Bias Correction Method of High-Resolution Satellite-Based Precipitation Product for Peninsular Malaysia

Zafar Iqbal
Universiti Teknologi Malaysia https://orcid.org/0000-0001-7727-7786
Shamsuddin Shahid (sshahid@utm.my)
Universiti Teknologi Malaysia https://orcid.org/0000-0001-9621-6452
Kamal Ahmed
Universiti Teknologi Malaysia
Xiaojun Wang
Nanjing Hydraulic Research Institute
Tarmizi Ismail
Universiti Teknologi Malaysia
Hamza Farooq Gabriel
National University of Sciences and Technology (NUST)

Research Article

Keywords: Satellite rainfall, bias correction, classifier, machine learning, extreme rainfall.

Posted Date: December 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1141784/v1

Version of Record: A version of this preprint was published at Theoretical and Applied Climatology on March 15th, 2022. See the published version at https://doi.org/10.1007/s00704-022-04007-6.

Bias Correction Method of High-resolution Satellite-based Precipitation Product for Peninsular Malaysia

Zafar Iqbal[1], Shamsuddin Shahid[1*], Kamal Ahmed[1,2], Xiaojun Wang[3,4], Tarmizi Ismail[1], Hamza Farooq Gabriel[5]

[1]School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM) 81310 Johor Bahru, Malaysia

[2]Faculty of Water Resource Management, Lasbela University of Agriculture Water and Marine Sciences (LUAWMS), 90150 Uthal, Balochistan, Pakistan

[3]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, 210029, China;

[4]Research Center for Climate Change, Ministry of Water Resources, Nanjing, 210029, China;

[5]School of Civil and Environmental Engineering (SCEE), National University of Sciences and Technology (NUST), Islamabad, Pakistan

*Corresponding Author Email: sshahid@utm.my

Dr. Shamsuddin Shahid
Associate Professor
Department of Water & Environmental Engineering
School of Civil Engineering, Faculty of Engineering
Universiti Teknologi Malaysia (UTM)
81310 Johor, Malaysia
Room No.: M46-332
Phone: +60 7 5531624 (office)
Mobile: +60 182051586
Website: http://civil.utm.my/sshahid/
Alternative E-mail: sshahid_ait@yahoo.com: sshahid@utm.my

Keywords: Satellite rainfall, bias correction, classifier, machine learning, extreme rainfall.

Bias Correction Method of High-resolution Satellite-based Precipitation Product for Peninsular Malaysia.

Abstract

Satellite-based precipitation (SBP) is emerging as a reliable source for high-resolution rainfall estimates over the globe. However, uncertainty in SBP is still significant, limiting their use without evaluation and often without bias correction. The bias correction of SBP remained a challenge for atmospheric scientists. In this study, the performance of six SBPs, namely, SM2RAIN-ASCAT, IMERG, GsMap, CHIRPS, PERSIANN-CDS and PERSIANN-CSS in replicating observed daily rainfall at 364 stations over Peninsular Malaysia was evaluated. The bias of the most suitable SBP was corrected using a novel machine learning (ML)-based bias-correction method. The proposed bias-correction method consists of an ML classifier to correct the bias in estimating rainfall occurrence and an ML regression model to correct the amount of rainfall during rainfall events. The performance of different widely used ML algorithms for classification and regression were evaluated to select the suitable algorithms. IMERG showed better performance, showing a higher correlation coefficient (R^2) of 0.57 and Kling-Gupta Efficiency (KGE) of 0.5 compared to the other products. The performance of random forest (RF) was better than the k-nearest neighbourhood (KNN) for both classification and regression. RF classified the rainfall events with a skill score of 0.38 and estimated the rainfall amount of a rainfall event with the modified Index of Agreement (md) of 0.56. Comparison of IMERG and bias-corrected IMERG (BIMERG) revealed an average reduction in RMSE by 55% in simulating observed rainfall. The proposed bias correction method performed much better when compared with the conventional bias correction methods such as linear scaling and quantile regression. The BIMERG could reliably replicate the spatial distribution of heavy rainfall events, indicating its potential for hydro-climatic studies like flood and drought monitoring in the study area.

Keywords: Satellite rainfall, bias correction, classifier, machine learning, extreme rainfall.

Significance Statement

The two-stage novel bias correction algorithm was applied to correct the best suitable satellite-based data for peninsular Malaysia to obtain a fine gridded high performing data set for hydrological modelling.

1 Introduction

Estimating rainfall's spatial distribution and temporal variability is vital for any hydrological or climatic study (Iqbal et al., 2020, Ahmed et al., 2017). In-situ observations are the most reliable precipitation data; however, they are insufficient to provide details of spatial rainfall distribution in most regions due to the sparse distribution of rain gauges (Dewan et al., 2019, Nashwan et al., 2019a). Satellite-driven products are emerging as a dependable source of high spatial and temporal resolution rainfall measurement globally (Bhatti et al., 2016, Noor et al., 2019a). Satellite-based precipitation products (SBP) have shown their potential in different hydro-climatic studies such as floods modelling, drought monitoring, water budgeting, and hydrological change assessment (Xie et al., 2011, Bitew et al., 2011, Yong et al., 2010, Nashwan et al., 2020).

Higher spatial resolution rainfall is vital for capturing the spatially high heterogeneous rainfall in the tropics (Tan et al., 2015). Highly localized intense rainfall is a common phenomenon in the tropical region due to higher convective activities (Sa'adi et al., 2020), which shares about 70% of the total rainfall (Badron et al., 2015). Such rainfalls are usually short, intense and have a smaller areal extent (Zafar et al., 2004). The convective events mostly have a cell diameter of <10 km, which is much lower than the density of rain gauges in most regions of the globe (Schroeer et al., 2018). Therefore, most rainfall activities in the tropical region cannot be detected using the existing rainfall monitoring network. It has been reported that rising temperature would enhance convective moisture convergence (Ahmed et al., 2015), which eventually increases the amount of convective rainfall and decreases their spatial extents (Wasko et al. Therefore, much higher spatial resolution rainfall data will be required for tropical rainfall analysis in the near future.

SBP are generally passive microwave (PMW) and infrared radiance (IR) precipitation retrievals. Several SBP datasets have been developed after the success of the Tropical Rainfall Measuring Mission (TRMM). The newly developed SBP products such as Integrated Multi-satellite Retrievals for global precipitation measurement (GPM) (IMERG) (Huffman et al., 2015), Goddard profiling algorithm (GPROF) (Kummerow et al., 2015) and SM2RAIN-ASCAT (Brocca et al., 2019) showed impressive rainfall estimation capability over various regions (Suliman et al., 2020).

However, SBP is an indirect estimation of precipitation, and the precision of SBP depends on the sensor used and the retrieval algorithm applied (Hsu et al., 1997). The capability of the sensors and retrieval algorithms varies widely with geography and climate. For example, many SBP products overestimate precipitation over the desert and underestimate precipitation over the forest. The bias in SBP also varies significantly between arid and tropical climate zones (Nashwan et al., 2019b, Ushio et al., 2009). Moreover, topography and cloud type add different types of bias in SBP (Sun et al., 2018, Serrat-Capdevila et al., 2016).

Biases in SBP in the tropical maritime continent in Southeast Asia is much more complex. Rainfall in this maritime continental region is defined by a complex interaction of ocean, irregular land-ocean interface and highly variable topography. Accurate estimation of such a complex rainfall process using the existing satellite sensors and retrieval algorithms is often impossible. Therefore, the most appropriate SBP selection for such a region and the correction of biases in SBP is important before their application in any hydro-climatological analysis (Katiraie-Boroujerdy et al., 2020).

Several studies have been conducted to assess the performance of SBP products over the maritime continent of Southeast Asia. Peña-Arancibia et al. (2013) evaluated three reanalyses and three satellite-based products. The study found that the reanalysis product performed well, except for the months affected by the Asia-Pacific monsoon. They also reported that satellite-based products are better in estimating convective rainfall in the South East Asian region. The evaluation of nine satellite products by Rauniyar et al. (2017) over the maritime continent revealed inconsistent results over mountain, ocean and coastal regions with an overall under or overestimation. Existing literature reported inconsistency in the performance of various SBP products, such as Tan et al. (2015a) found TRMM to perform better in Malaysia, while Soo et al. (2020) showed CMORPH as the best SBP product. Soo et al. (2019) found GSMap-NRT a better product in simulating the flow of the Kelantan River, Malaysia. Semire et al. (2012) reported that among the three TRMM products (3B42 V6, 3B43 V6, 3A12 V6), 3B43 V6 showed a better correlation with the observed data in Peninsular Malaysia. A comparative study made by Tan et al. (2018) found IMERG real-time has better performance than other SBPs in Peninsular Malaysia. Tan et al. (2017) found IMERG with the least systematic bias in detecting daily precipitation in Singapore. Overall, the review of the performance of SBP in Peninsular Malaysia indicates IMERG as the most reliable for rainfall

estimation, though it has a significant bias. Recently, a new SBP SM2RAIN-ASCAT showed its potential in rainfall estimation in different regions globally (Gupta et al., 2019, Paredes-Trejo et al., 2018). However, the capability of SM2RAIN-ASCAT in the tropical maritime continental region has not been evaluated yet along with its performance as compared to other products such as CHIRPS, PERSIANN andGsMap.

