

# **Correcting bias of satellite rainfall data using physical empirical model**

Ghaith Falah Ziarh<sup>1</sup>, Shamsuddin Shahid<sup>1</sup>, Tarmizi bin Ismail<sup>1</sup>, Md Asaduzzaman<sup>2</sup> and Ashraf Dewan<sup>3</sup>

<sup>1</sup>Department of Water and Environmental Engineering, School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia. E-mail: [eng.ghaith.ziarh@gmail.com](mailto:eng.ghaith.ziarh@gmail.com); [sshahid@utm.my](mailto:sshahid@utm.my); [tarmiziismail@utm.my](mailto:tarmiziismail@utm.my);

<sup>2</sup>Department of Engineering, School of Digital, Technologies and Arts, Staffordshire University, Stoke-on-Trent ST4 2DE, UK. Email: [md.asaduzzaman@staffs.ac.uk](mailto:md.asaduzzaman@staffs.ac.uk)

<sup>3</sup>Spatial Sciences Discipline, School of Earth and Planetary Sciences, Curtin University, Perth, WA 6102, Australia. Email: [a.dewan@curtin.edu.au](mailto:a.dewan@curtin.edu.au)

# Correcting bias of satellite rainfall data using physical empirical model

## Abstract

The provision of high resolution near real-time rainfall data has made satellite rainfall products very potential for monitoring hydrological hazards. However, a major challenge in their direct-use can be problematic due to measurement error. In this study, an attempt was made to correct the bias of Global Satellite Mapping of Precipitation near-real-time (GSMaP\_NRT) product. Physical factors, including topography, season, windspeed and cloud types were accounted for correcting bias. Peninsular Malaysia was used as the case study area. Gridded rainfall, developed from 80 gauges for the period 2000-2018, was used along with physical factors in a two-stage procedure. The model consisted of a classifier to categorise rainfall of different intensity and regression models to predict intensity class of different rainfall amount. An ensemble tree-based learning algorithm, called random forest, was used for classification and regression. The results revealed a big improvement of near-real-time GSMaP\_NRT product after bias correction (GSMaP\_BC) compared to the gauge corrected version (GSMaP\_GC). Accuracy evaluation for complete time series indicated about 110% reduction of normalized root-mean-square error (NRMSE) in GSMaP\_BC (0.8) compared to GSMaP\_NRT (1.7) and GSMaP\_GC (1.75). On the other hand, the bias of GSMaP\_BC became nearly zero (0.3) compared to 2.1 and -3.1 for GSMaP\_NRT and GSMaP\_GC products. The spatial correlation of GSMaP\_BC was  $>0.7$  with observed rainfall data for all months compared to 0.2–0.78 for GSMaP\_NRT and GSMaP\_GC, indicating capability of GSMaP\_BC to replicate spatial pattern of rainfall. The bias-corrected near-real-time GSMaP data can be used for monitoring and forecasting floods and hydrological phenomena in the absence of dense rain-gauge network in areas, frequently experience hydro-meteorological hazards.

Keywords Near-real-time rainfall, satellite precipitation, bias correction, ensemble learning algorithm, physical-empirical model

## 1. Introduction

Precipitation is a major input for understanding hydrological processes, monitoring hydrological hazards, and evaluating different climatic studies. In-situ gauges records are generally used for reliable and accurate estimation of precipitation. However, space-time variation of several gauge records is often inadequate for hydrological study, particularly in areas having rainfall heterogeneity (Nashwan et al., 2019a; Shiru et al., 2018). Besides, installation of in-situ instruments in inaccessible locations such as mountainous regions and large waterbodies precludes reliable estimation of precipitation (Ahmed et al., 2017; Ji et al., 2020). Remote sensing technique for global precipitation observation helped to overcome such difficulty. Satellite-based precipitation products are not only capable of providing large synoptic coverage, they are cost-effective relative to in-situ networks (Nashwan et al, 2019b; Suliman et al., 2020).

Several satellite precipitation products have been developed over the last two decades and have become important data source for a wide range of applications. The products are mainly based on infrared or microwave sensors and observed precipitation data (Nashwan et al., 2019b). However, their performance vary significantly according to areas of interest. Global Satellite Mapping of Precipitation (GSMaP) (Kubota et al.2007) is one of such products which is shown to be more accurate compared to other satellite-based products (Gosset et al., 2018; Islam, 2018; Reddy et al., 2019; Shawky et al., 2020). Even though the performance of GSMaP is better compared to other remote sensing precipitation products for many regions, studies showed that this product is far less accurate compared with in-situ records (Islam, 2018; Tan et al., 2018; Aslami et al., 2019; Chen et al., 2020; Shawky et al., 2020). Chen et al. (2020), for instance, reported that GSMaP overestimates light rainfall globally, with an error of  $>2$  mm, on daily scale. Shawky et al. (2020) showed an overestimation of light rain and underestimation of heavy rainfall over arid environments. Aslami et al. (2019) revealed that GSMaP overestimates precipitation in the northwest of Iran. Mastrantonas et al. (2018) noted that GSMaP performs better for estimating low-intensity rainfall events but underestimation ( $>50\%$ ) was prevalent for high-intensity rainfall in Kinu basin of Japan. Tan et al. (2018) showed underestimation of rainfall amounts with GSMaP products over major river basins of China. Islam (2018) reported that GSMaP perform poorly in detecting extreme rainfall accumulations in Bangladesh. Ning et al. (2017) reported GSMaP tends to overestimate light rainfall ( $<16$  mm/day) and underestimate precipitation with rain rate

>16 mm/day in Eastern China. These studies emphasize the need for bias correction of GSMaP product prior to use in hydrological studies.

Several attempts have been made to correct the bias of satellite-based rainfall products (Chaudhary and Dhanya, 2019; Yeh et al., 2019; Gebremedhin et al., 2020). Approaches to correcting bias include nonparametric quantile mapping (Yang et al., 2016; Alharbi et al., 2018), cumulative distribution function matching (Mastrantonas et al., 2018), linear scaling (Hashemi et al., 2017; Chaudhary and Dhanya, 2019), intensity thresholds (Saber and Yilmaz, 2018), non-linear Power Transformation (Pratama et al., 2018), local correction (Yeh et al., 2019), censored shifted mixture distribution (Ma et al., 2019), ratio bias correction (Mastrantonas et al., 2018) and regression analysis (Yeh et al., 2019) and dynamic bayesian model (Ma et al., 2018). Use of these methods indicated that the performance of satellite-based precipitation products improves significantly. However, most of these studies reported unacceptable performance even after correcting bias, particularly in replicating extreme rainfall amounts. Mastrantonas et al. (2018) reported correction of bias only improves the performance for high temporal scale and no improvement can be found for <6-hour rainfall amount. Deng et al. (2018) showed that bias correction can reduce error of GSMaP estimates though bias removal can also lead to disappearing of some rain events. Findings of previous studies therefore indicated the necessity of a more efficient method for bias correction of satellite precipitation.

Biases with remotely sensed rainfall products vary with geography, topography, season, windspeed and cloud type (Zeng and Yong, 2019; Saber and Yilmaz, 2018; Chen et al., 2019). Yeh et al. (2019) mentioned that bias in GSMaP rainfall estimation depends on the area being investigated together with elevation, and season. Relatively high uncertainty is found to be associated with GSMaP precipitation at high altitude regions compared with flat topography (Zeng and Yong, 2019). Utsumi and Kim (2018) reported IR-based precipitation product considerably overestimates rainfall intensity, as a result, bias tends to be related to cloud types. Saber and Yilmaz (2018) revealed that bias in GSMaP depends on rainfall thresholds, i.e., higher rainfall amount presents larger bias. Trinh-Tuan et al. (2019) showed the dependency of GSMaP bias on elevation and zonal windspeed in central Vietnam. Therefore, consideration of different factors can improve bias of satellite rainfall products. However, only a few attempts have been made to correct for bias using physical factors. Velasquez et al. (2020) proposed a method by taking orographic factor into account. Yeh et al. (2020) corrected bias of GSMaP data using elevation. To the best of our knowledge, no study utilises multiple physical factors to account for bias of satellite rainfall products such as GSMaP.

In this study, a novel bias correction method is proposed. Hourly data from Global Satellite Mapping of Precipitation (GSMaP) v.7 is used and elevation, windspeed, season and rainfall types are considered. An ensemble algorithm, based on a large number of classification and regression trees (CARTs) called Random Forest (RF) (Breiman, 2001), is employed to correct raw GSMaP precipitation data by taking in-situ precipitation, elevation, distance to coast, month of the year, and threshold of precipitation into account. As different cloud types are accountable to produce amounts and intensities of precipitation, different intensity categories are considered as a proxy to cloud type. For example, stratus clouds produce light precipitation while nimbostratus clouds yield elevated precipitation. Likewise, distance from coast was used as a proxy to windspeed due to the fact that speed of wind usually decreases from the coast to inland (Kubota and Ahmad, 2006). Separate models were developed for different months in order to consider seasonality. Observed hourly precipitation, recorded at 80 locations over peninsular Malaysia, is used for model development and validation of the bias correction models.

## **2. Materials and methods**

### **2.1 Peninsular Malaysia**

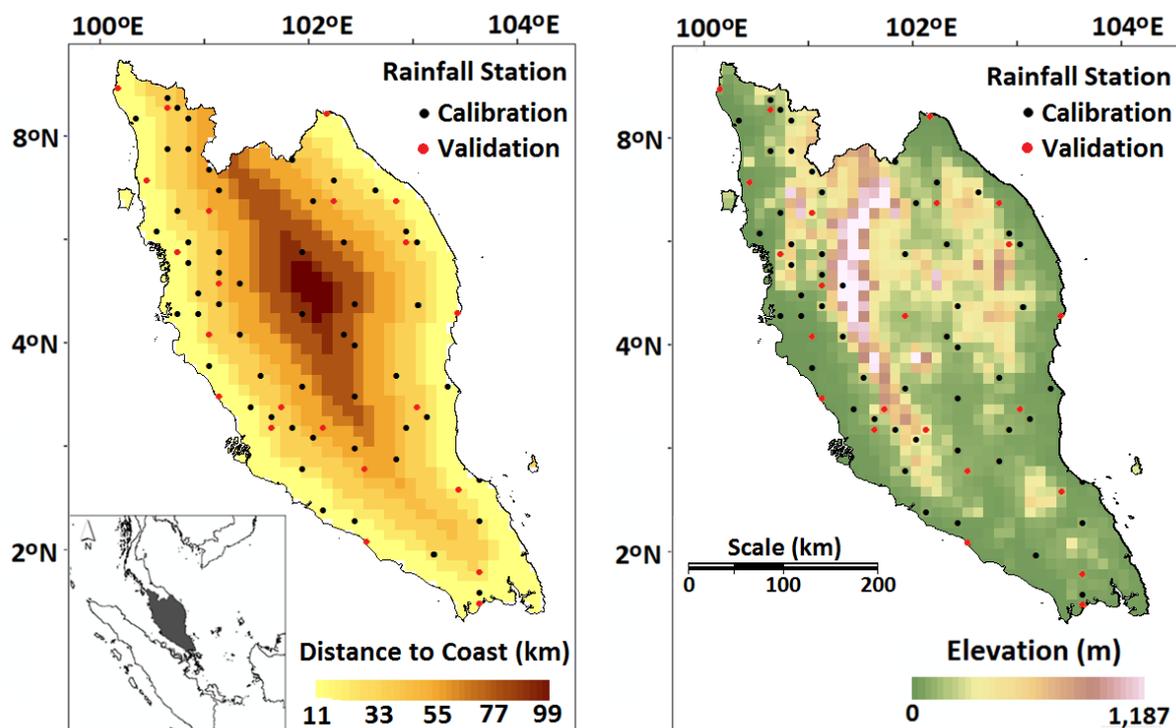
Peninsular Malaysia (latitude 1.20°–6.40°N; longitude 99.35°–104.20°E) encompasses an area of 130,598 km<sup>2</sup> (Figure 1). Topography of the country is composed of an irregular mountainous region in the middle and flat seashores around. It has a tropical climate with year-round high temperatures and humidity. Annual mean temperature in the peninsula varies between 21° C and 32° C (Pour et al., 2020). Rainfall climatology of the region is controlled by two monsoons and their interactions with heterogeneous land surface and coastlines. Most of the rainfall occurs during two monsoons, i.e, northeast monsoon (NEM) ranges from November to February, and southwest monsoon (SWM) occurs between May and August (Muhammad et al., 2019). Between the two monsoons, NEM is intense. Several extreme rainfall events often occur in consecutive days during NEM, resulting in floods, particularly to the west of peninsular Malaysia (Nashwan et al., 2018).

