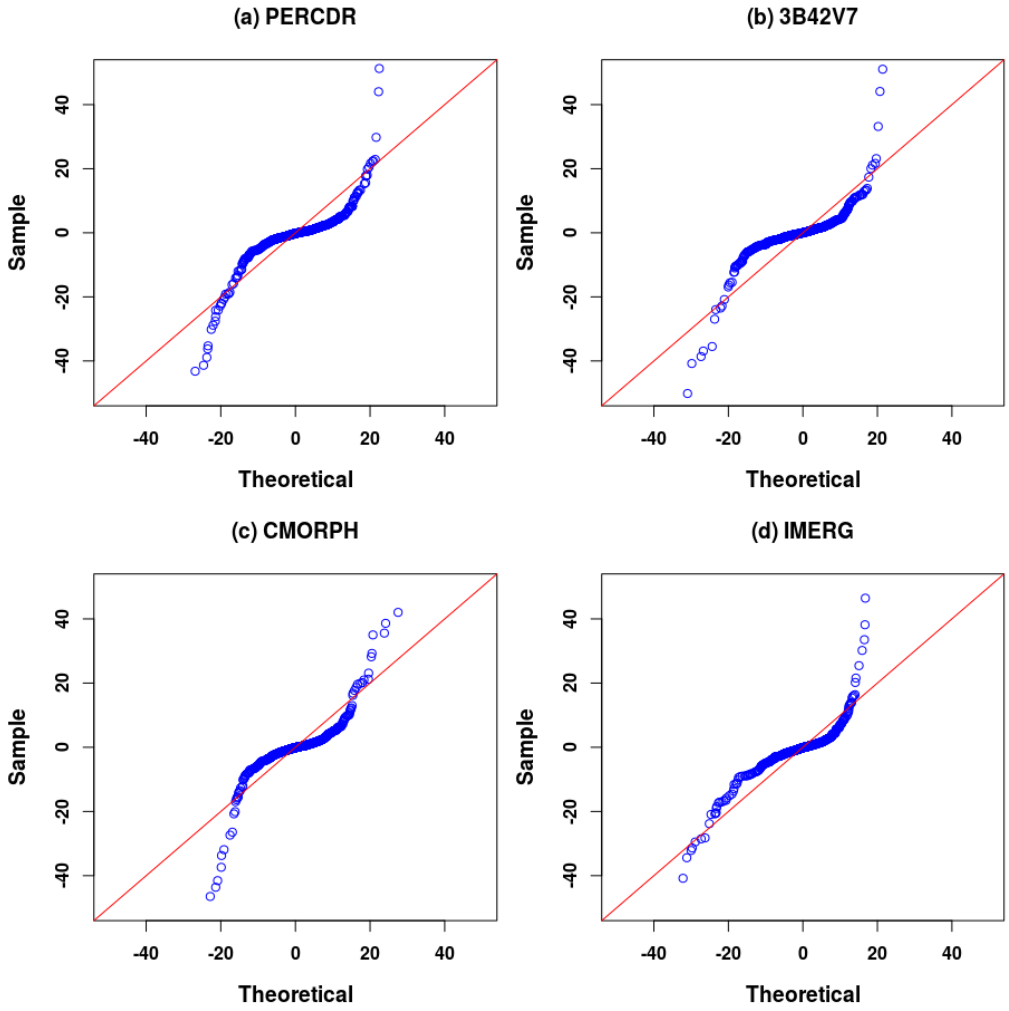