Though a large bias in the most suitable SBP in Malaysia (IMERG) has been reported, no attempt has been made to correct the bias in SPBs in Peninsular Malaysia. Several efforts have been made to correct SBP biases in other parts of the world using different methods, including linear regression (Yang et al., 2016, Alharbi, 2019), distribution function matching (Mastrantonas et al., 2019), mean bias correction (Hashemi et al., 2017, Chaudhary et al., 2019), distribution mapping (Katiraie-Boroujerdy et al., 2020) and Bayesian algorithm (Ma et al., 2018). Recently, machine learning algorithms have also been introduced for satellite precipitation bias correction. Pratama et al. (2018) combined genetic algorithm with a power transformation method for satellite precipitation bias correction. Le et al. (2020) used a neural network to correct satellite precipitation bias in the Mekong River basin. Studies revealed improvement in SBP performance after bias correction. However, significant bias still exists in replicating different rainfall extremes such as consecutive dry days and extreme rainfall amounts, which are most important for estimating dry spells and floods. Therefore, there is a need for a better bias correction technique of SBP.

The objective of the present study are: (1) to compare the performance of a recently available SBP, namely SM2RAIN-ASCAT, IMERG, GsMap, CHIRPS, PERSIANN-CDS and PERSIANN-CSS (2) to reduce bias in the best performing SBP using a two-stage bias correction approach consists of a classifier and a regressor. The novelty of the present study is a machine learning-based double bias-correction approach for bias correction of satellite rainfall. The bias-corrected IMERG data generated in this study can be used for hydrological and metrological studies at fine resolution. The bias adjustment methodology is developed in two main steps as classification (rain/no rain) and regression on rainy days.

2 Study area and Data

2.1 Geography of Peninsular Malaysia

The methodology opted in the paper was applied to Peninsular Malaysia, which is situated in South East Asia with a latitude 1.20° to 6.40° N and longitude 99.35° to 104.20° E (Noor et al., 2019a). Situated near the equator, the climate of Malaysia is humid and hot. The Rainforest climate of the region is highly influenced by the Asian–Australian atmospheric dynamics along with land-sea interaction, varying topography and monsoon winds (Webster et al., 1998). The average daily temperature varies between 21 and 32º C, with an average annual variation of 3ºC. The annual average rainfall is about 2000-4000 mm, with an average of 150 to 200 rainy days per year (Tan et al., 2014, Noor et al., 2019b). The distribution pattern of precipitation in the region is established with the integrated response of local topography and wind flow direction.

Peninsular Malaysia has two seasons throughout the year, i.e. Southwest Monsoon (SWM), which prevails from May to August, and Northeast Monsoon (NEM) exists between November and February. Extreme rainfall events are generally observed during NEM, whereas the weather is dry during SWM. Coastal areas are under the influence of NEM, whereas higher altitudes have less influence on the monsoon. Peninsular Malaysia represents the humid weather with recorded maximum precipitation during the 'inter-monsoon period'. Rain gauge density has been reported to be the major source of uncertainty in evaluating any rainfall product. Better satellite precipitation evaluation is possible with dense observation (Gadelha et al., 2019). World Meteorological Organization (WMO) recommended one station per 575 km^2 as the optimum threshold of rain-gauge density (WMO, 1994). In this study, 364 stations were used with a gauging density of 363 km^2

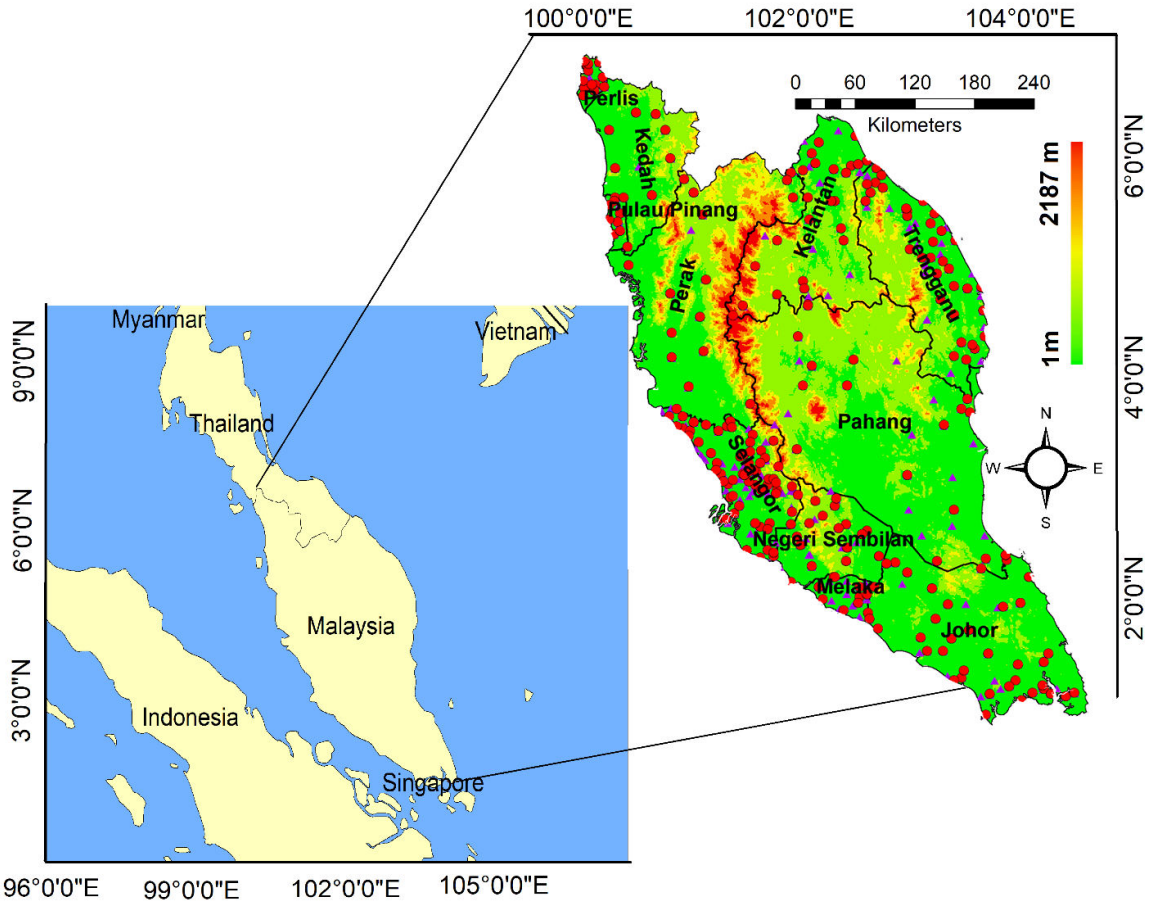

Fig. 1 Location of Study area showing the rainfall stations. Red Circles shows the observed stations used for calibration while the purple triangle shows the stations used for validation.

2.2 Data and sources

The data recorded at 432 rainfall gauges were acquired from the Department of Irrigation and Drainage (DID) Malaysia. The stations having less than 10% missing values (total 364 stations) were only considered in this study. The location of the observed station is shown in Fig 1. Inverse distance weighting (IDW) method was used to interpolate all the SBP rainfall at the selected 364 stations to evaluate their performance. IDW considers the influence of neighbouring points

according to their distance from the station location. It provides better interpolation when densely gauged data is available.

SM2RAIN-ASCAT is a global-scale precipitation product generated from the European Space Agency (ESA) Climate Change Initiative (CCI) project soil moisture data (Brocca et al. (2014). SM2RAIN-ASCAT is available on a daily scale at a spatial resolution of 12.5 km for the period of 2007–2019 (Ciabatta et al., 2018, Liu et al., 2011, Liu et al., 2012). Brocca et al. (2014) estimated rainfall through inversion of a soil moisture estimation equation,

$$p(t) = Z^{*}ds(t)/dt + as(t)^{b} \quad (1)$$

where, *p(t)* represents the computed precipitation, *Z** is the soil moisture capacity, *s(t)* shows the soil saturation at a time, *t*, and *a* and *b* are the parameters showing the relation between drainage and soil saturation, which are calculated following a calibration procedure. The algorithm showed accurate results for global and regional scales (Ciabatta et al., 2015, Ciabatta et al., 2017, Abera et al., 2016, Brocca et al., 2014). The assumption used is surface flow is only possible when rainfall completely saturates the soil. The data sets were obtained from (https://doi.org/10.5281/zenodo.3635932).