Being a tropical country, peninsular Malaysia is more subtle to alteration in climatic characteristics (Shahid et al., 2017). Recent works reported increasing frequency and severity of extreme events, causing societal and economic disruptions to peninsular Malaysia (Noor et al., 2018, Noor et al., 2020, Abdullah et al., 2019). For instance, an increase in monsoonal floods due to rainfall extremes is already evident in the region (Noor et al., 2018). The most

recent flood (December 2014) affected thousands of people and registered huge economic damage (Shahid et al., 2017). Thus, reliable estimation of high-resolution real-time precipitation is very important for peninsular Malaysia, which could allow efficient monitoring of floods and consequent reduction of damage to society and economy.

## 2.2 Data sources

Hourly observed rainfall records for the period 2000-2018 at 80 locations, distributed over peninsular Malaysia, were collected from Department of Irrigation and Drainage (DID), Malaysia. DID employs automatic tipping bucket rain gauge for measuring and recording of rainfall data. Figure 1 shows locations of rainfall stations in the study area.



**Figure 1** (a) Distance to coast; and (b) elevation of peninsular Malaysia. Point features represent rainfall gauges

GSMaP near-real-time (GSMaP\_NRT) and GSMaP gauge corrected (GSMaP\_GC) hourly precipitation data (v.7) were acquired from 2000-2018 (<https://sharaku.eorc.jaxa.jp/GSMaP/index.htm>). It has a spatial resolution of  $0.1^\circ \times 0.1^\circ$  (latitude  $\times$  longitude) and developed by Japan Science and Technology Agency (Okamoto et al., 2005; Ushio et al., 2009; Yamamoto et al., 2017). GSMaP\_NRT is developed by integrating

precipitation rates, extracted from passive microwave radiometers and cloud images (Nashwan and Shahid, 2019b). GSMaP\_GC is generated by correcting bias of GSMaP\_NRT using observed precipitation database of climate prediction center (CPC) of the National Oceanic and Atmospheric Administration (NOAA). The CPC database includes 14 rain gauges records, mostly located in the coastal region of peninsular Malaysia. They are used to correct for GSMaP\_NRT rainfall.

Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) was re-sampled to the spatial resolution of GSMaP ( $0.1^{\circ} \times 0.1^{\circ}$ ) and shown in Figure 1(a). Distance to coast at each GSMaP grid over peninsular Malaysia was estimated using proximity tool, available in ArcGIS (v. 10.7.1). Distance map was then converted to raster, maintaining resolution of GSMaP ( $0.1^{\circ} \times 0.1^{\circ}$ ) (Figure 1(b)).

### **3. Methodology**

Observed hourly rainfall data were used as reference for correcting bias of GSMaP\_NRT product. The GSMaP\_NRT data was interpolated to station locations using an ordinary kriging technique as it is one of the most suitable methods for rainfall interpolation in Peninsular Malaysia (Jamaludin and Suhaimi, 2013; Hassim et al., 2020). A two-stage bias correction method was proposed in this study. In the first step, a classifier was used to categorise rainfall of different intensity based on observed data, and then a regression algorithm was used to predict rainfall for each intensity category. An algorithm, based on an ensemble of classification and regression trees (CART), known as random forest (RF), was used to classify rainfall categories and prediction of rainfall amounts for each category. A separate model was developed for each month to account for the influence of wind in different months. Out of 80 stations, 70% of stations (56 stations) was used for model development and the rest (24 stations) was used for validation. Stations used for model calibration and validation are shown in Figure 1 (black and red dots).

Methodological procedure of this study is presented in Figure 2. GSMaP\_NRT rainfall, elevation of each station, and distance to coast were used as inputs from which observed rainfall was derived. A classification model is developed first to classify GSMaP\_NRT rainfall to different intensity categories. Thresholds used for classification of rainfall are given in Table 1. Note that different regions classify rainfall intensity differently. For example, American Meteorological Society classifies  $<2.5$  mm/hour as light rainfall and  $>7.5$  mm/hour as heavy rain, while UK Meteorological Office categorises  $< .5$  mm/hour as light and  $> 4.0$  mm/hour as heavy rainfall (Barthiban et al., 2012). There is no universal definition to classify

rainfall intensity for peninsular Malaysia. In Indonesia, rainfall intensity is classified as 0-1 as no rain, >1 -5 as light rainfall, >5-10 as moderate and >10 as heavy and >20 as very high rainfall. In Malaysia, Tan et al. (2015) classified hourly rainfall <2mm as low based on probability distribution function. Therefore, classification developed by Indonesian Meteorological Department was adopted here with a slight modification of Tan et al. (2015) to define low rainfall intensity.

Table 1 Thresholds used to classify hourly rainfall in Peninsular Malaysia

<b>Rainfall categories</b>	<b>Threshold (mm)</b>
Low	$\leq 2$
Moderate	$> 2$ to $\leq 5$
Moderately high	$> 5$ to $\leq 10$
High	$> 10$ to $\leq 20$
Extreme	$> 20$

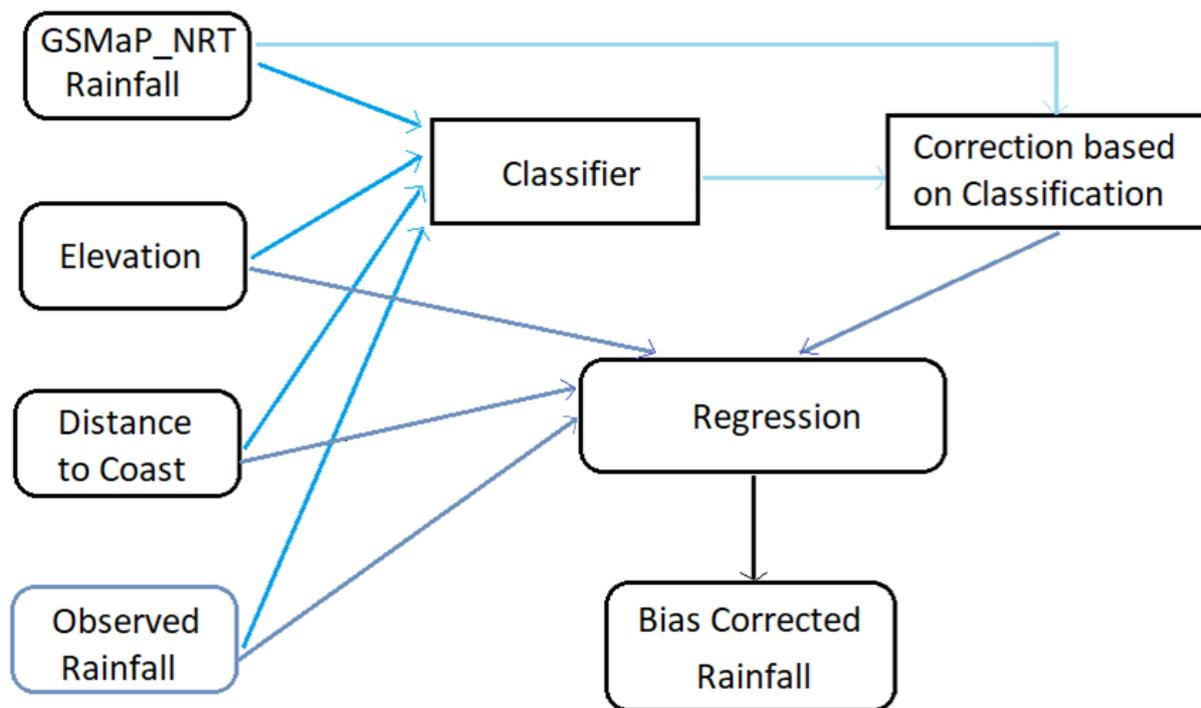
Classified values of GSMaP\_NRT rainfall, elevation of rainfall location, and distance to coast were used as inputs and classified values of observed rainfall was outputted to develop classification model. The output was then compared with GSMaP\_NRT rainfall to make the first stage of the model. If GSMaP\_NRT rainfall was within the range of classified model by the classifier, no correction is made. However, if it was found more than the upper range of the category modelled by the classifier, GSMaP\_NRT rainfall was considered equal to the upper bound of the category or vice versa. The correction based on classified outputs can be expressed as:

$$R_{c-c} = \begin{cases} U_T & \text{If } R_G > U_T \\ R_G & \text{If } U_T > R_G \leq U_T \\ L_T & \text{If } R_G < U_T \end{cases} \quad (1)$$

where  $R_{c-c}$  is the corrected rainfall based on classification,  $R_G$  is GSMaP\_NRT rainfall;  $U_T$  is the upper bound; and  $L_T$  is the lower bound of a category. For example, if a GSMaP\_NRT rainfall amount of 7 mm was classified as moderate category ( $> 2$  to  $\leq 5$  mm), then it was corrected as equal to the upper bound of that category (or 5 mm).

In the next step, a regression model was developed to predict the amount of rainfall in each class. However, separate model was used for each category. The model was developed as described above, except for rainfall amount. In this case, classified value of rainfall amount was used to predict depth of rainfall.

Calibrated models were used to correct the bias of GSMaP\_NRT rainfall during the validation period (April 2012 to Dec 2018). Results from validation for all months were merged to assess the performance of the bias-correction method. The bias-corrected rainfall using the model proposed in this study is termed as GSMaP\_BC. The GSMaP\_BC was then compared with GSMaP\_NRT and GSMaP\_GC during the validation period to indicate the efficacy of proposed method. The performance of the bias correction method was measured using a set of statistical metrics. A brief description of the RF algorithm and the statistical metrics used for evaluating the performance are described below.



**Figure 2** Workflow, used for bias correction of GSMaP\_NRT rainfall product

### 3.1 Random Forest

Random Forest is a promising classification and regression method, developed by Breiman (2001) by combining previously developed ‘random decision trees’ by Ho (1995) and his own method ‘bagging of predictors’ (Breiman 1996). In brief, a random forest method uses a series or ‘forests’ of classification or regression trees (CART) that partition a set of explanatory

variables recursively to predict a response variable. The algorithm first creates a training dataset by sampling cases randomly from a large dataset with replacement, popularly known as ‘bootstrap samples’ from which the trees are ‘grown’. Then RF selects a subset of explanatory variables and finds a variable (predictor) that partitions the response variable the ‘best’ in the training dataset. The partitioning is performed based on the identification of predictor that minimizes error if the response variable is continuous. However, if the response variable is categorical then it minimizes within-group variance. This optimal prediction forms the first node in the tree and also splits the data. In the same way, the algorithm randomly selects a subset of predictors each time at each node and split the data. The process stops when further split is not possible. The random selection method of samples and explanatory variables reduces bias and variation within and between trees, resulting in an increase in the predictive power of the technique (Rokach, 2016). For classification trees, samples those are not included in the training data are referred to as ‘out-of-bag’ (OOB) and they are then used to predict their response using the tree. The misclassification of cases is subsequently used as an estimate of the predictive error rate known as ‘out-of-bag error’ (OOB-ER).