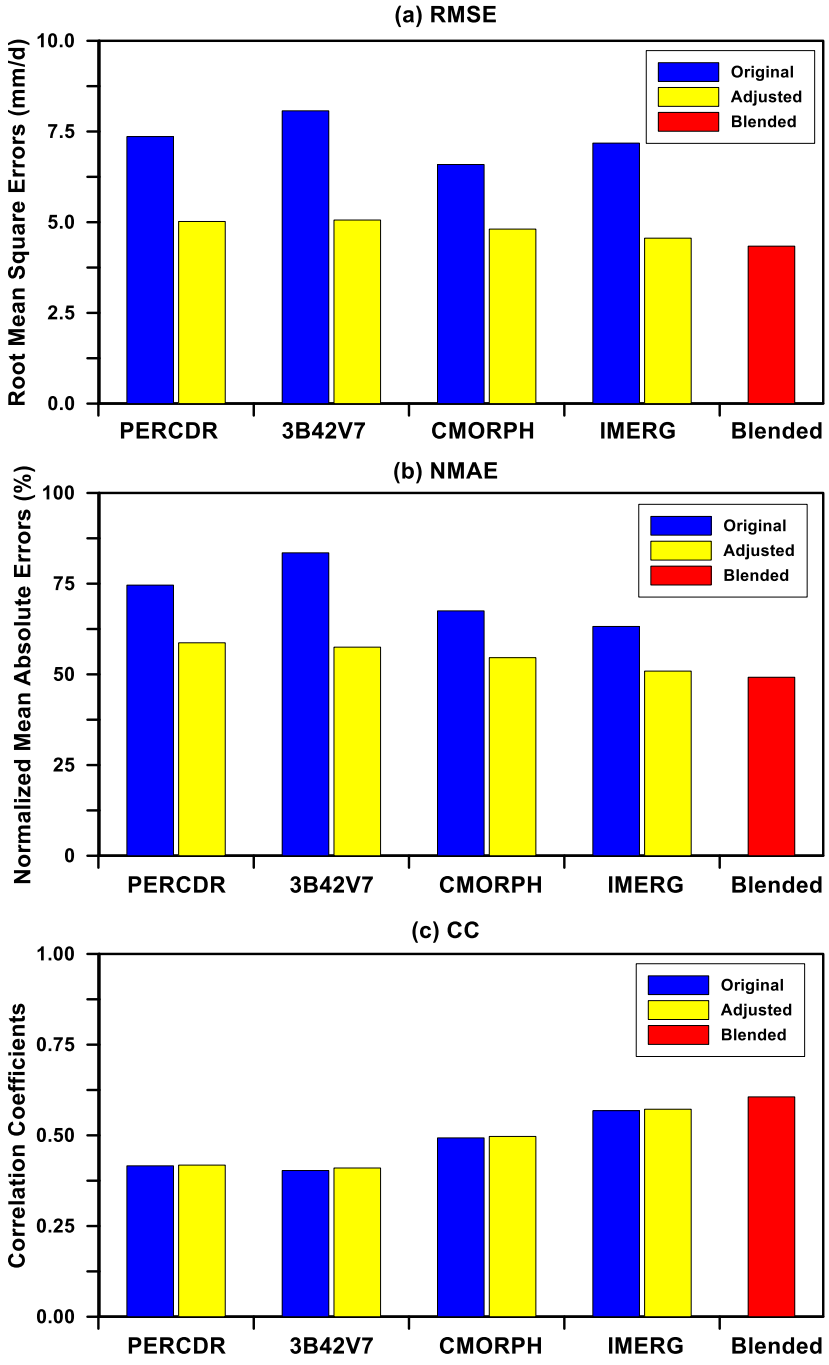