IMERG provides different versions of multi-satellite precipitation data. In this study, IMERG Version 06 Level 3 daily precipitation data of 10 km resolution was employed. The dataset is developed by NASA GES data and Information Services Centre (DISC) from the half-hourly precipitation data of GPM_3IMERGHH by summing the daily precipitation with a latency of 2-3 months. The data were obtained from the GPM website (http://pmm.nasa.gov/data-access/downloads/gpm/). Three types of products are generated from the IMERG algorithms (Early, Late and Final run). GPCC (Global Precipitation Climatology Centre) data are infused with the IMERG to produce the IMERG final run used in this study (Huffman et al., 2020).

GsMap is a multi-satellite product having global coverage developed under the Global precipitation measurement (GPM) mission. The global precipitation map has been developed using GPM core satellite data, Dual-frequency Precipitation Radar (DPR) and other GPM constellation satellites. The GsMap have two products which are GsMap NRT (Near Real-Time) and GsMap GC

(Gauge Calibrate). The former product is the result of integration between the output of passive microwaves radiometers and infrared images. Whereas, the latter is a bias-adjusted product using the NOAA CPC data to bias correct the GsMap-NRT by an algorithm developed by (Mega et al., 2019). In this study, GsMap gauge calibrated product was used having a spatial resolution of 0.1^{o} x 0.1^{o}. The data is freely available at https://sharaku.eorc.jaxa.jp/GSMaP/index.htm.

Precipitation Estimation from Remotely Sensed Information using Artificial Neural Netwroks (PERSIANN) developed by applying artificial neural network (ANN) on the infrared brightness temperature images of geostationary satellites by Center for Hydrometeorology and Remote sensing (CHRS) the University of California, Irvine (UCI). PERSIANN-CSS uses the threshold cloud segmentation algorithm to separate and classify the cloud patches whereas PERSIANN-CDR is developed using GridSat-B1 infrared data and bias-adjusted using GPCP product. PERSIANN-CCS have a spatial resolution of $0.04^{o}\times0.04^{o}$ whereas PERSIANN-CDS have a resolution of $0.25^{o}\times0.25^{o}$ (Nguyen et al., 2018). These two products were obtained from https://chrsdata.eng.uci.edu/.

Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) developed to support the United States Agency for International Development Famine Early Warning Systems Network (FEWS NET) has a spatial resolution of 0.25^{o} and 0.05^{o} (Funk et al., 2015). In this study, the CHIRPS data having 0.05 resolution was used to evaluate the performance of SBP.

3 Methodology

The Satellite products were evaluated using categorical and continuous indices to assess their relative performance in replicating spatial and temporal variability of observed rainfall of Peninsular Malaysia. A two-stage ML-based bias correction method is proposed in this study to correct the bias of the best SBP product against the observed rainfall to improve its performance. The methods used in this study are elaborated in the following subsections

3.1 Measuring Data Performance

The performance of SBP products was evaluated employing four statistical metrics, coefficient of determination (R^2), percentage of bias (Pbias), and normalized root mean square error (NRMSE), modified Index of Agreement (md) and Kling Gupta Efficiency (KGE) as given below:

$$NRMSE = \frac{\left[\frac{1}{N}\sum_{i=1}^{N}\left(x_{sat,i} - x_{obs,i}\right)^2\right]^{1/2}}{x_{max} - x_{min}} \quad (2)$$

$$PBIAS = 100\ \frac{\sum_{i=1}^{N}\left(x_{sat,i} - x_{obs,i}\right)}{\sum_{i=1}^{N} x_{obs,i}} \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^{N}\left(x_{obs,i} - \overline{x_{obs}}\right)\left(x_{sat,i} - \overline{x_{sat}}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_{sat,i} - \overline{x_{sat}}\right)^2 \sum_{i=1}^{N}\left(x_{obs,i} - \overline{x_{obs}}\right)^2}} \quad (4)$$

$$KGE = 1 - \sqrt{(\mathrm{r} - 1)^2 + \left(\frac{\sigma\mathrm{sat}}{\sigma\mathrm{obs}} - 1\right)^2 + \left(\frac{\overline{x_{sat}}}{\overline{x_{obs}}} - 1\right)^2} \quad (5)$$

$$md = 1 - \frac{\sum_{i=1}^{N}\left(x_{obs,i} - x_{sat,i}\right)^j}{\sum_{i=1}^{N}\left(\left|\left(x_{sat,i} - \overline{x_{obs}}\right)\right| + \left|\left(x_{obs,i} - \overline{x_{obs}}\right)\right|\right)^j} \quad (6)$$

where, $x_{obs,i}$ and $x_{sat,i}$ are the i_{th} observed and satellite data; $\overline{x_{obs}}$ and $\overline{x_{sat}}$ are mean observed and mean satellite data, N is the sample size equal to the daily observations whereas σ_{sat} and σ_{obs} are the standard deviation of the satellite and observed data where j is an arbitrary power (positive integer) used to calculate md in this case we use j = 1. The values of NRMSE and Pbias near to 0 and R^2, md and KGE near to 1 represent better matching with the observed data.

Five categorical metrics, namely Hit rate (HR), Heidke Skill Score (HSS), Gerrity Skill Score (GSS), Hit Bias (HB) and Pierce Skill Score (PSS), was used to measure the capability of

SPB to identify rainfall/no rainfall days. The contingency table used for estimating categorical matrices is given in Table 1. The explanation of the joint distribution of the output of the contingency table is given in Table 2.

Table 1 The contingency table for calculation of categorical matrices

		Observed	
		> 0	= 0
IMERG	> 0	Hits	False alarm
	= 0	misses	Correct Negatives

Table 2 The joint distribution of the categorical matrices output

Term	**Description**
Hit	Event forecasted to happen, and it happened
Misses	Event forecasted not to occur, but it occurred
False alarm	Event forecast but it not occurred
Correct Negative (CN)	The forecasted not to occur, and it not occurred

The equation used to calculate the metrics are as below:

$$HR = \frac{Hits}{Misses + Hits} \tag{7}$$

$$HSS = \frac{2\,(Hits\ \times CN\ -FA\ \times Misses\,)}{[(Hits + Misses)(Misses + CN) + (Hits + FA)(False\ Alarm + CN]} \tag{8}$$

$$GSS\ = \sum_{i=3}^{4}\sum_{j=1}^{4} p_{ij}s_{ij} \tag{9}$$

$$HB = \frac{Hits + False\ Alarms}{Hits + Misses} \tag{10}$$

$$PSS = \frac{(Hits\ \times CN) - (FA\ \times Misses)}{(FA + CN) + (Hits + Misses)} \tag{11}$$

where FA means false alarm, CN indicates correct negative, and s_{ij} is the scoring matrix. The HR, GSS and PSS range between 0 and 1, and HB from 0 to ∞, where 1 represents a perfect forecast (Gerrity Jr, 1992). The HSS ranges from -1 to +1 where a value closer to +1 shows better performance (Heidke, 1926).

3.2 Bias Correction

A double correction approach is introduced in this study, where correction for zero rainfall days was first done using a classifier and then rainfall amount in rainy days were corrected using regression. Classified values of the best SBP rainfall based on rainfall/no rainfall was considered as input, and observed rainfall/no rainfall was used as an output to develop a classification model. The rainfall values classified as rainy days were then corrected using a regression model. Individual model was developed for each month to consider the seasonal rainfall variability. Rainfall data of 70% of stations or 255 stations were randomly selected for model calibration, while the model performance was evaluated at the remaining 109 stations. Rainfall data of all days of a month at all the stations, considered for calibration, were merged for model calibration. The validation was also performed in each station separately.

This study evaluated the performance of two Machine Learning algorithms, random forest (RF) and k-nearest neighbourhood (KNN), to select the best classifier. The performance of two regression algorithms, RF and artificial neural network (ANN), was evaluated to find the best method to predict rainfall amount on rainy days. RF and KNN were chosen as classifiers as those have been reported as the most suitable classifier among many others (Fernández-Delgado et al., 2014). Two classical ML algorithms (RF and ANN) were used due to their ability to simulate complex nonlinear relationships effectively (Pour et al., 2020, Sa'adi et al., 2017). The recent studies which used RF regression model to correct SBP (King et al., 2020, Beck et al., 2020) found it very effective in correcting the biases. They applied RF model on SNOw Data Assimilation System (SNODAS) and found an improvement of 86% in RMSE. The ML methods are described in the following sections. Details of the algorithms can be found in the literature cited in their description.

3.2.1 Random Forest

RF (Breiman, 2001) is amongst the most effective ML algorithms for predictive modelling. Rigorous improvements have made it more applicable in various research fields and improved its significance in ML powerhouse (Fawagreh et al., 2014). It is a flexible algorithm that can be used for both classification and regression. It creates randomness in decision trees using bootstrapping for sampling, which helps RF to be more insensitive to overfitting (Heung et al., 2014). The RF regression model is a kind of additive model that predicts the decision from the sequence of basic models shown in Eq (12).

$$\hat{f} = \frac{1}{N}\sum_{i=1}^{N} f_i(\acute{x}) \tag{12}$$

Given a training set for bootstrap aggregating, X= $x_1, x_2, x_3 \ldots\ldots x_n$ having a response of Y = $y_1, y_2, y_3 \ldots\ldots y_n$. These samples are bagged N times to find fitting trees to these training sets. Where $f_i(\acute{x})$ represents the random forest trees and $\hat{f}$ is the prediction of an unknown point x' using the bootstrapped data.