Mathematically, a random forest (Breiman, 2001) can be defined as a collection of decision tree predictors  $g(x; \theta_k), k = 1, 2, \dots, K$  where,  $x$  is the observed set of explanatory variables (predictors) associated with random vector  $X$ ,  $\theta_k$  is the parameter vector assumed to be independent and identically distributed random vector. The training data (observed) are assumed to be independently drawn from a joint distribution of  $(X, Y)$  and forms  $n(p + 1)$  tuples  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ . In case of regression, random forest prediction is the average over the set as:

$$\bar{g}(X) = \frac{1}{K} \sum_{k=1}^K g(x; \theta_k) \quad (2)$$

The average prediction error (PE) for an individual tree can be written as:

$$PE = E_{\theta} E_{XY} (Y - g(x; \theta_k))^2 \quad (3)$$

RF is employed in such a way that it (i) grows trees to maximum depth keeping individual error low; (ii) grows each tree based on a bootstrap sample from the training data; (iii) selects  $m$  covariates and chooses the best split of that node, based on these covariates at each node of every tree.

RF employs many decision trees and combines them to provide a more precise and stable prediction. Therefore, RF is a better prediction compared to a single model. RF randomizes in two stages, i.e., selecting training subset and then select variable at each node of a tree. This helps RF to overcome issue of overfitting which is usually common in other tree-based algorithms. Additionally, random subsampling helps RF to generalize data in an efficient way to obtain a better prediction. Internal validation and relative importance of the variables are evaluated in RF during data sampling which is also used to estimate optimum number of trees. Therefore, optimization of hyperparameter of RF (number of trees) is not a pre-requisite, common to other machine learning algorithms. The RF can be used for both classification and regressions, consequently, it is flexible and used in diverse applications. Fern´andez-Delgado et al. (2014) evaluated 179 classifiers and showed that RF is the best among classifiers. Sa'adi et al (2017) summarized that RF is capable of avoiding overfitting, handling a large amount of data, and has better analytical flexibility. The bias correction algorithm proposed in this study needs both classification and regression algorithms. Considering its robust prediction capability and diverse applicability, RF was therefore chosen. The “randomForest” package in R program was used for model development.

### 3.2 Performance Evaluation of Prediction Models

To evaluate the performance of the proposed method, observed rainfall data at each grid was compared with GSMaP\_NRT, GSMaP\_BC, and GSMaP\_rainfall. Four categorical indices, namely Probability of Detection (POD), Critical Success Index (CSI), hit BIAS (HB) and False Alarm Ratio (FAR) (Rozante et al., 2018; Nashwan et al., 2019) were used to evaluate the performance of the classifier. Four statistical indices, namely normalized root mean square error (NRMSE), Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), modified index of agreement (md) (Willmott, 1981) and Kling-Gupta Efficiency (KGE) index (Kling et al., 2012) were used to evaluate accuracy in replicating observed rainfall amount. Detail description of the metrics used in this study can be found in Nashwan et al. (2019). The four categorical indices are defined as:

$$POD = \frac{Hits}{Hits+Misses} \quad (2)$$

$$CSI = \frac{Hits}{Hits+Misses+ False Alarms} \quad (3)$$

$$HB = \frac{Hits+Flase Alarms}{Hits+Misses} \quad (4)$$

$$FAR = \frac{False\ Alarms}{Hits+False\ Alarms} \quad (5)$$

where, *Hits* represent number of success in categorizing the GSMaP\_NRT rainfall to observed class while *Misses* indicate vice versa. *False Alarms* indicate a prediction of rainfall when no rainfall occurred. The ideal value for POD, CSI and HB is 1 while it is 0 for FAR. The statistical indices used here are defined as:

$$NRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - O_i)^2}}{\sigma} \quad (6)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (x_{sim,i} - x_{obs,i})^2}{\sum_{i=1}^n (x_{obs,i} - \bar{x}_{obs})^2} \quad (7)$$

$$MD = 1 - \frac{\sum_{i=1}^n (x_{obs} - x_{sim})^j}{\sum_{i=1}^n (|x_{sim} - \bar{x}_{obs}| + |x_{obs} - \bar{x}_{obs}|)^j} \quad (8)$$

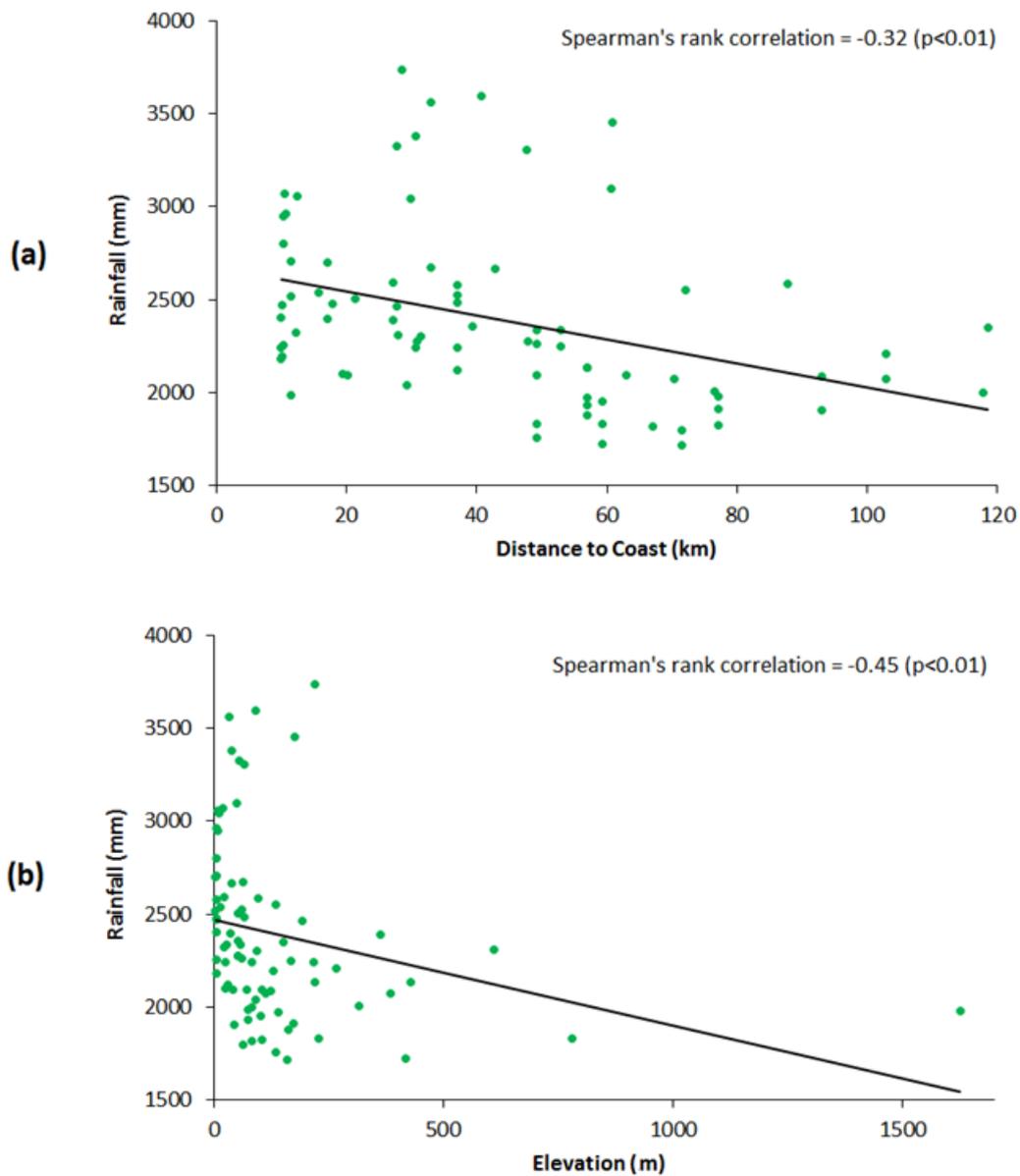
$$KGE = 1 - \sqrt{(r - 1)^2 + \left(1 - \frac{\mu_{sim}}{\mu_{obs}}\right)^2 + \left(\frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}}\right)^2} \quad (9)$$

where  $x_{sim,i}$  and  $x_{obs,i}$  are the  $i$ -th GSMaP and observed data;  $n$  is the sample size;  $r$  is Pearson's correlation;  $\mu$  and  $\sigma$  are mean and standard deviation, respectively of bias-corrected (*sim*) and observed (*obs*) data. The values of NRMSE ranges between 0 and  $\infty$ , NSE between 1.0 and  $-\infty$ , MD between 1.0 and -1.0, and KGE between  $-\infty$  and 1.0. The optimum value of NSE, MD and KGE is 1.0 while NRMSE is 0.

## 4. Results

### 4.1 Correlation between rainfall and distance to coast and elevation

The relationship of annual mean rainfall with distance to coast and elevation is presented in Figure 3 which shows a decrease in rainfall according to two parameters. Considering non-normal distribution of distance to coast and elevation variables, Spearman's rank correlation was performed. The correlation coefficients between rainfall and distance to coast and elevation were -0.32 and -0.45 ( $p < 0.01$ ), suggesting an influence of these two physical parameters on rainfall in the study area.