Figure 3: Quantile-quantile plots at the training sets for the bias between GR and SPE, where (a) to (d) shows PERCDR, 3B42V7, CMORPH, and IMERG, respectively.



515

Figure 4: Intercomparison of statistical error indices for the original, bias-corrected, and blended SPE at the validation grids in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

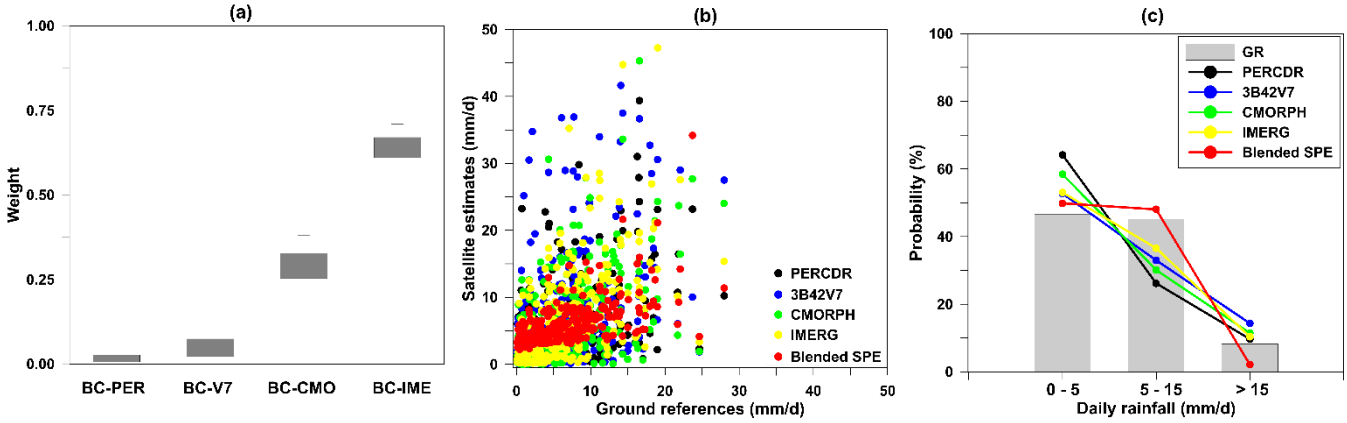


Figure 5: (a) The Box-Whisker plots of relative weights of the bias-corrected SPE in Stage 2; (b) Scatter plots between GR and various SPE (original and blended) at the validation grids in the warm season of 2014; (c) The PDF of daily rainfall in terms of the GR, original and blended SPE with various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

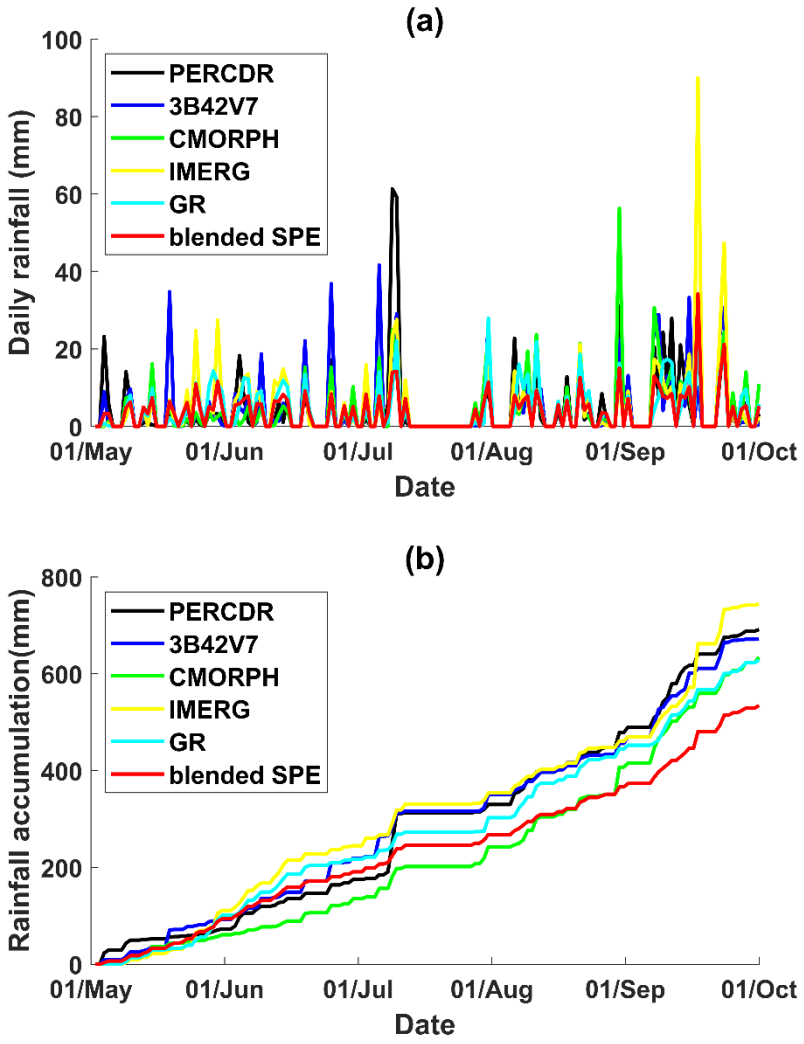


Figure 6: Time series of daily rainfall estimates and rainfall accumulations at a selected validation grid with the maximum rainfall record in the warm season of 2014: (a) daily rainfall estimates, and (b) rainfall accumulations.