3.2.2 K-Nearest Neighbour

The KNN is an efficient nonparametric classification algorithm that assigns data to a class based on its nearest neighbours (Huang et al., 2017). In the particular classification problem, assuming that $T = \{x_n \in R^d\}_{n=1}^{N}$ indicates a training set comprises of N samples within each M class in d-dimension; the sample x_n is assigned the class mark "c_n", the distance between the unknown point x and x_i^{NN} is estimated using Euclidean distance method,

$$d(x, x_i^{NN}) = \sqrt{(x - x_i^{NN})^T (x - x_i^{NN})} \quad (13)$$

Next, the class name of the query point x is estimated based on the majority voting of its neighbours,

$$\acute{c} = \arg max_c \sum_{(x_i^{NN}, c_i^{NN}) \in \bar{T}} \delta(c = c_i^{NN}) \quad (14)$$

where c is a class label and c_i^{NN} is the class label of i-th nearest neighbour. $\delta(c = c_i^{NN})$, an indicator function, can have a value of one of the class c_i^{NN} of the neighbour x_i^{NN}.

3.3 Artificial Neural Network

ANN is one of the most prevalent ML algorithms which has been extensively studied and commonly used in many fields. The ANN used in this study is a multilayer feed-forward network with the backpropagation learning algorithm (Sivapragasam et al., 2010). ANN generally consists of an input, a hidden and an output layer (Urbanczik, 1996). The inputs (IMERG rainfall) through the input layer are fed into the ANN. In ANN, the difficult task is to assign the number of hidden layers. Generally, number of layers are selected based on trial and error processes (Ghaffari et al., 2006, Shankar et al., 2007). The best performance in this study was found for an ANN with a single hidden layer. The hidden layer blends weights and bias with the inputs by employing activation functions to generates output. A single input is used in this study (x) which is blended using a vector of weight w_j to determine the simulated output, y_j using Eq. 15.

$$y_j = f\left(x.w_j - b_j\right) \quad (15)$$

where, $f(.)$ is the activation function, w_j is the weight vector, b_j is the bias at node j. $y_j$is the only input at node j to the hidden layer, or it acts as output in the case of a single layer, and j represents the number of nodes. Each neuron has a transfer function to represent the internal activation level of a neuron. Various transfer functions such as hyperbolic, tangent linear and sigmoidal are used for different relationships (Shankar et al., 2007).
The generalized form of a sigmoidal transfer function is given in Eq 15.

$$z_j = \frac{1}{1 + e^{-y_y}} \tag{15}$$

In this equation $z_j$represents the output at node j. In a multilayer neural network, the weight w_j is assigned by the backpropagation method at each node to determine the unknown data correctly.

4 Results

4.1 Evaluation of Satellite Data

The spatial distribution of annual average rainfall estimated using observed and SBPs are presented in Fig. 2. Observed and all SBP were interpolated to a resolution of 0.1°×0.1° using IDW method to generate a spatial distribution map of observed rainfall in Fig 2. Both IMERG and SM2RAIN-ASCAT indicated the high rainfall regions in the northeast coastal region and low rainfall zone in the central part of the peninsula whereas CHIRPS and PERSIANN-CSS are showing overestimation in the overall study area and GsMap showed relatively higher rainfall in the northeast coastal region. However, the comparison of the maps revealed the better capability of IMERG to replicate the spatial distribution of annual average rainfall. The assessment of map similarity using the correlation analysis showed a high R^2 of IMERG with observed rainfall (0.56) compared to SM2RAIN-ASCAT (0.15), GsMap (0.18), PERSIANN-CDR (0.14), PERSIANN-CSS (0.10) and CHIRPS (0.13) with a significance level > 0.001.

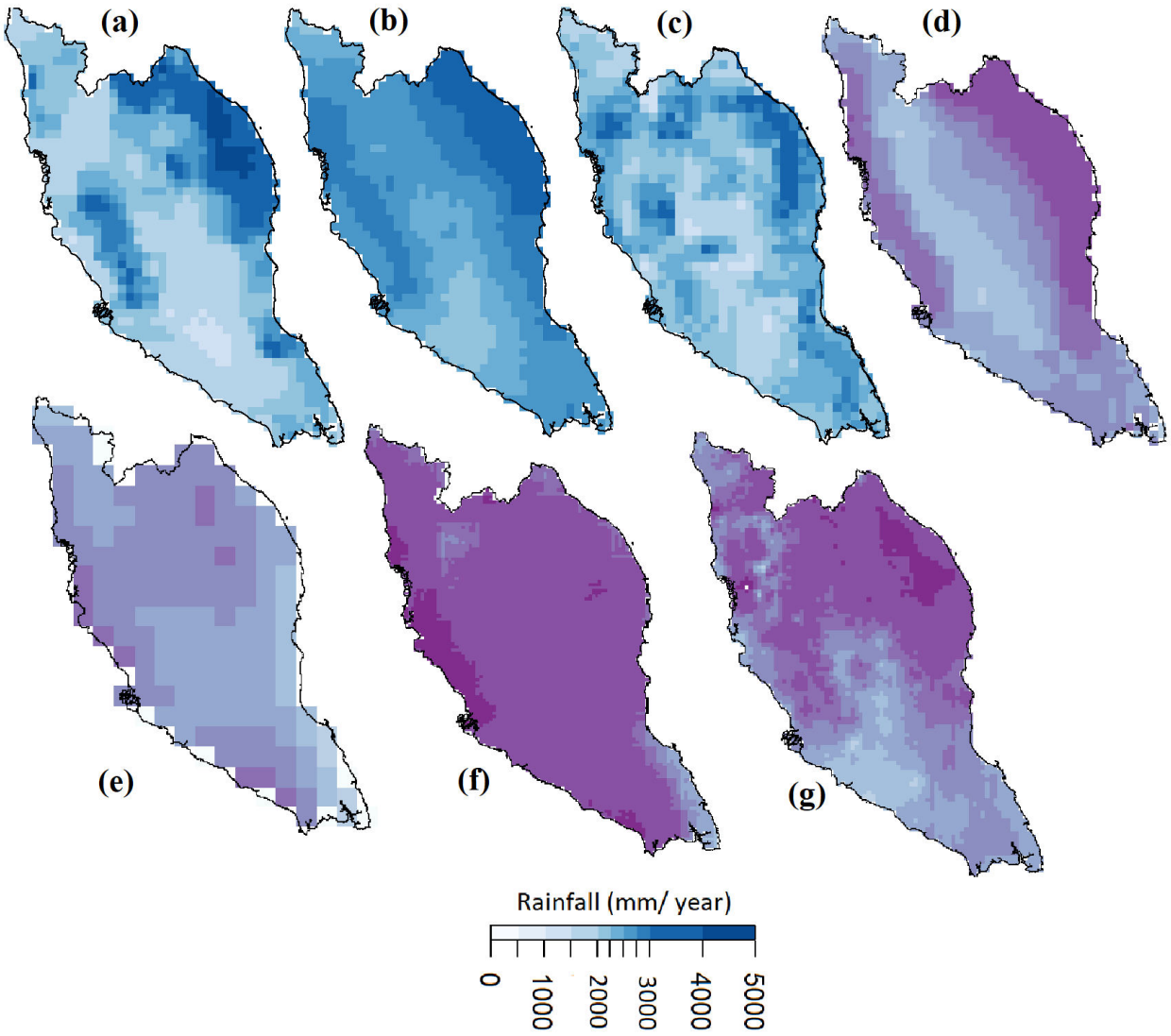

Fig. 2 Annual average rainfall estimated using (a) observed (b) IMERG (c) SM2RAIN-ASCAT (d) GsMap (e) PERSIANN-CDR (f) PERSIANN-CSS and (g) CHIRPS.

The rainfall time series of six SBPs were also compared with the observed rainfall time series at each station for the period 2007–2019 to show their relative performance. Four statistics metrics were used for this purpose. Results are shown using box-whisker diagrams in Fig 3. The box-whisker diagrams in the figure were prepared using the statistical metrics estimated at all the

364 stations, where the horizontal line inside the box indicates a median value of a metric while the height of the box represents an inter-quartile range of the metric.