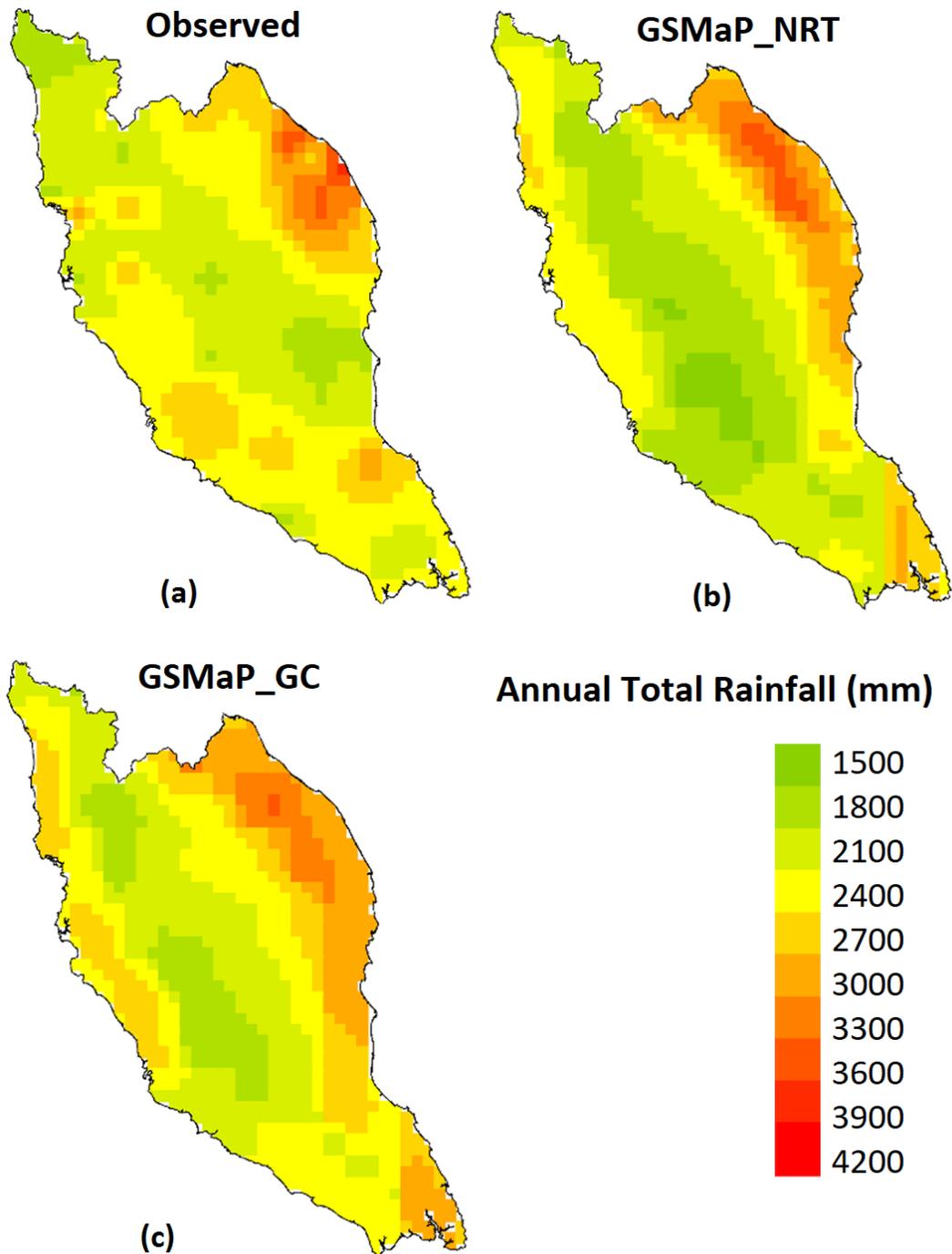
**Figure 3** Correlations between annual mean rainfall: (a) distance to coast; and (b) elevation

#### 4.2 GSMaP in reconstructing spatial distribution of mean annual rainfall

Annual average precipitation of peninsular Malaysia estimated with observed, GSMaP\_NRT, and GSMaP\_GC rainfall is presented in Figure 4. Observed precipitation data at 80 locations were gridded to the same resolution of GSMaP ( $0.1^{\circ} \times 0.1^{\circ}$ ).

Both GSMaP\_NRT and GSMaP\_GC were able to show high rainfall in the northeastern coast and low rainfall area of central region. However, low rainfall region in Peninsular Malaysia is elongated from northwest corner to central east coast, while both GSMaP\_NRT and GSMaP\_GC showed that the extent of low rainfall zone stretches between central north

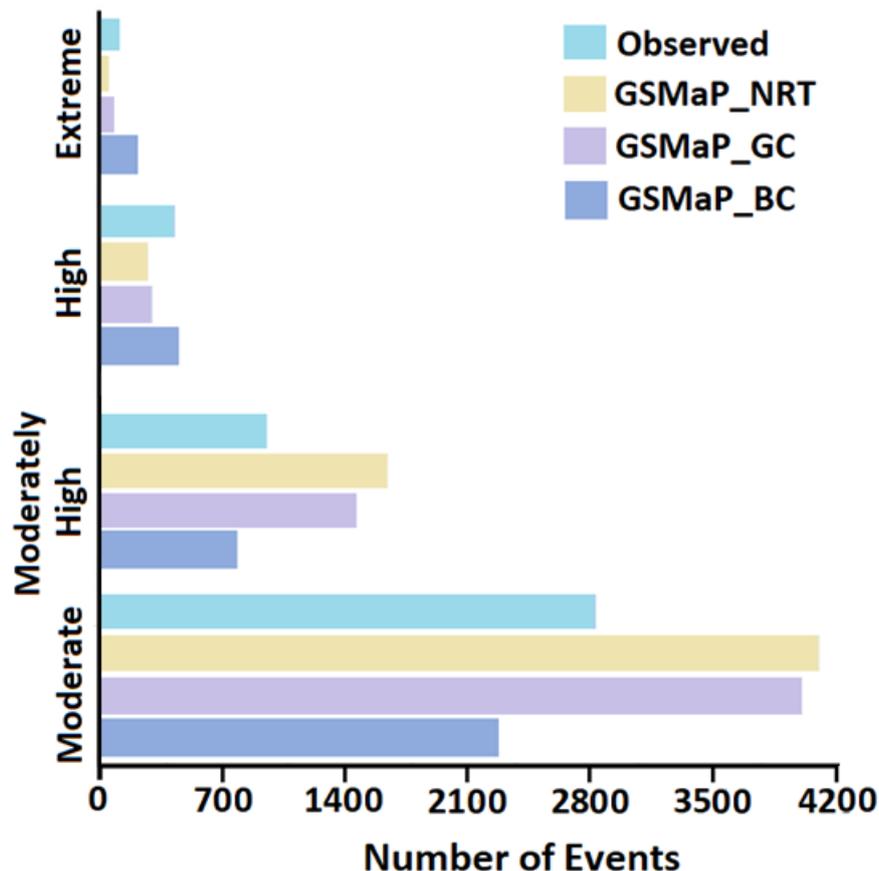
and west coast. Both products also failed to replicate low rainfall zone in the south of peninsula. This means that gauge correction was insufficient to improve the performance of GSMaP\_NRT precipitation in the region.



**Figure 4** Annual average precipitation of peninsular Malaysia, derived from observed records, GSMaP\_NRT and GSMaP\_GC during 2000–2018

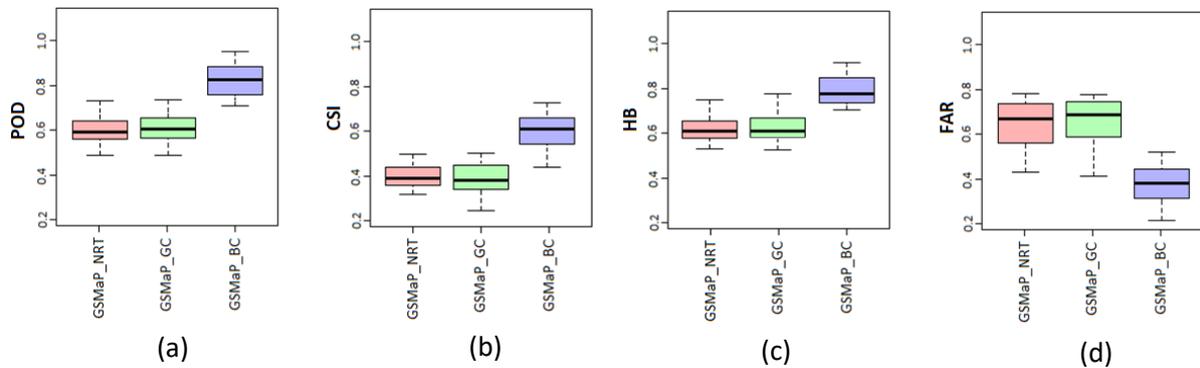
### 4.3 Performance in reconstructing different rainfall categories

The performance of classification of GSMaP\_NRT rainfall was estimated by comparing the number of events belongs to different categories detected using categorical metrics. The total number of observed events belongs to different categories and those that are detected by different rainfall products are presented in Figure 5. Rainfall data at all stations used for validation (24 stations) was used to estimate rainfall events. The number of events belongs to low rainfall category was very high compared to other categories for all products. Therefore, low rainfall events are not presented in Figure 5 in order to have a useful comparison of rainfall events that belong to different categories of different products. The number of observed low rainfall events was 118,321, while detected number of low rainfall events by GSMaP\_NRT, GSMaP\_GC, and GSMaP\_BC were 125,241, 124,674, and 116,133, respectively. Figure 5 shows that the number of rainfall events of other categories estimated by GSMaP\_BC was also closest to the observed value. The GSMaP\_NRT estimated a higher number of moderate and moderately high rainfall events, while less number of high and extreme rainfall events compared to GSMaP\_GC. This indicates that mean annual rainfall estimated by GSMaP\_NRT was close to GSMaP\_GC. Therefore, spatial distribution of annual average rainfall of GSMaP\_NRT and GSMaP\_GC resembled to each other (Figure 2).



**Figure 5** Categories of rainfall events calculated from observed rainfall, GSMaP\_NRT, GSMaP\_GC and GSMaP\_BC data

The performance of classified GSMaP\_NRT rainfall at each grid point was estimated using categorical indices and results are presented in Figure 6 which showed a large improvement of bias-corrected rainfall in terms of categorical indices. The POD for GSMaP\_NRT and GSMaP\_GC were estimated to be 0.59 and 0.60, while it was 0.85 for GSMaP\_BC. A similar improvement was also noticed in terms of other metrics. The CSI, HB and FAR of GSMaP\_BC were 0.68, 0.79 and 0.38, respectively, much better than that of GSMaP\_NRT and GSMaP\_GC.



**Figure 6** Performance of the classifier in classifying GSMaP\_NRT rainfall to different observed rainfall categories estimated using four categorical metrics: (a) probability of detection; (b) critical success index (CSI); (c) hit bias (HB); and (d) false alarm ratio (FAR)

#### 4.4 Performance in estimating rainfall depths

The performance of bias-corrected rainfall compared to GSMaP\_NRT and GSMaP\_GC was evaluated for each rainfall category and for complete timeseries. The performance was evaluated using both statistical metrics and visual inspection of scatterplot and the Taylor diagram.