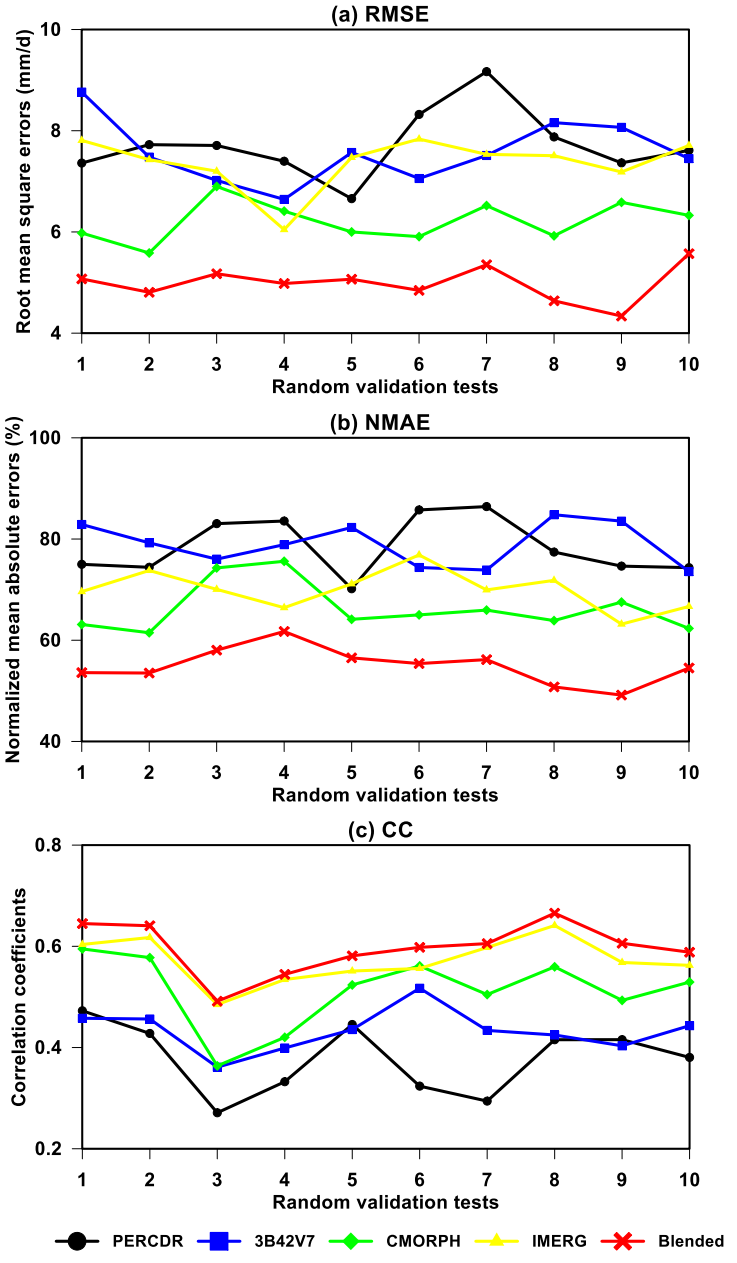


Figure 7: Statistical error indices of the original and blended SPE for 10 random tests in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