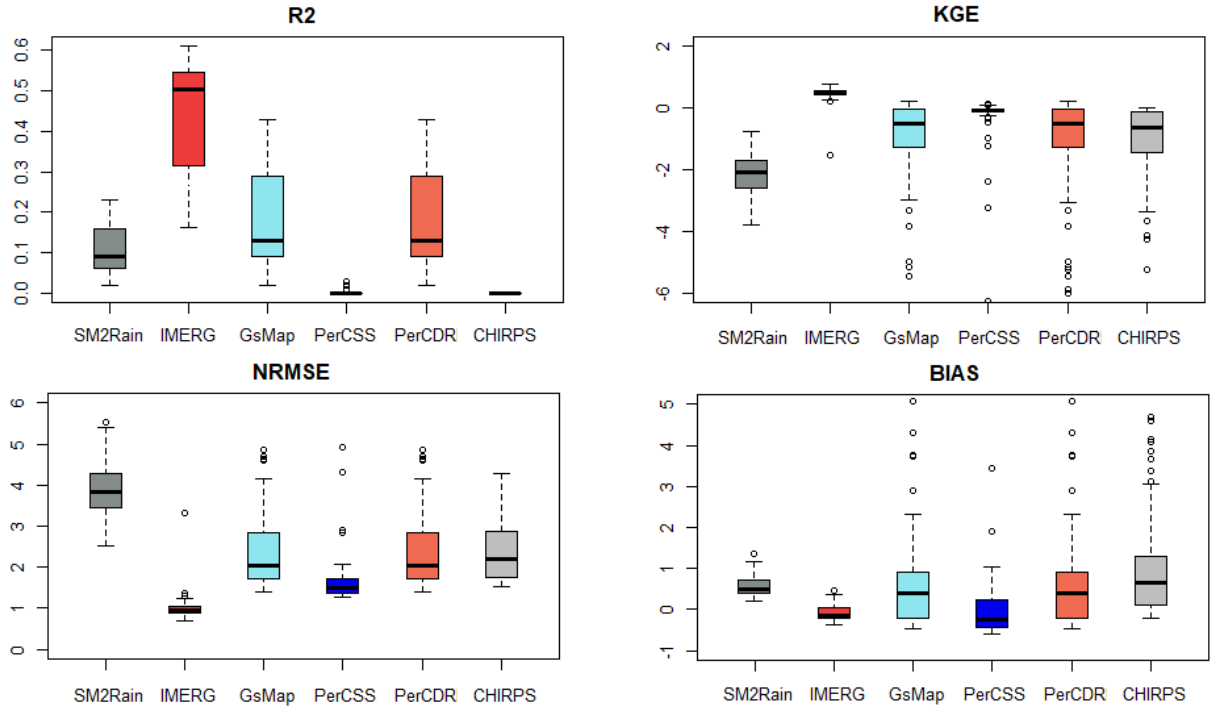

Fig. 3 Boxplot showing the relative performance of all six SBP in term of (1) Mean Square Error; (2) percentage of bias; (3) coefficient of determination; and (4) Kling–Gupta efficiency.

The results showed that the lowest median NRMSE value was achieved by IMERG (0.9) and the highest value (3.8) by SM2RAIN-ASCAT, indicating that the absolute error for IMERG is much less than SM2RAIN-ASCAT and all other products. IMERG also showed a good correlation with the observed station data as compared to the other 5 products. The R^2 value of IMERG was found 0.57 compared to 0.1 for SM2RAIN-ASCAT, 0.15 GsMap, 0.05 PERSIANN-CSS, 0.19 PERSIANN-CDR and 0.04 CHIRPS, while KGE of IMERG was 0.5 compared to -2.2 for SM2RAIN-ASCAT when calculated against the observed rainfall. The metrics show a better capacity of IMERG compared to SM2RAIN-ASCAT, CHIRPS, GsMap, PERSIANN_CDS and PERSIANN-CDR in estimating daily rainfall in Peninsular Malaysia.

Fig. 4 shows the probability distribution function (PDF) of areal average monthly observed and satellite rainfall for the period 2007-2019. The results showed a better match of observed PDF with the IMERG PDF as compared to the other SBP. PERSIANN-CDS was found to perform better after IMERG whereas SM2Rain was the worst in replicating the rainfall probability as compared to the observed data. The IMERG showed a bit higher mean than the observed, but it replicated the observed rainfall variability reliably.

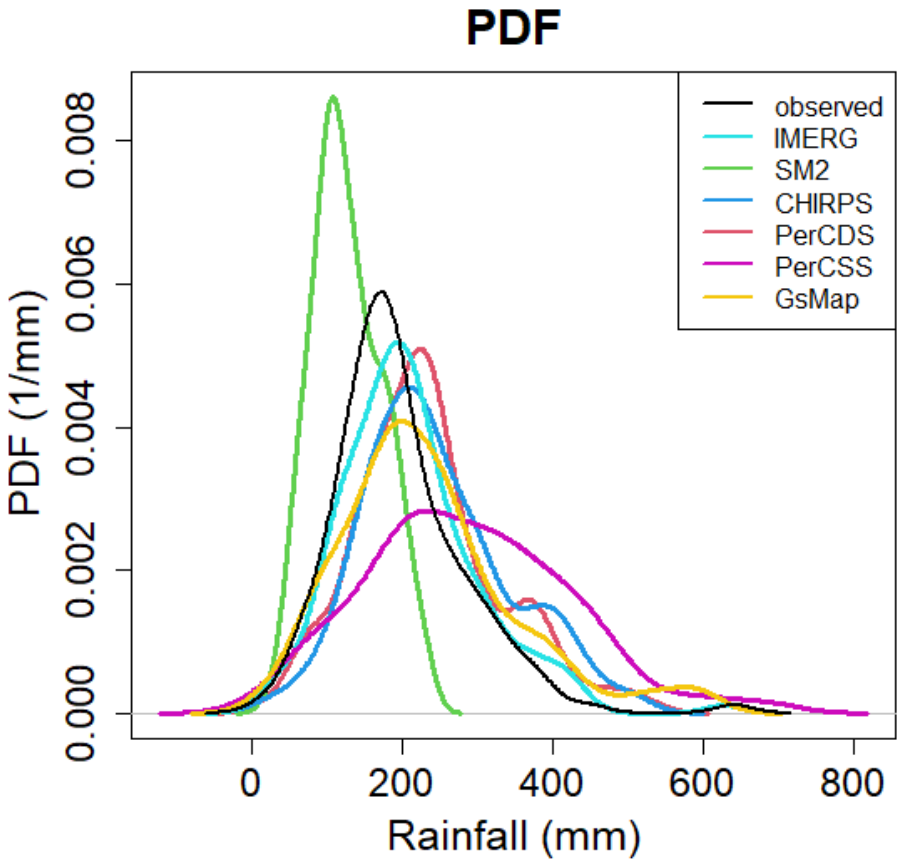

Fig. 4 Probability Distribution function for Observed, IMERG, SM2RAIN-ASCAT, CHRIPS, PERSIANN-CDS, PERSIANN-CSS and GsMap.

4.2 Bias correction

Though IMERG better replicated the gauged rainfall according to different statistics, a significant bias was still in IMERG rainfall. The percentage of bias in IMERG was in the range of -0.39 to 0.64 with a median of -0.16, as shown in Fig 3. This indicates the need for improvement of the performance of IMERG rainfall. The performance of the proposed bias correction technique in detecting rainfall/no rainfall days and estimating the amount of rainfall in rainy days are discussed in the following subsections.

4.2.1 Performance of classifiers

Classifiers were developed to correct the number of rainfall days in IMERG data. RF and KNN were applied to SBP to correct the zero rainfall days based on observed data. The zero rainfall corrected data were compared with the observed data using the categorical indices at all the stations used for model validation. The IMERG rainfall days before and after classification using KNN and RF are presented using boxplots in Fig. 5. The results revealed that both the classification algorithms increased the performance of IMERG in terms of all statistics. The performance comparison of RF and KNN revealed the better performance of RF compared to KNN. For example, HB of IMERG, IMERG(RF) and IMERG(KNN) were 1.20, 1.1 and 1.15, respectively, and the PSS were 0.31, 0.38 and 0.35, respectively. It means that the prediction of rainfall days was more accurately using RF compared to KNN.