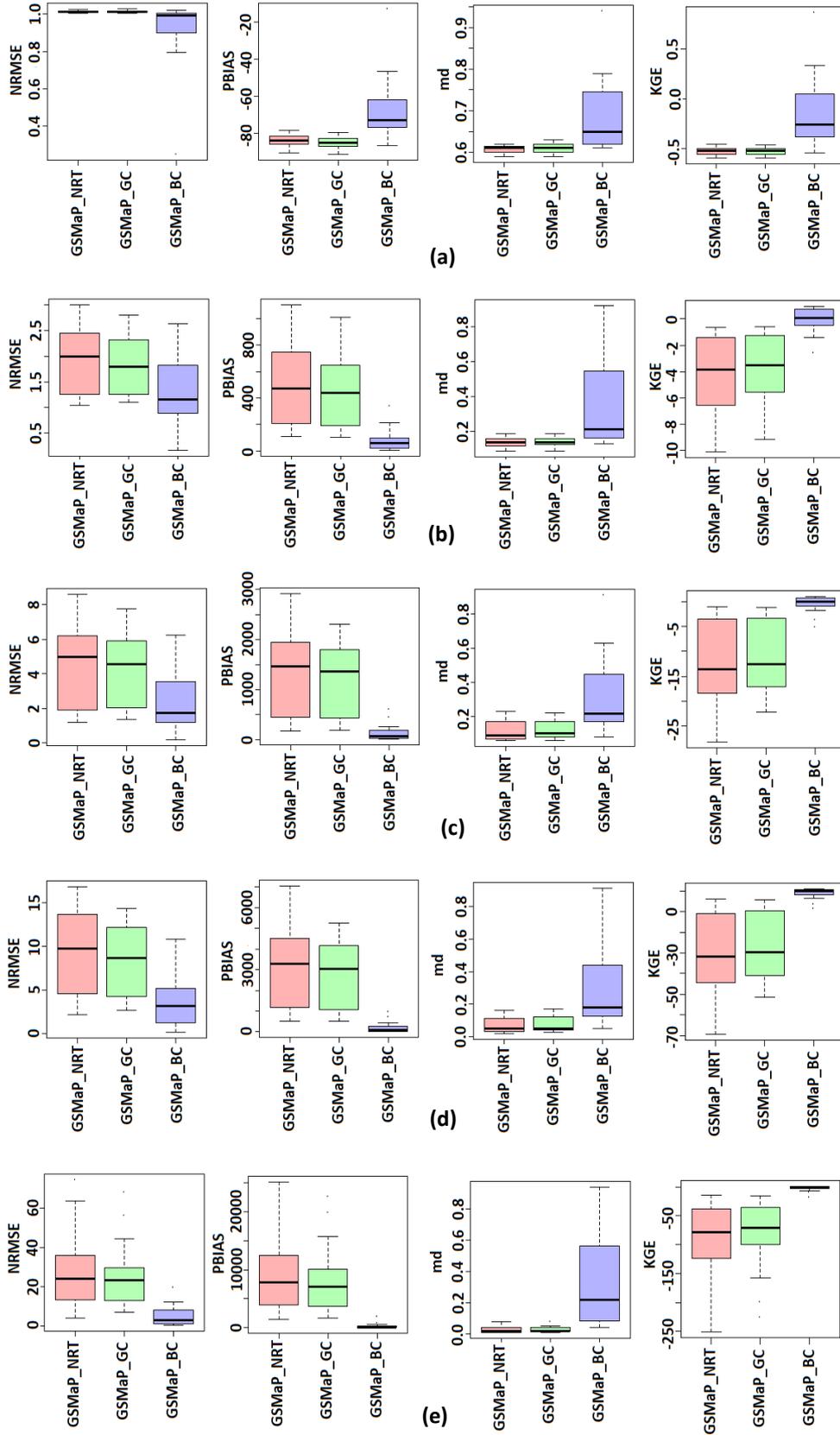
The performance of different rainfall products compared to observed data for rainfall events  $\leq 2$  mm/hour, based on four statistics, are presented in Figure 7. In this figure, median (horizontal line in the box), interquartile range (height of the box), and total range with outliers (whiskers and dots) of statistical metrics, estimated at different grids are shown. The boxplots of the metrics clearly indicated a higher performance of GSMaP\_BC compared to

GSMaP\_NRT and GSMaP\_GC. The median of NRMSE, PBIAS, md and KGE for GSMaP\_BC were much higher than those for GSMaP\_NRT and GSMaP\_GC.

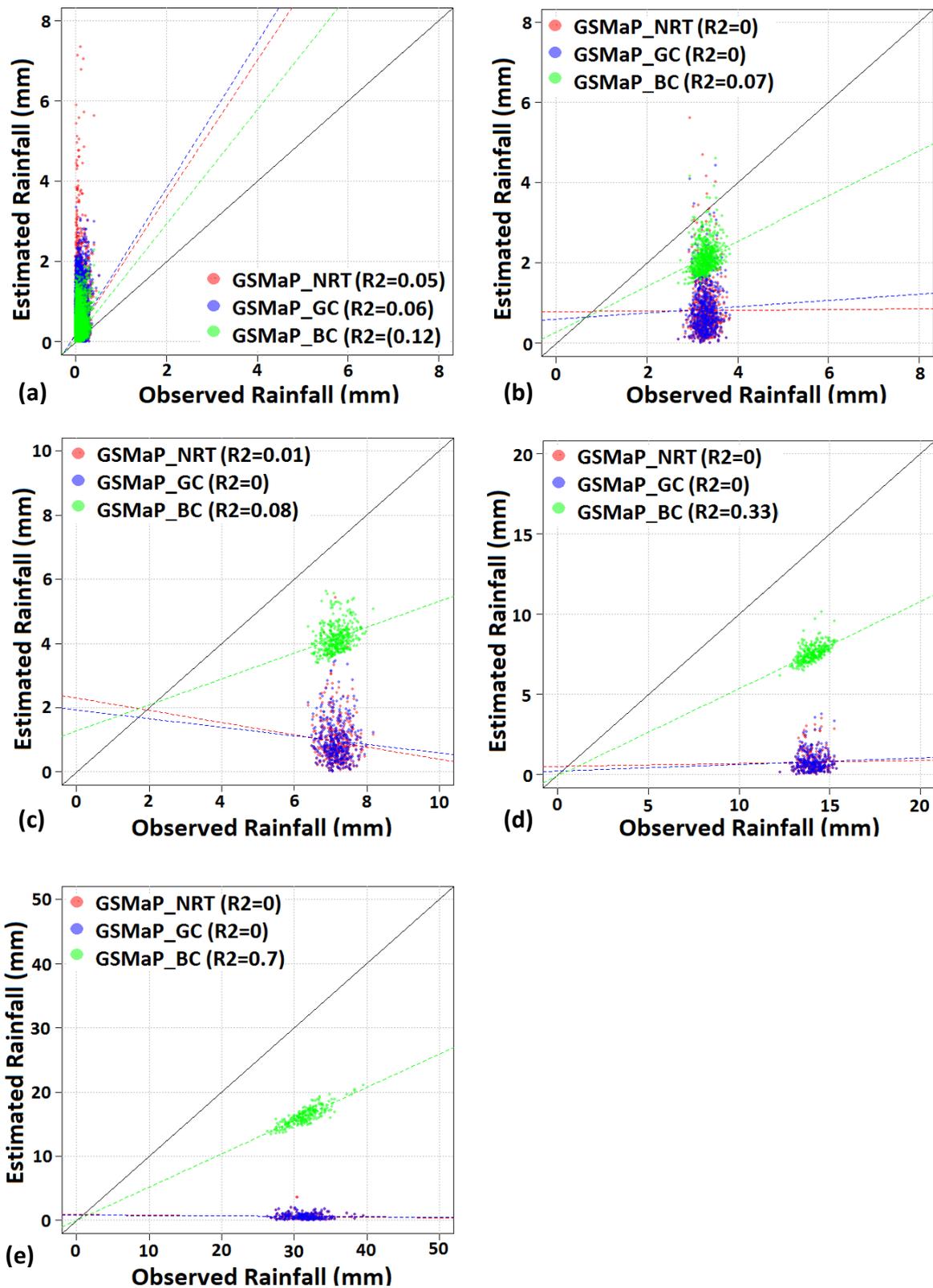
Results for other rainfall categories were also found more or less similar. GSMaP\_BC performed much better than GSMaP\_NRT and GSMaP\_GC. The median NRMSE for low, moderate, moderately high, high, and extreme rainfall categories were 0.9, 1.0, 1.8, 3.6, and 4.9 for GSMaP\_BC while they were 1.1, 2.0, 4.9, 9.1 and 23.1 for GSMaP\_NRT and 1.1, 1.9, 4.7, 7.9 and 22.1 for GSMaP\_GC. Similar improvements were noticed in terms of other statistics.

The height of the box for GSMaP\_BC was found less compared to GSMaP\_NRT and GSMaP\_GC for all statistics except md, which indicated a similar performance of the GSMaP\_BC for most of the grid points. The height of the box for md was much higher for GSMaP\_BC, but location of the median line in the lower side of the box indicated near to the perfect value of md in some locations.

The scatterplot of observed and estimated rainfall by the three products is shown in Figure 8. The GSMaP\_NRT, GSMaP\_GC, and GSMaP\_BC rainfall were presented using orange, blue and green dots. The GSMaP\_NRT and GSMaP\_GC rainfall appeared to be completely scattered. The correlation coefficient for both products was always  $<0.06$ , suggesting a poor outcome. The correlation lines for the products (orange and blue lines) were far apart from the diagonal line of the plot. The performance appeared to improve significantly after bias correction method. The GSMaP\_BC aligned with diagonal line of the scatterplot. The green line represents that the GSMaP\_BC was much closer to the diagonal line with a coefficient value of 0.12, 0.07, 0.08, 0.33, and 0.7 for low, moderate, moderately high, high, and extreme rainfall categories, much higher than GSMaP\_NRT and GSMaP\_GC.

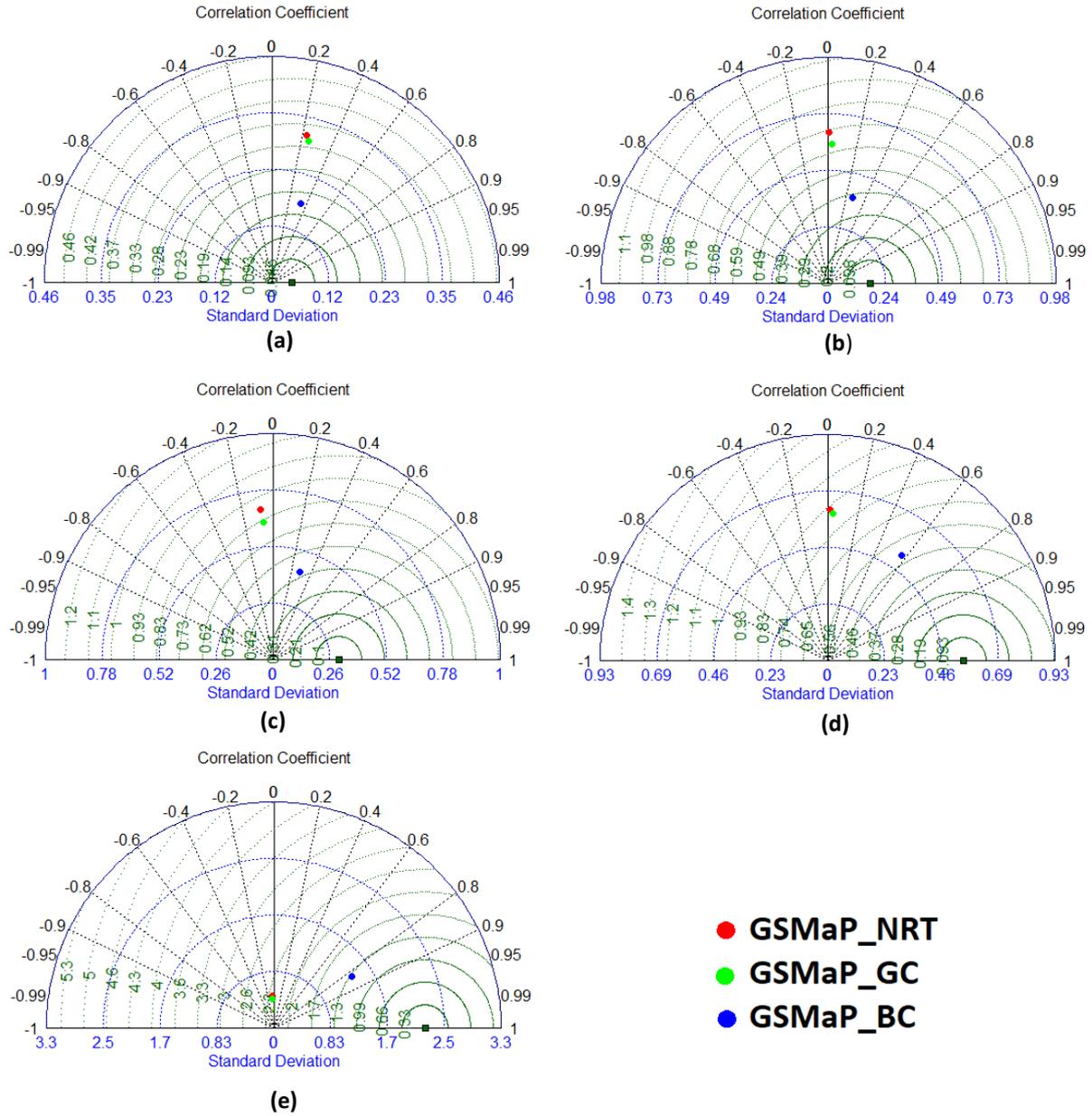


**Figure 7** Performance of bias correction method in replicating five categories of observed rainfall: (a) low; (b) moderate; (c) moderately-high; (d) high; and (e) extreme



**Figure 8** Performance of bias correction method in replicating five categories of observed rainfall: (a) low; (b) moderate; (c) moderately-high; (d) high; and (e) extreme

The areal average of observed and three rainfall datasets was used to compare their performance using the Taylor diagram. Three statistical metrics (correlation, standard deviation (SD) and root-mean-square (RMS) error)) are presented to compare timeseries rainfall. The Taylor diagram presented in Figure 9 shows that the points representing GSMaP\_BC rainfall are much closer to observed values (deep green square box in x-axis) compared to GSMaP\_NRT and GSMaP\_GC. The difference in SD of GSMaP\_BC and observed rainfall was less than 0.1 while RMS difference was only 0.15 for low rainfall category. Rainfall values of all other categories also improved significantly after accounting bias. In all cases, the point representing bias-corrected rainfall was much closer to observed value.