530

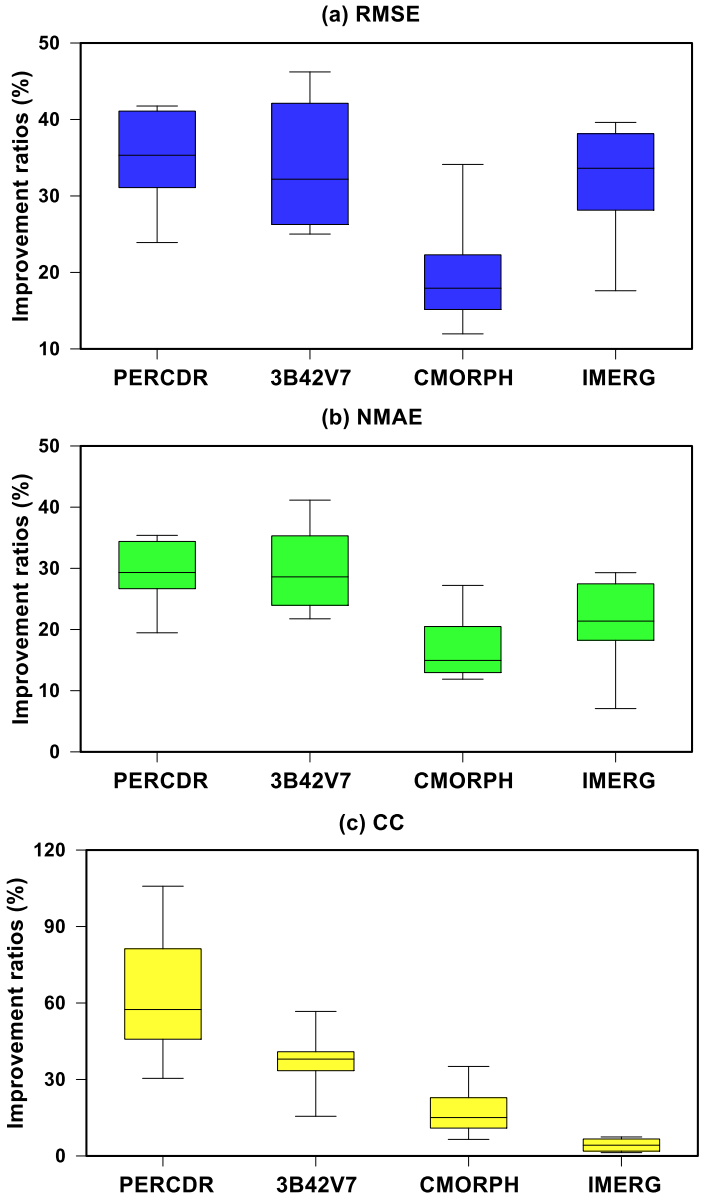


Figure 8: The Box-Whisker plots of improvement ratios of statistics for the blended SPE compared with the original SPE, including PERCDR, 3B42V7, CMORPH, and IMERG for 10 random tests in the warm season of 2014: (a) RMSE, (b) NMAE, and (c) CC.

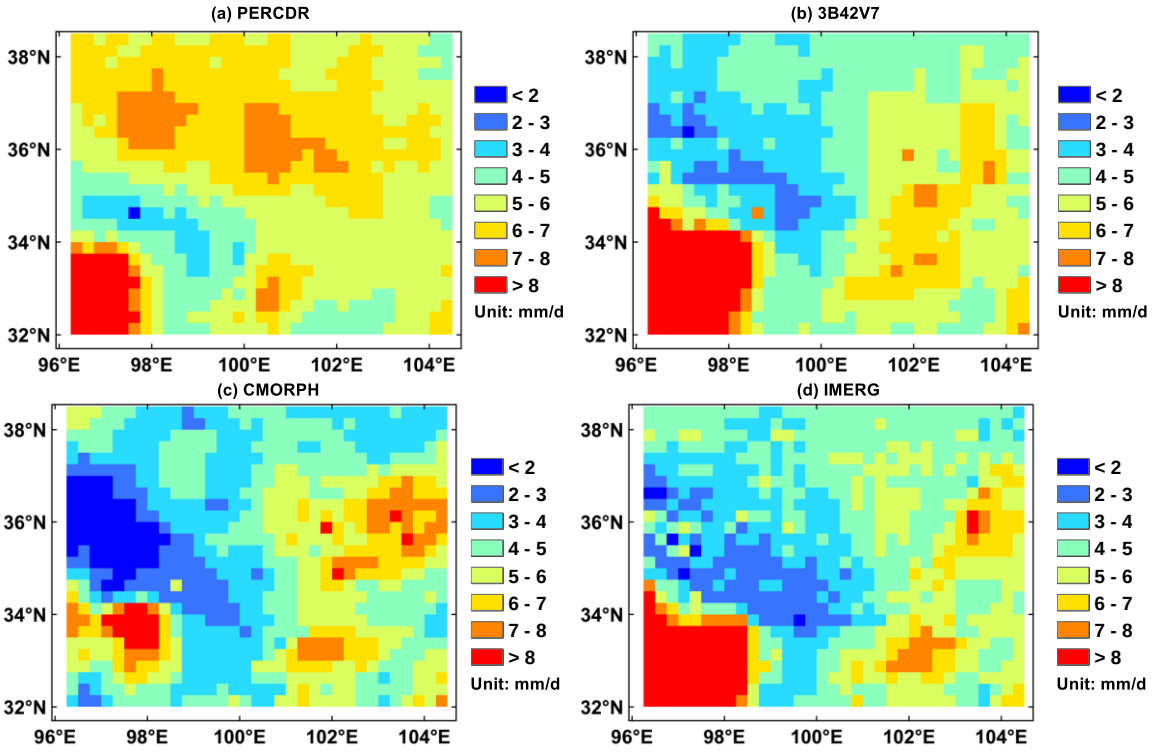


Figure 9: Spatial patterns of the daily mean precipitation in terms of the original SPE in the warm season of 2014: (a) PERCDR, (b) 3B42V7, (c) CMORPH, and (d) IMERG.