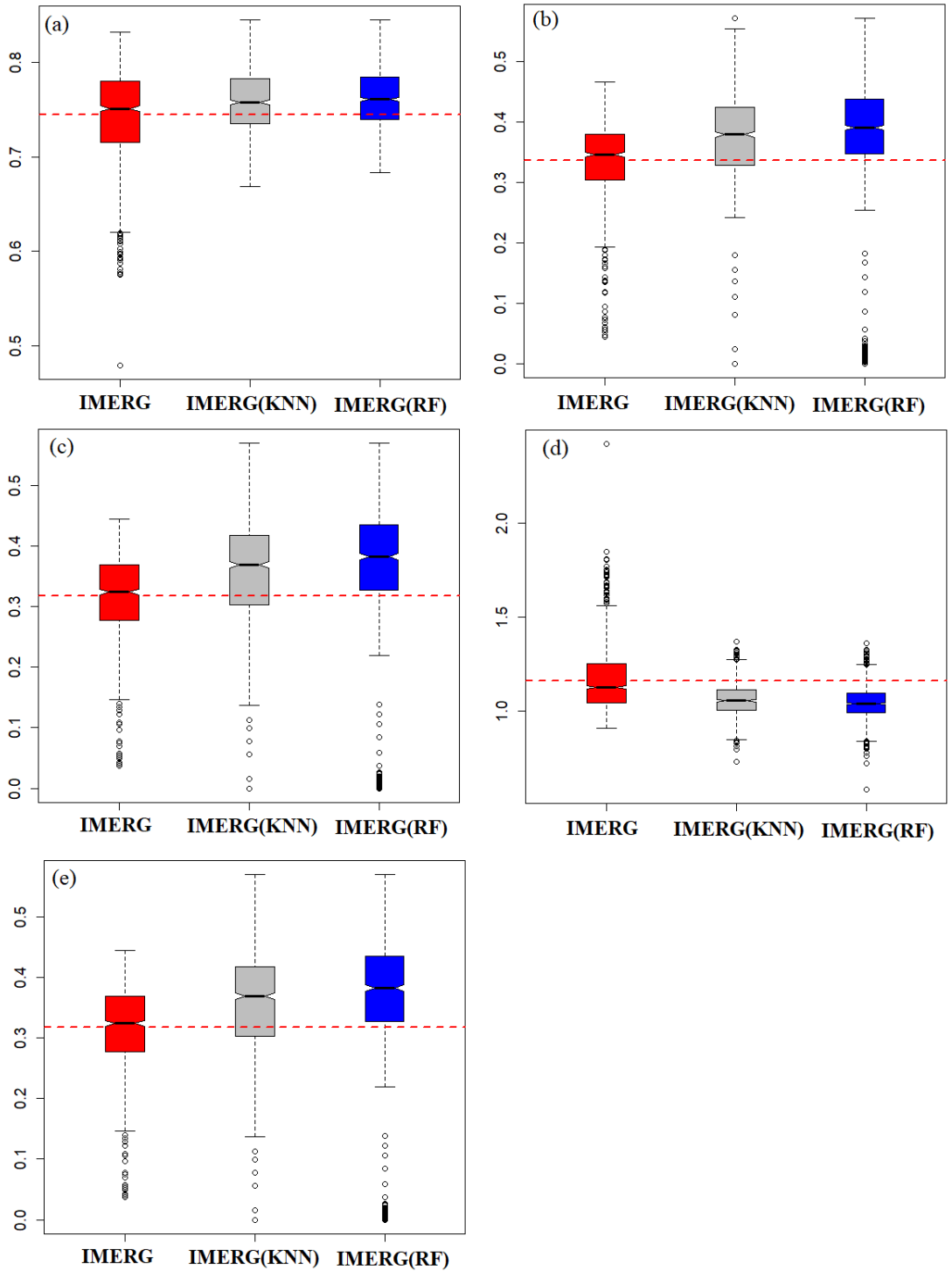

Fig. 5 Comparison IMERG rainfall with KNN and RF corrected data in estimation of rainfall events (a) Hit Rate (b) Heidke Skill Score (c) Gerrity Score (d) Bias and (e) Pierce Skill Score.

4.2.2 Performance of regression models

Regression models were developed to predict the rainfall amount on rainy days. Models were developed using both ANN and RF. The model's performance in estimating rainfall amount in rainfall days at all the stations used for model validation are presented using a boxplot in Fig 6. The performance of RF and ANN regression models was evaluated using different statistical indices to determine the most suitable model. The relative performance of RF and ANN clearly shows the superiority of RF in estimating rainfall amount on rainy days. The NRMSE, R^2, PBIAS and md for RF estimated rainfall was 1.4, 0.34, 0.02, and 0.56 compared to 2.5, 0.13, -0.01 and 0.43, respectively, for ANN. Therefore, RF was used as the regression model for estimating rainfall in the proposed bias removal technique. The estimated rainfall by RF model in rainfall days were merged with classifier output to generate the entire time series.

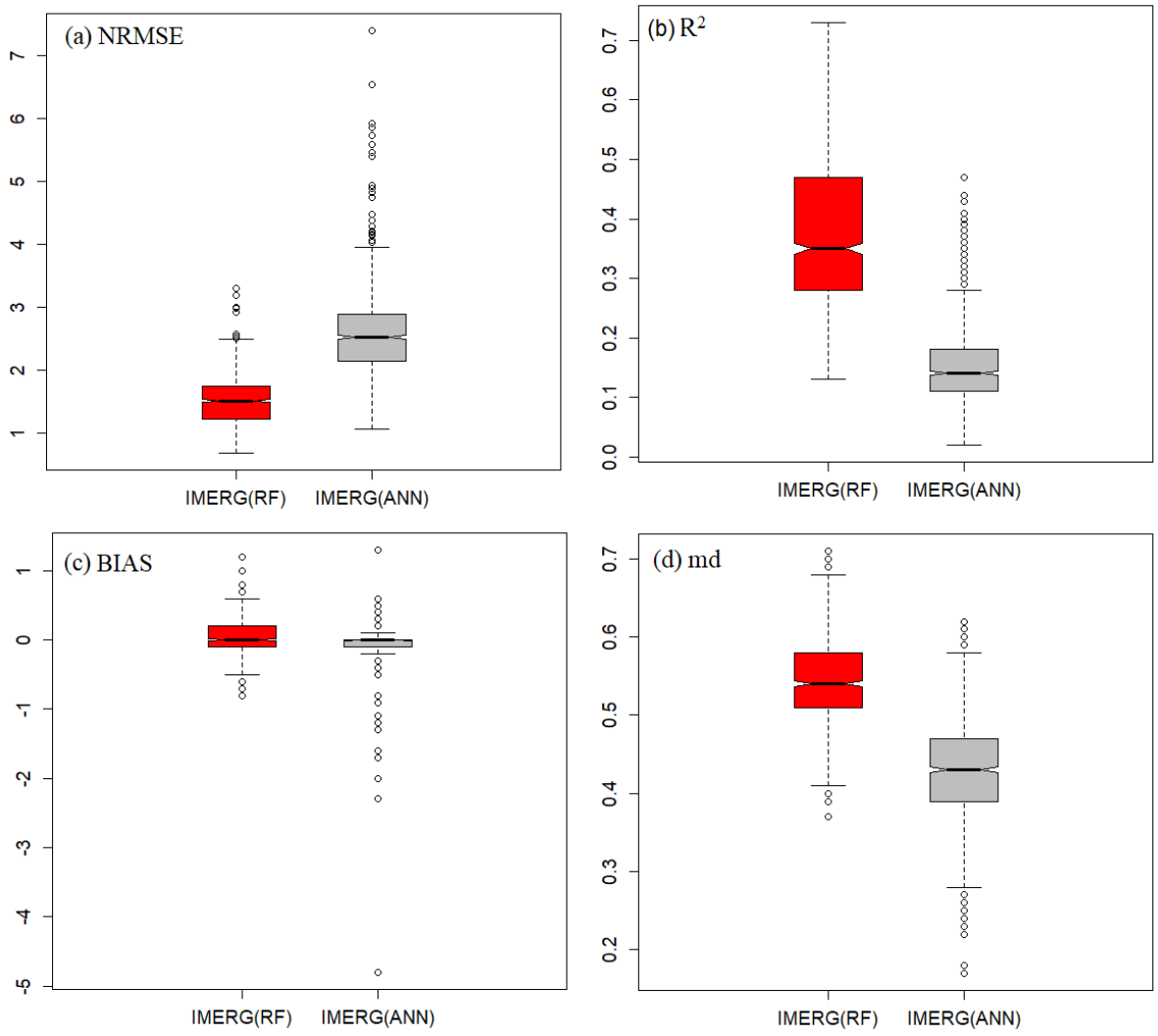

Fig. 6 Performance of regression algorithms in the correction of IMERG rainfall amount on rainy days

4.3 Comparison of performance with the conventional model

The performance of newly developed bias correction was compared with two widely used conventional bias correction methods, linear scaling (LS) and quantile regression (QR). The relative performance of the methods was estimated using two categorical and two continuous indices and presented in Fig 7. The newly developed method showed higher performance than the conventional

methods in correcting SBP bias. The RF reduced RMSE in rainfall by 55% compared to 20 % by LS and 24% by QR, indicating about 125% better performance of the newly developed model than LS compared to the IMERG.