**Figure 9** Taylor diagram, showing the performance of bias correction in replicating five categories of observed rainfall: (a) low; (b) moderate; (c) moderately-high; (d) high; and (e) extreme

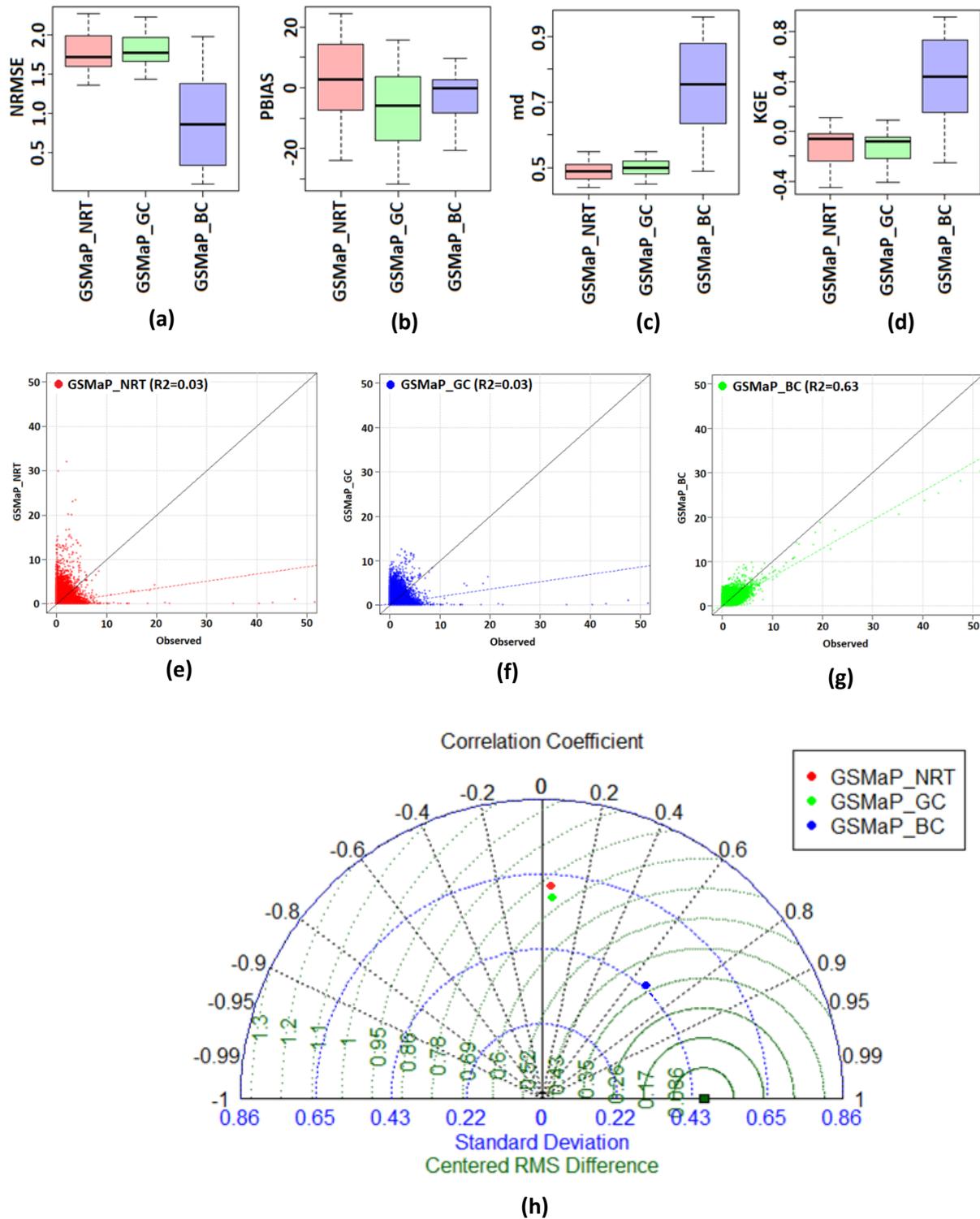
#### 4.5 Performance in reconstructing complete rainfall time series

The performance of bias-correction was further evaluated for whole timeseries at each grid location and results are presented in Figure 10. The statistical metrics estimated for complete time series revealed about 110% reduction of NRMSE in GSMaP\_BC (0.8) compared to GSMaP\_NRT (1.7) and GSMaP\_GC (1.75) (Figure 10a). The PBIAS of GSMaP\_BC was found near to zero (0.3) while it was 2.1 and -3.1 for GSMaP\_NRT and GSMaP\_GC. The md values for GSMaP\_NRT, GSMaP\_GC and GSMaP\_BC were 0.48, 0.51 and 0.75 and the KGE

as -0.31, -0.33 and 0.51, signifying a big improvement in terms of the performance of GSMaP\_NRT rainfall, after bias correction. The performance was also higher compared to gauge-corrected rainfall product of GSMaP.

The scatterplot of GSMaP datasets against observed rainfall are presented in Figure 10(b). Complete timeseries at all 24 stations used for validation were averaged to prepare the scatterplot. The range of rainfall was reduced (0 to ~50 mm) due to averaging. The scatterplot showed a complete scattering of GSMaP\_NRT and GSMaP\_GC rainfall with near to zero correlation coefficient (0.03). In contrast, GSMaP\_BC rainfall was aligned with the diagonal line. The  $R^2$  value of 0.63 for GSMaP\_BC, indicating a significant improvement of rainfall values with the proposed method.

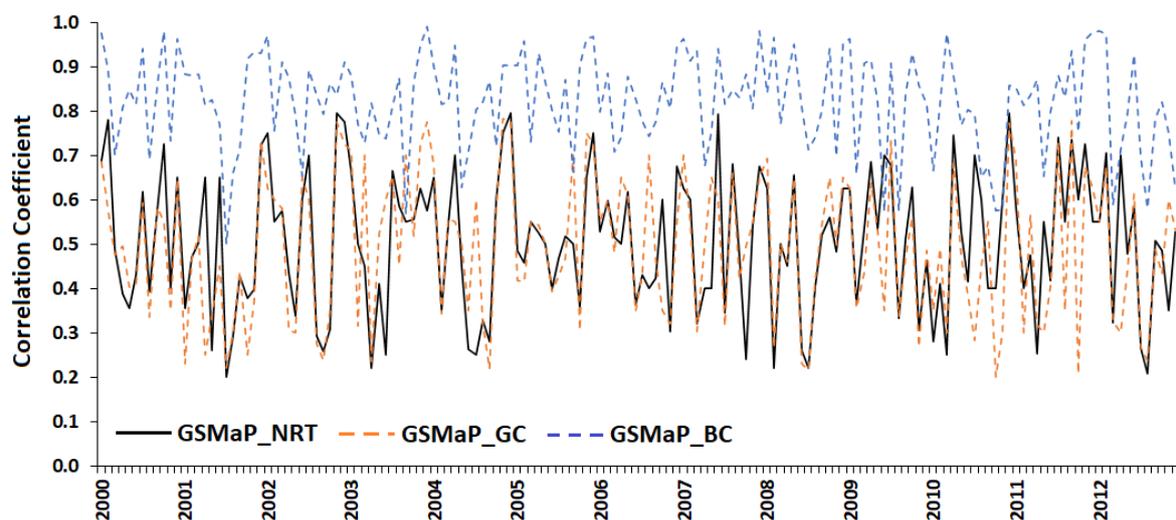
The Taylor diagram in Figure 10(c) also revealed that the point representing GSMaP\_BC rainfall was much closer to observed rainfall compared to GSMaP\_NRT and GSMaP\_GC. The difference in SD of GSMaP\_BC and observed rainfall was less than 0.4 while the RMS difference was only 0.5.



**Figure 10** Taylor diagram, showing the performance of bias correction method in replicating complete timeseries of observed rainfall in terms of: (a) NRMSE; (b) PBIAS; (c) md; and (d) KGE: scatter plot of observed rainfall against (e) GSMaP\_NRT; (f) GSMaP\_GC; and (g) GSMaP\_BC; and (h) Taylor diagram

#### 4.6 Performance in replicating spatial pattern of rainfall

The performance of the model in replicating spatial pattern of observed rainfall was finally evaluated. The correlation of monthly observed and GSMaP estimated rainfall at each grid point was estimated for all months during the validation process. The Pearson's correlation coefficient for different months for three products are presented in Figure 11, which revealed higher  $R^2$  of GSMaP\_BC rainfall with observed rainfall compared to GSMaP\_NRT and GSMaP\_GC for all months. The  $R^2$  values for GSMaP\_BC were mostly  $>0.7$  while they were in the range of 0.2–0.78 for GSMaP\_NRT and GSMaP\_GC. The  $R^2$  values for GSMaP\_BC were higher than the critical value of 0.505 ( $p=0.01$ ). This indicated capability of GSMaP\_BC to replicate spatial pattern of observed rainfall accurately, and therefore, capable of monitoring hydrological events such as floods.



**Figure 11** Spatial association of different rainfall products with observed rainfall over peninsular Malaysia for all months

## 5. Discussion

Estimation of precipitation with spatial and temporal accuracy is very indispensable for hydro-meteorological studies (Iqbal et al., 2020, Ahmed et al., 2017). Satellite-based products are becoming reliable source to obtain spatially distributed rainfall estimates at regional and global scales (Bhatti et al., 2016, Noor et al., 2019a). The requirement of high spatial resolution precipitation is more demanding in tropical region because of its large variability over a short distance. Therefore, several studies have been conducted to evaluate the performance of different satellite-based precipitation products, including IMERG-E, IMERG-L, IMERG-F, GSMaP\_NRT, GSMaP\_GC, TRMM 3B42RT, CMORPH, TMPA 3B42V7, PERSIANN-CDR and GPCP-1DD in peninsular Malaysia. Comparing the results of this study with existing

studies revealed better performance of GSMaP rainfall products for peninsular Malaysia (Tan et al., 2015; Paska et al., 2017; Tam et al., 2019; Soo et al., 2020; Noor et al., 2020). Tam et al. (2019) showed GSMaP-NRT is the best product for hydrological simulation in the Kelantan River basin, located in northeast high rainfall region. Noor et al. (2020) showed GSMaP\_GC is the best product for estimating rainfall intensity-duration-frequency curve for peninsular Malaysia. Despite better performance of GSMaP products in peninsular Malaysia, they still showed large bias. Noor et al. (2020) showed that GSMaP\_GC underestimated rainfall intensity in the range of 8–27% while underestimation by GSMaP\_NRT was in the range of 35–49%. Therefore, this study attempted to improve the performance of GSMaP\_NRT rainfall in peninsular Malaysia.