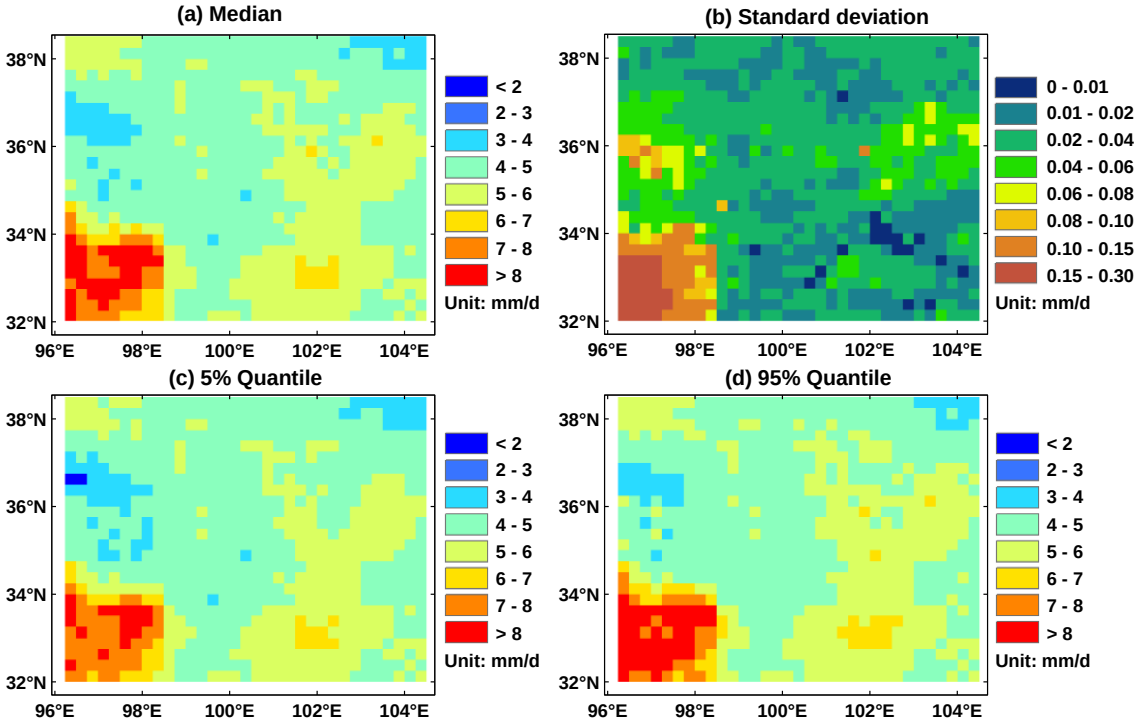


Figure 10: Spatial patterns of the blended SPE in terms of (a) median, (b) SD, (c) 5% and (d) 95% quantiles of daily mean precipitation in the warm season of 2014.