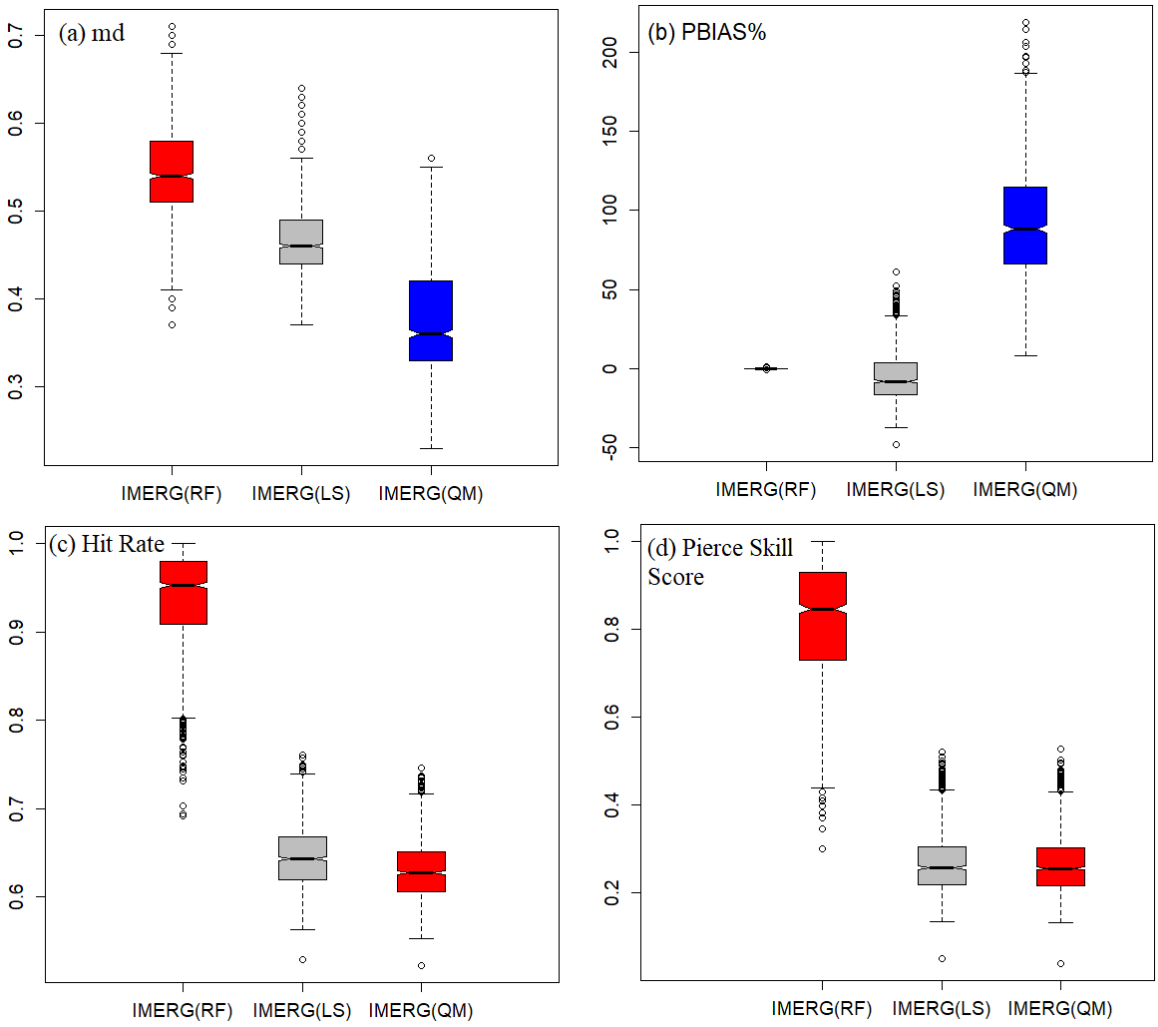

Fig. 7 Comparison of newly developed bias-correction method with conventional methods in correcting bias in IMERG rainfall

4.4 Performance in rainfall amount estimation

The corrected IMERG data were compared with observed data to estimate PBIAS and R^2 for each calendar day. Obtained results are presented in Fig 8. The metrics were estimated using

observed with IMERG and BIMERG data at all the locations utilized for validation. The PBIAS and R^2 for each calendar day were calculated for all grids in the study area. PBIAS and R^2 values of each calendar day are plotted in Fig 8. The blue lines in the figure represent the bias-corrected IMERG and the orange lines represent the IMERG before correction. The results revealed a large reduction of bias and a significant improvement in correlation after bias correction. The mean bias of IMERG for all the days considered for analysis was 158%, which was reduced to -13.3% in BIMERG. The mean value of the R^2 for IMERG was 0.03 which was increased to 0.24 after correction using RF regression model. The R^2 was also found significant at $p<0.05$. Therefore, BIMERG can be recommended for hydro-climatic studies in Peninsular Malaysia.

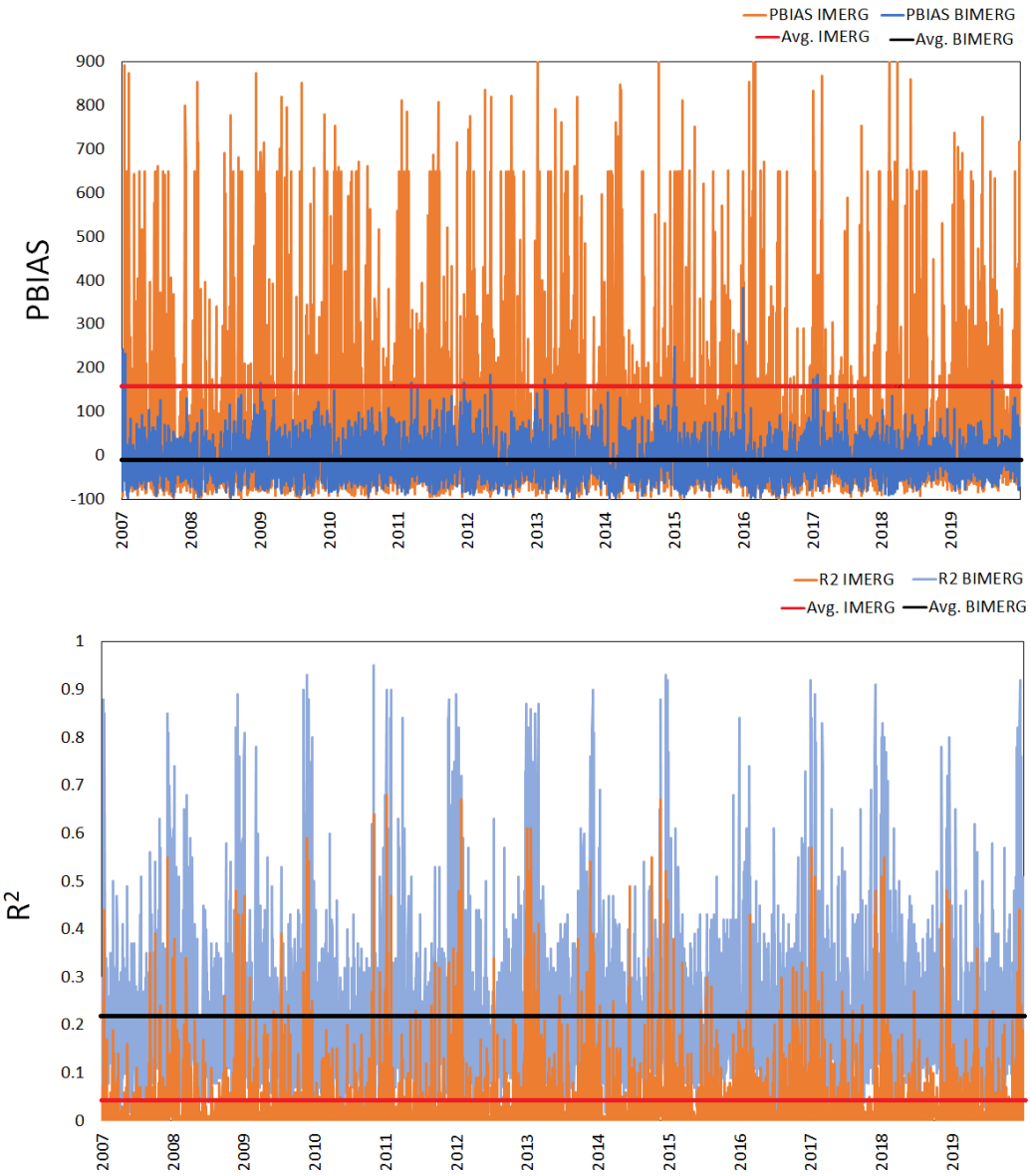

Fig. 8 Spatial Comparison of the Raw and Corrected IMERG

4.5 Performance in the estimation of spatial distributions of rainfall

The maps showing the spatial distribution of observed and BIMERG annual and seasonal rainfall were prepared to show the ability of BIMERG to replicate the spatial variability of rainfall (Fig. 9). The northeast of the peninsula receives the highest amount of rainfall (3500 to 4000 mm/year). BIMERG also replicated this zone with an average annual rainfall of 3500–4000 mm/year. BIMERG also able to reconstruct the low rainfall patches reliably. Similarly, the BIMERG was capable of reconstructing the spatial variability of NEM and SWM rainfall. The higher amount of NEM rainfall in the coastal region of peninsular Malaysia is in the range of 1000-1400 mm was well simulated by BIMERG. Relatively low rainfall received in Peninsular Malaysia during SWM was also found in BIMERG.