Two physical factors, distance to coast and elevation were considered to correct for bias in GSMaP\_NRT rainfall. Mean annual rainfall in peninsular Malaysia was found to have a significant negative association with distance to coast and elevation parameters. Northeast monsoon (NEM) brings moist air from the South China sea that produces intense rainfall in west coast of Malaysia. The southwest monsoon (SWM) flows from the Indian Ocean to eastern coast of the peninsula. Therefore, western coastal region receives higher rainfall during NEM and eastern coastal region receives during the SWM. Rainfall gradually decreases from coast to inland. Therefore, a significant correlation ( $p < 0.05$ ) between annual mean rainfall and distance to coast was observed in this study. Negative association between rainfall and elevation in peninsular Malaysia is due to less inter-monsoonal rainfall in high elevation zones (Wong et al., 2009). Besides two monsoons, a major portion of rainfall occurs during two inter-monsoonal break periods (April and October). Thunderstorm is a major cause of intense rainfall during inter-monsoonal periods. Thermally driven local circulation develops convective clouds and causes localized rainfall in the late afternoon, mostly in the valleys and in flat coastal regions (Joseph et al., 2008). In addition, most of the elevated interior regions receive less rainfall during NEM due to blockage of westward monsoon progression by the Titiwangsa mountain range (Juneng et al., 2007). The relationship of annual mean rainfall with elevation (Figure 3) showed a very weak association between rainfall and elevation for low altitude and high for elevated zones. As most of the stations used to derive the relationship are located at low elevation, overall correlation of rainfall with elevation was weak. However, the correlation was still significant at  $p < 0.05$ .

A major drawback of satellite precipitation products is overestimation of the occurrence of rainfall. This was also observed in this work. Both GSMaP\_NRT and GSMaP\_GC estimated

more low and moderate rainfall events and underestimated high and extreme rainfall events. Classification of rainfall in the first stage of bias correction approach, proposed in this study, helped correcting classification of rainfall, and thus, reduces uncertainty in the estimation compared to other bias correction approaches that are based on mean, variability and distribution matching. Pour et al. (2016) showed better performance in bias correction of global circulation model (GCM) precipitation, especially when a classification is used to identify rainfall days before estimating rainfall amounts using a regression model.

This study revealed a large improvement in the performance of GSMaP\_NRT rainfall in terms of all statistics, and the proposed bias correction method with two physical parameters is highly effective. The PBIAS in GSMaP\_NRT was in the range of -24–23% with a median of 2.1%. After correction, the bias in GSMaP\_BC was in the range of -10–9.5% with a median of 0.3%, much lower than the median bias of GSMaP\_GC (-3.1%). The use of physical factors significantly affected the prediction of rainfall which can be employed for reducing bias in satellite-based rainfall products. However, the proposed method is highly dependent on the classification scheme of rainfall events. The use of RF also improved the performance of GSMaP\_NRT rainfall. In future, other classification and regression algorithms can be used to compare the performance of the proposed bias correction method.

The bias correction model was developed using rainfall data of 70% of the stations selected randomly. The performance of any prediction model varies according to the percentage of data used for model development. For developing IDF models at ungauged locations in peninsular Malaysia, Noor et al. (2020) employed 80% of data for model development and 20% for validation. For the prediction of seasonal extreme rainfall events in peninsular Malaysia, Pour et al. (2020) used 70% of data for model development and the remaining 30% data for validation. A review of previous studies indicated that no certain rule can be used for model calibration and validation. In future, the performance of bias correction method for different data classification can be evaluated.

Rainfall in peninsular Malaysia, particularly convective rainfall during inter-monsoonal periods, is highly localized. The cell diameters of such convective rainfall are usually <10 km (Schroerer and Sungmin, 2018), which may not be detected by satellite sensors, having a spatial resolution of 10 km. However, a comparison was made in this study among rainfall products having the same spatial resolution ( $0.1^\circ \times 0.1^\circ$ ). The categorical metrics like FAR may not provide an accurate estimation of the rainfall events, however, they still can be used to evaluate

relative performance of different rainfall products with the same spatial resolution. This fact should be considered in the interpretation of the results, presented in this work.

## **6. Conclusion**

A physical empirical model was developed for correcting bias of GSMaP precipitation product. The performance of the proposed method was evaluated by comparing it with GSMaP\_NRT and GSMaP\_BC. Results revealed a significant improvement in GSMaP\_BC rainfall product compared to GSMaP\_NRT and GSMaP\_GC. The two-step bias correction approach, based on classification and regression techniques, was able to reconstruct categories of different rainfall intensity using observed data with a reasonable error. The bias-corrected rainfall was able to replicate spatial distribution of observed rainfall. Accurate near-real-time high-resolution hourly rainfall is very important for monitoring and forecasting floods and landslides, triggered by intense rainfall. Accurately corrected near real-time GSMaP data can be used for this purpose in the absence of dense rain-gauge network. This can be particularly important for peninsular Malaysia, where flash floods and landslides, driven by short-lived intense rainfall, are common. The proposed methods can be applied to other near-real-time satellite precipitation datasets to find the best bias-corrected rainfall product for peninsular Malaysia. Besides, other classification and regression algorithms, particularly those based on intelligence algorithms can be used to explore further improvement of the model.

## References

- Abdullah, F., Mohammad, S.N., Mohamad, J., Ahmad, M. (2019) The Economic Model for Flood Damage Cost in Retailing Business in Malaysia. In: Mat Noor A., Mohd Zakuan Z., Muhamad Noor S. (eds) Proceedings of the Second International Conference on the Future of ASEAN (ICoFA) 2017 - Volume 1. Springer, Singapore. [https://doi.org/10.1007/978-981-10-8730-1\\_51](https://doi.org/10.1007/978-981-10-8730-1_51)
- Ahmed, K., Shahid, S., Ali, R.O., Harun, S.B., Wang, X.J. (2017) Evaluation of the performance of gridded precipitation products over Balochistan Province, Pakistan. *Desalination and Water Treatment* 1, 14. doi:10.5004/dwt.2017.20859
- Alharbi, R., Hsu, K., Sorooshian, S. (2018) Bias adjustment of satellite-based precipitation estimation using artificial neural networks-cloud classification system over saudi arabia. *Arabian Journal of Geosciences*, 11(17) doi:10.1007/s12517-018-3860-4
- Aslami, F., Ghorbani, A., Sobhani, B., Esmali, A. (2019) Comprehensive comparison of daily IMERG and GSMaP satellite precipitation products in ardabil province, iran. *International Journal of Remote Sensing*, 40(8), 3139-3153. doi:10.1080/01431161.2018.1539274
- Barthiban, S., Lloyd, B.J., Maier, M. (2012) Sanitary Hazards and Microbial Quality of Open Dug Wells in the Maldives Islands. *Journal of Water Resource and Protection*, 4, 474-486. doi:10.4236/jwarp.2012.47055
- Breiman, L. (1996) Bagging predictors. *Machine Learning* 24, 123-140.
- Breiman, L. (2001) Random forests. *Machine Learning* 45, 5-32.
- Chaudhary, S., Dhanya, C. T. (2019) Investigating the performance of bias correction algorithms on satellite-based precipitation estimates. Paper presented at the Proceedings of SPIE - the International Society for Optical Engineering, , 11149 doi:10.1117/12.2533214
- Chen, H., Yong, B., Gourley, J. J., Liu, J., Ren, L., Wang, W., Zhang, J. (2019) Impact of the crucial geographic and climatic factors on the input source errors of GPM-based global satellite precipitation estimates. *Journal of Hydrology*, 575, 1-16. doi:10.1016/j.jhydrol.2019.05.020
- Chen, H., Yong, B., Shen, Y., Liu, J., Hong, Y., Zhang, J. (2020) Comparison analysis of six purely satellite-derived global precipitation estimates. *Journal of Hydrology*, 581 doi:10.1016/j.jhydrol.2019.124376

- Deng, P., Zhang, M., Guo, H., Xu, C., Bing, J., Jia, J. (2018) Error analysis and correction of the daily GSMaP products over hanjiang river basin of china. *Atmospheric Research*, 214, 121-134. doi:10.1016/j.atmosres.2018.07.022
- Enderwin Hadi, S., Bayong, T.H.K., Gumawan, D., Siswanto (2014) Modified Convective Stratiform Technique (CSTm) Performance on Rainfall Estimation in Indonesia. *J. Math. Fund. Sci.*, 46(3): 251-268
- Fernández-Delgado, M., Cerndas, E., Barro, S. (2014) Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research* 15 (2014) 3133-3181
- Gebremedhin, M.A., Lubczynski, M.W., Maathuis, B.H.P., Teka, D. (2020) Novel approach to integrate daily satellite rainfall, with in-situ rainfall, Upper Tekeze Basin, Ethiopia. *Atmospheric Research*, 105135. <https://doi.org/10.1016/j.atmosres.2020.105135>
- Gosset, M., Alcoba, M., Roca, R., Cloché, S., Urbani, G. (2018) Evaluation of TAPEER daily estimates and other GPM-era products against dense gauge networks in west africa, analysing ground reference uncertainty. *Quarterly Journal of the Royal Meteorological Society*, 144, 255-269. doi:10.1002/qj.3335
- Hashemi, H., Nordin, M., Lakshmi, V., Huffman, G.J., Knight, R. (2017) Bias correction of long-term satellite monthly precipitation product (TRMM 3B43) over the conterminous united states. *Journal of Hydrometeorology*, 18(9), 2491-2509. doi:10.1175/JHM-D-17-0025.1
- Hassim, M., Yuzir, A., Razali, M.N., Ros, F.C., Chow, M.F., Othman, F. (2020) Comparison of Rainfall Interpolation Methods in Langat River Basin. *The 7th AUN/SEED-Net Regional Conference on Natural Disaster (RCND 2019) IOP Conf. Series: Earth and Environmental Science* 479,012018IO. Doi:10.1088/1755-1315/479/1/0120182
- Ho, T.K. (1995) Random Decision Forests. In: *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, pp. 278-282, Montreal, QC.
- Islam, M.A. (2018) Statistical comparison of satellite-retrieved precipitation products with rain gauge observations over bangladesh. *International Journal of Remote Sensing*, 39(9), 2906-2936. doi:10.1080/01431161.2018.1433890
- Jamaludin, S.S.S., Suhaimi, H. (2013) Spatial Interpolation on Rainfall Data over Peninsular Malaysia Using Ordinary Kriging. *Jurnal Teknologi* 63(2): 51-58. DOI: 10.11113/jt.v63.1912