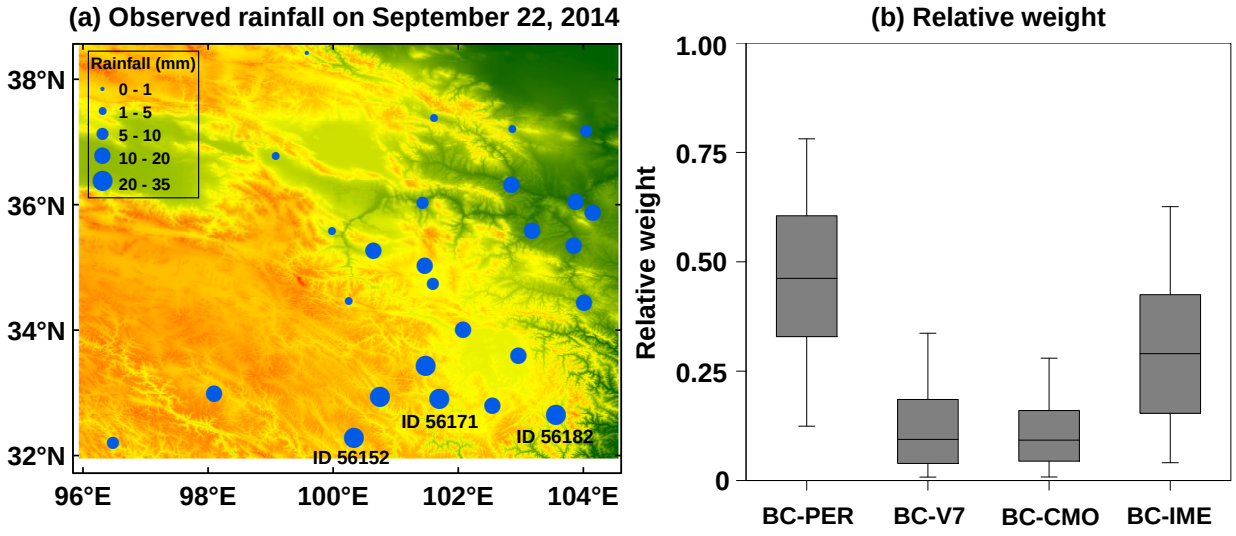


Figure 11: (a) Spatial pattern of gauge-based measurements during a heavy rainfall case over the NETP on September 22, 2014, where the site IDs 56171, 56152 and 56182 report the top three daily rainfall amounts of 30.4 mm, 24.6 mm and 23.1 mm, respectively; (b) the corresponding Box-Whisker plots of relative weights of the individual SPE in the Stage 2 process.

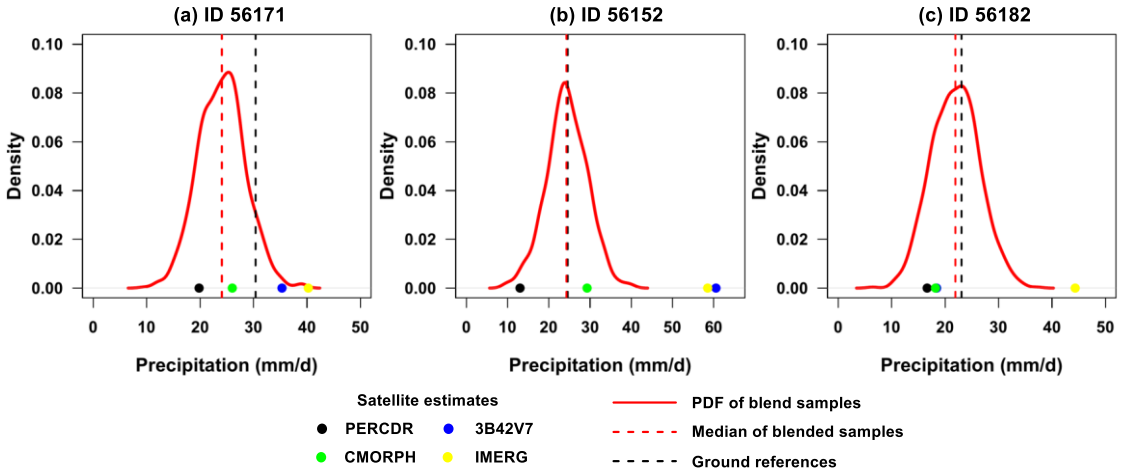


Figure 12: The PDF curves of blended SPE and the corresponding median value at three gauge sites during a heavy rainfall case on September 22, 2014: (a) ID 56171, (b) ID 56152, and (c) ID 56182. The original SPE and GR at each pixel are also indicated in each subfigure.