- Ji X, Li Y, Luo X, He D, Guo R, Wang J, Bai Y, Yue C, Liu C (2020) Evaluation of bias correction methods for APHRODITE data to improve hydrologic simulation in a large Himalayan basin. *Atmospheric Research*, 242, 104964. <https://doi.org/10.1016/j.atmosres.2020.104964>
- Joseph, B., Bhatt, B.C., Koh, T.Y., Chen, S. (2008) Sea breeze simulation over the Malay Peninsula in an intermonsoon period. *Journal of Geophysical Research*, 113, D20122, doi:10.1029/2008JD010319.
- Juneng, L., Tangang, F.T., and Reason, C.J.C. (2007) Numerical case study of an extreme rainfall event during 9-11 December 2004 over the east coast of Peninsular Malaysia, *Meteorol. Atmos. Phys.*, 98, 81–98, 2007.
- Kubota, T., Ahmad, S. (2006) Wind Environment Evaluation of Neighborhood Areas in Major Towns of Malaysia, *Journal of Asian Architecture and Building Engineering*, 5:1, 199-206, DOI: 10.3130/jaabe.5.199
- Latupapua, H., Latupapua, A.I., Wahab, A., Alaydrus, M. (2018) Wireless Sensor Network Design for Earthquake's and Landslide's Early Warnings. *Indonesian Journal of Electrical Engineering and Computer Science*, 11(2): 437~445. DOI: 10.11591/ijeecs.v11.i2.pp437-445
- Ma, Q., Xiong, L., Xia, J., Xiong, B., Yang, H., Xu, C. (2019) A censored shifted mixture distribution mapping method to correct the bias of daily IMERG satellite precipitation estimates. *Remote Sensing*, 11(11) doi:10.3390/rs11111345
- Ma, Y., Yang, Y., Han, Z., Tang, G., Maguire, L., Chu, Z., Hong, Y. (2018) Comprehensive evaluation of ensemble multi-satellite precipitation dataset using the dynamic bayesian model averaging scheme over the tibetan plateau. *Journal of Hydrology*, 556, 634-644. doi:10.1016/j.jhydrol.2017.11.050
- Mastrantonas, N., Bhattacharya, B., Shibuo, Y., Rasmy, M., Espinoza-Dávalos, G., Solomatine, D. (2019) Evaluating the benefits of merging near-real-time satellite precipitation products: A case study in the kinu basin region, japan. *Journal of Hydrometeorology*, 20(6), 1213-1233. doi:10.1175/JHM-D-18-0190.1
- Mayowa, O.O., Pour, S.H., Shahid, S., Mohsenipour, M., Harun, S.B., Heryansyah, A., Ismail, T. (2015) Trends in rainfall and rainfall-related extremes in the east coast of peninsular Malaysia. *Journal of Earth System Science* 124 (8), 1609-1622

- Muhammad, M.K.I., Nashwan, M.S., Shahid, S., Ismail, T.B., Song, Y.H., Chung, E.-S. (2019) Evaluation of Empirical Reference Evapotranspiration Models Using Compromise Programming: A Case Study of Peninsular Malaysia. *Sustainability* 2019, 11, 4267
- Nashwan, M.S., Shahid, S., Abd Rahim, N. (2019a) Unidirectional trends in annual and seasonal climate and extremes in Egypt. *Theoretical and Applied Climatology* 136 (1-2), 457-473
- Nashwan, M.S., Shahid, S., Chung, E.S., Ahmed, K. and Song, Y.H. (2018) Development of Climate-Based Index for Hydrologic Hazard Susceptibility. *Sustainability* 10 (7), 2187
- Nashwan, M.S., Shahid, S., Wang, X.J. (2019b) Assessment of satellite-based precipitation measurement products over the hot desert climate of Egypt. *Remote sensing* 11 (5), 555
- Ning, S., Song, F., Udmale, P., Jin, J., Thapa, B. R., & Ishidaira, H. (2017). Error analysis and evaluation of the latest GSMaP and IMERG precipitation products over eastern china. *Advances in Meteorology*, 2017 doi:10.1155/2017/1803492
- Noor, M., Ismail, T., Chung, E.S., Shahid, S., Sung, J.H. (2018) Uncertainty in rainfall intensity duration frequency curves of peninsular Malaysia under changing climate scenarios. *Water* 10 (12), 1750
- Noor, M., Ismail, T., Shahid, S., Asaduzzaman, M., Dewan, A. (2020) Evaluating intensity-duration-frequency (IDF) curves of satellite-based precipitation datasets in Peninsular Malaysia. *Atmospheric Research* 105203. DOI: 10.1016/j.atmosres.2020.105203
- Okamoto, K., Ushio, T., Iguchi, T., Takahashi, N., Iwanami, K. (2005) The global satellite mapping of precipitation (GSMaP) project. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '05)*, pp. 3414–3416.
- Paska, J., Lau, A.M.S., Tan, M.L., Tan, K.C. (2017) Evaluation of TRMM 3B42 V7 product on extreme precipitation measurements over Peninsular Malaysia, in *Proceedings of The International Society for Optical Engineering*, 10421, 104210D.
- Pour, S.H., Shahid, S., and Chung, E.S. (2017) A hybrid model for statistical downscaling of daily rainfall *Procedia Engineering* 154, 1424-1430
- Pour, S.H., Wahab, A.K.A., Shahid, S. (2020) Physical-empirical models for prediction of seasonal rainfall extremes of Peninsular Malaysia. *Atmospheric Research* 233, 104720

- Pratama, A.W., Buono, A., Hidayat, R., Harsa, H. (2018). Bias correction of daily satellite precipitation data using genetic algorithm. Paper presented at the IOP Conference Series: Earth and Environmental Science, 149(1) doi:10.1088/1755-1315/149/1/012071
- Reddy, M.V., Mitra, A.K., Momin, I.M., Mitra, A.K., Pai, D.S. (2019) Evaluation and inter-comparison of high-resolution multi-satellite rainfall products over india for the southwest monsoon period. *International Journal of Remote Sensing*, 40(12), 4577-4603. doi:10.1080/01431161.2019.1569786
- Rokach, L. (2016) Decision forest: Twenty years of research. *Information Fusion* 27, 111-125.
- Sa'adi, Z., Shahid, S., Chung, E.S., Ismail, T. (2017) Projection of spatial and temporal changes of rainfall in Sarawak of BorneoIsland using statistical downscaling of CMIP5 models. *Atmospheric Research*, 197:446-460
- Saber, M., Yilmaz, K.K. (2018) Evaluation and bias correction of satellite-based rainfall estimates for modelling flash floods over the mediterranean region: Application to karpuz river basin, turkey. *Water (Switzerland)*, 10(5) doi:10.3390/w10050657
- Shahid, S., Pour, S.H., Wang, X., Shourav, S.A., Minhans, A. Ismail, T.b. (2017) Impacts and adaptation to climate change in Malaysian real estate. *International Journal of Climate Change Strategies and Management*, 9(1): 87-103. <https://doi.org/10.1108/IJCCSM-01-2016-0001>
- Shawky, M., Moussa, A., Hassan, Q. K., El-Sheimy, N. (2019) Performance assessment of sub-daily and daily precipitation estimates derived from GPM and GSMaP products over an arid environment. *Remote Sensing*, 11(23) doi:10.3390/rs11232840
- Shiru, M.S., Shahid, S., Alias, N., Chung, E.S. (2018) Trend analysis of droughts during crop growing seasons of Nigeria. *Sustainability* 10 (3), 871
- Soo, E.Z.X., Jaafar, W.Z.W., Lai, S.H., Islam, T., Srivastava, P. (2019) Evaluation of satellite precipitation products for extreme flood events: case study in Peninsular Malaysia. *Journal of Water and Climate Change*, 10, 871-892.
- Suliman, A.H.A., Awchi, T.A., Al-Mola, M., Shahid, S. (2020) Evaluation of remotely sensed precipitation sources for drought assessment in Semi-Arid Iraq. *Atmospheric Research*, 242, 105007. <https://doi.org/10.1016/j.atmosres.2020.105007>
- Tam, T.H., Abd Rahman, M.Z., Harun, S., Hanapi, M.N., Kaoje, I.U. (2019) Application of Satellite rainfall products for flood inundation modelling in Kelantan River Basin, Malaysia. *Hydrology*, 6, 95.

- Tan, M.L., Ibrahim, A.L., Duan, Z., Cracknell, A.P., Chaplot, V. (2015) Evaluation of Six High-Resolution Satellite and Ground-Based Precipitation Products over Malaysia. *Remote Sens.* 2015, 7, 1504-1528.
- Tan, X., Yong, B., & Ren, L. (2018) Error features of the hourly GSMaP multi-satellite precipitation estimates over nine major basins of china. *Hydrology Research*, 49(3), 761-779. doi:10.2166/nh.2017.263
- Trinh-Tuan, L., Matsumoto, J., Ngo-Duc, T., Nodzu, M. I., Inoue, T. (2019) Evaluation of satellite precipitation products over central vietnam. *Progress in Earth and Planetary Science*, 6(1) doi:10.1186/s40645-019-0297-7
- Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K, Aonashi, K. (2009) A Kalman Filter Approach to the Global Satellite Mapping of Rainfall (GSMaP) from Combined Passive Microwave and Infrared Radiometric Data. *Journal of Meteorological Society of Japan*, 87A, 137-151
- Utsumi, N., Kim, H. (2018) Warm season satellite precipitation biases for different cloud types over western north pacific. *IEEE Geoscience and Remote Sensing Letters*, 15(6), 808-812. doi:10.1109/LGRS.2018.2815590
- Velasquez, P., Messmer, M., Raible, C.C. (2020) A new bias-correction method for precipitation over complex terrain suitable for different climate states: a case study using WRF (version 3.8.1). *Geosci. Model Dev.*, 13, 5007–5027, 2020. <https://doi.org/10.5194/gmd-13-5007-2020>
- Wong, C.L., Venneker, R., Uhlenrook, S., Jamil, A.B.M., Zhou, Y. (2009) Variability of rainfall in Peninsular Malaysia. *Hydrol. Earth Syst. Sci. Discuss.*, 6, 5471–5503.
- Yamamoto, M. K., Shige, S., Yu, C., Cheng, L. (2017) Further improvement of the heavy orographic rainfall retrievals in the GSMaP algorithm for microwave radiometers. *Journal of Applied Meteorology and Climatology*, 56(9), 2607-2619. doi:10.1175/JAMC-D-16-0332.1
- Yang, Z., Hsu, K., Sorooshian, S., Xu, X., Braithwaite, D., Verbist, K. M. J. (2016) Bias adjustment of satellite-based precipitation estimation using gauge observations: A case study in chile. *Journal of Geophysical Research*, 121(8), 3790-3806. doi:10.1002/2015JD024540
- Yeh, N., Chuang, Y., Peng, H., Hsu, K. (2019) Bias adjustment of satellite precipitation estimation using ground-based observation: Mei-yu front case studies in taiwan. *Asia-Pacific Journal of Atmospheric Sciences*, doi:10.1007/s13143-019-00152-7

Yeh, N.C., Chuang, Y.C., Peng, H.S. (2020) Bias Adjustment of Satellite Precipitation Estimation Using Ground-Based Observation: Mei-Yu Front Case Studies in Taiwan. *Asia-Pacific J Atmos Sci* 56, 485–492. <https://doi.org/10.1007/s13143-019-00152-7>

Zeng, S., Yong, B. (2019) Evaluation of the GPM-based IMERG and GSMaP precipitation estimates over the Sichuan region. *Acta Geographica Sinica*, 74(7), 1305-1318. [doi:10.11821/dlxb201907003](https://doi.org/10.11821/dlxb201907003)