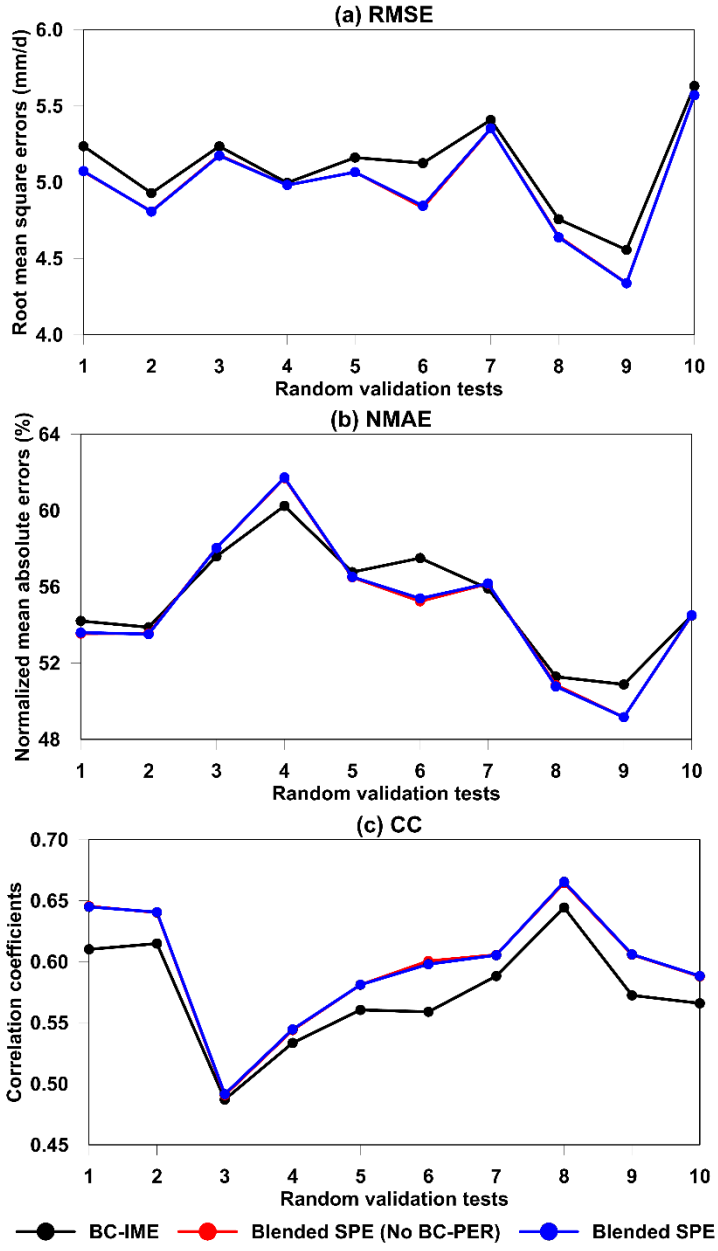


Figure 13: Statistical error indices (i.e., RMSE, NMAE, and CC) of the best-performed bias-corrected SPE (i.e., BC-IME, black) and blended SPE before (red) and after (blue) removing the worst-performed BC-PER for 10 random verified tests in the warm season of 2014 in the NETP.

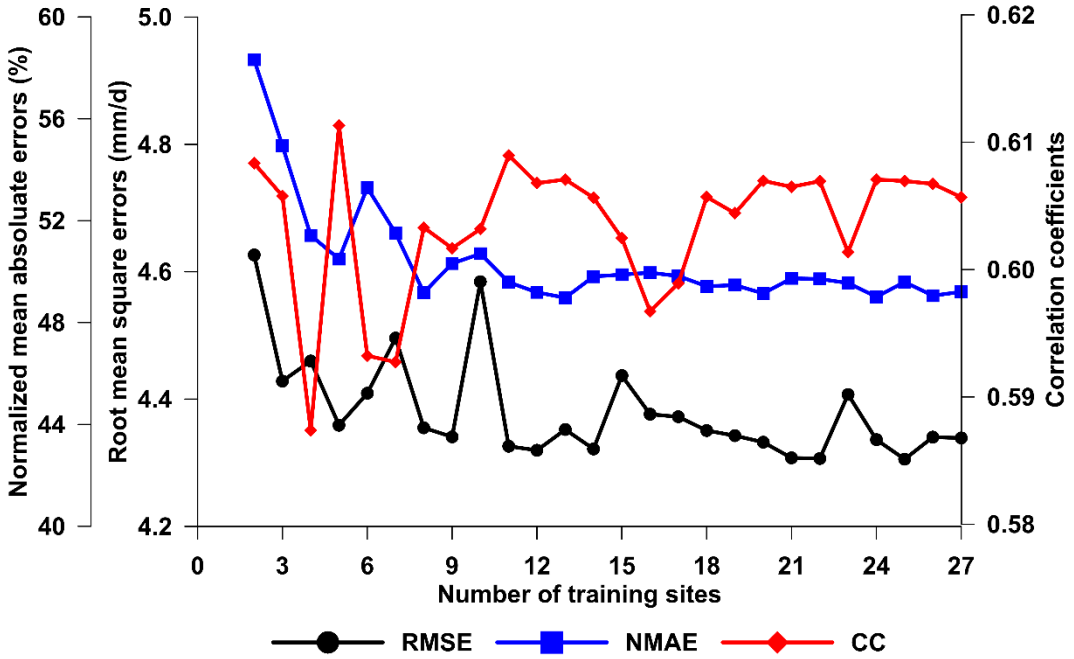


Figure 14: Statistical error indices (i.e., RMSE, NMAE, and CC) of the blended SPE at the validation grid locations in terms of different number of training sites in the warm season of 2014 in the NETP.