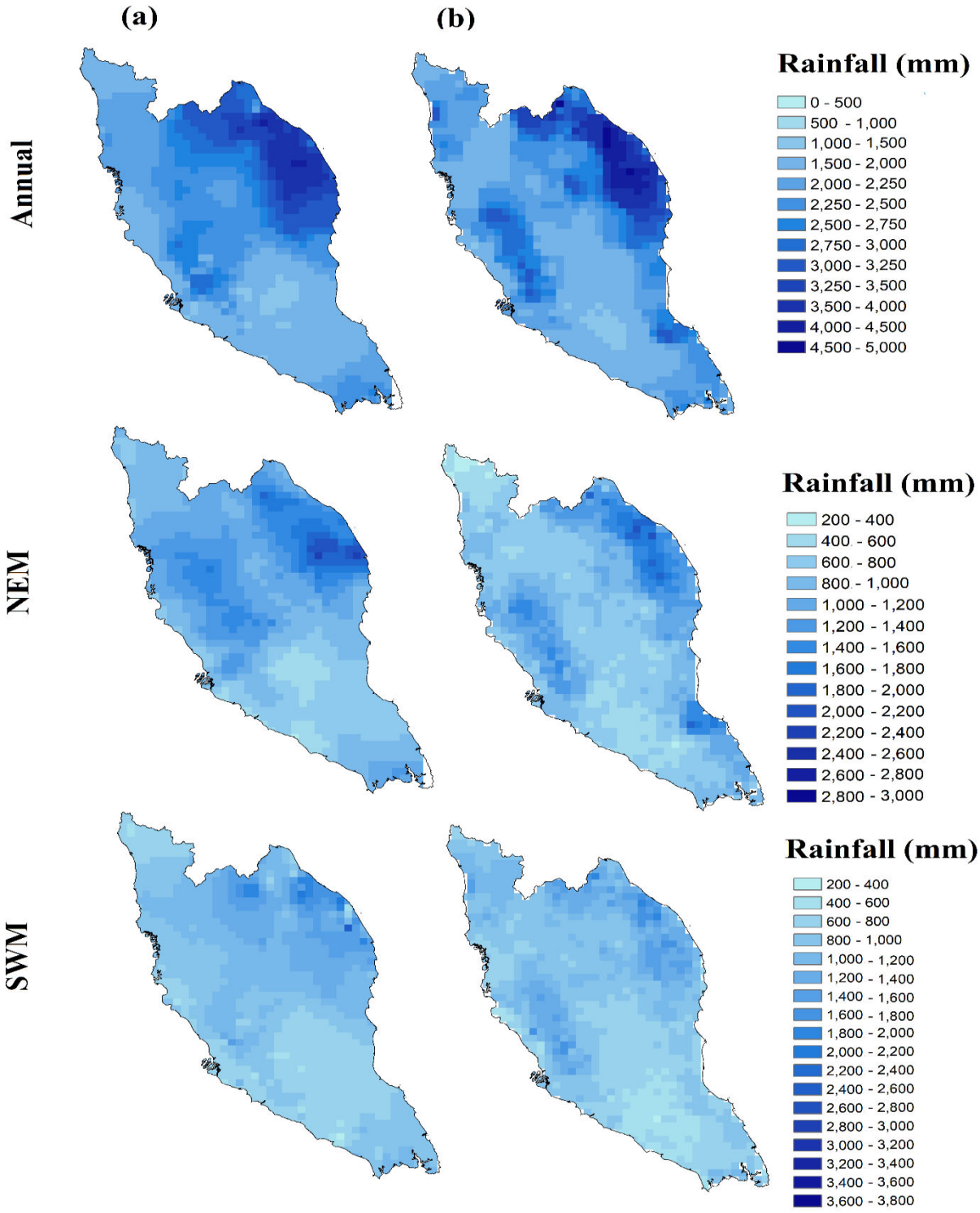

Fig. 9 Spatial distribution of annual, Northeast Monsson and Southwest Monsoon (a) observed, and (b) bias-corrected IMERG rainfall

4.6 Performance in constructing extreme rainfall events

The performance of BIMERG in reconstructing extreme rainfall events was also evaluated to assess its applicability in hydro-climatic studies in Peninsular Malaysia. Two heavy rainfall events were shown as examples: (1) Event-1 (December 09, 2007): several areas of Perak and Kelantan in the northern peninsula received a high rainfall ranging between 37.5 and 137 mm, which caused a flash flood in several regions. (2) Event-2 (December 18, 2014): heavy rainfall causes flood in Kuala Terengganu, located in the northeast of the Peninsular when the peak rainfall at some of the stations was estimated above 300 mm (405 mm at Kg. Keruak di Ulu Besut and 354.5 mm at Ibu Bekalan Sg. Angga, Ulu Besut station).

Fig 10 shows the ability of IMERG and BIMERG in reconstructing Event-1. IMERG underestimated rainfall at high rainfall regions while overestimated at low rainfall. The BIMERG estimated the extreme rainfall of 137 mm at Machang and the observed rainfall variability (37.5 to 98 mm) in other places as 40 to 80 mm.

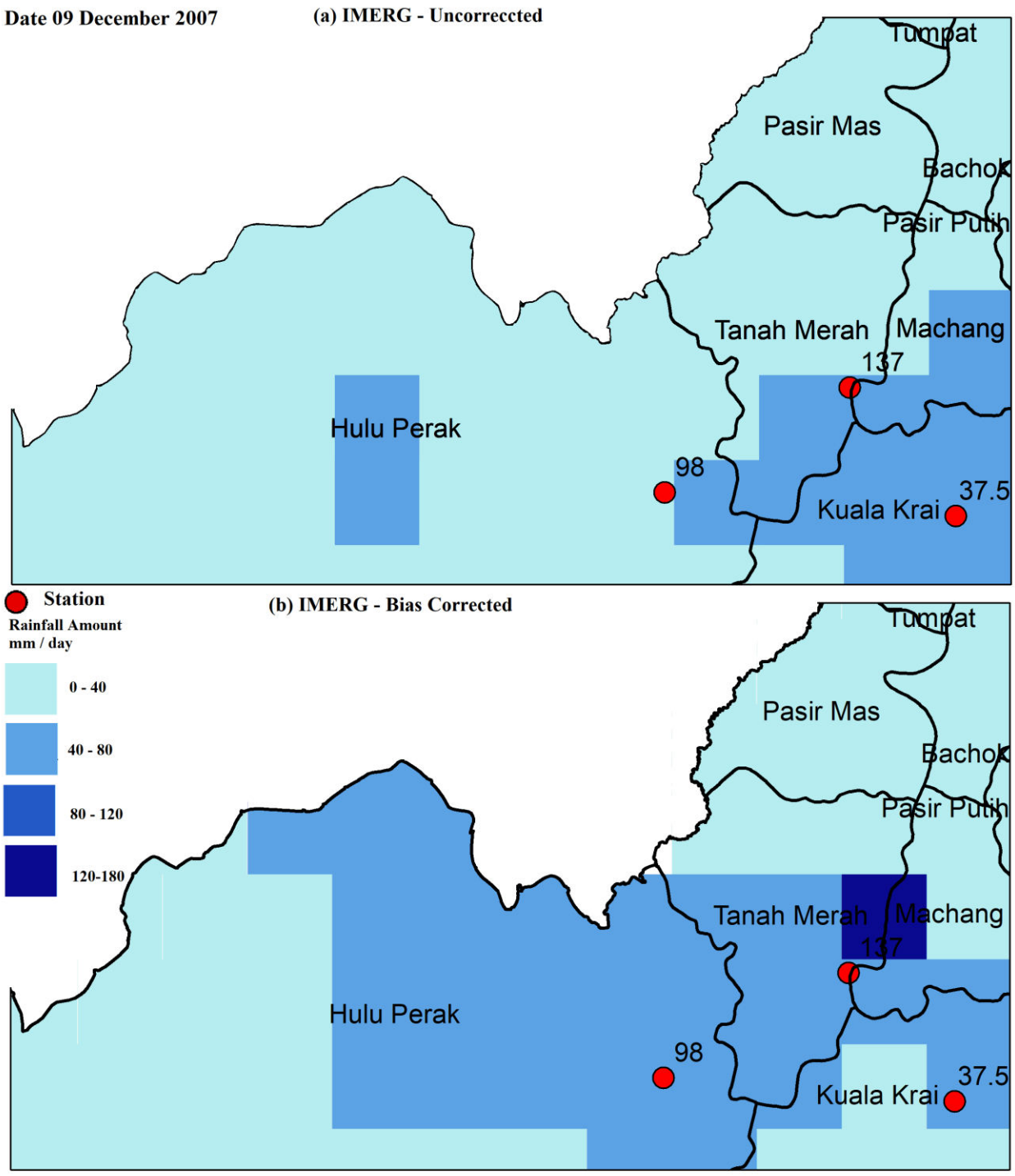

Fig. 10 Performance of IMERG and bias-corrected IMERG in reconstructing heavy rainfall event on December 09, 2007.

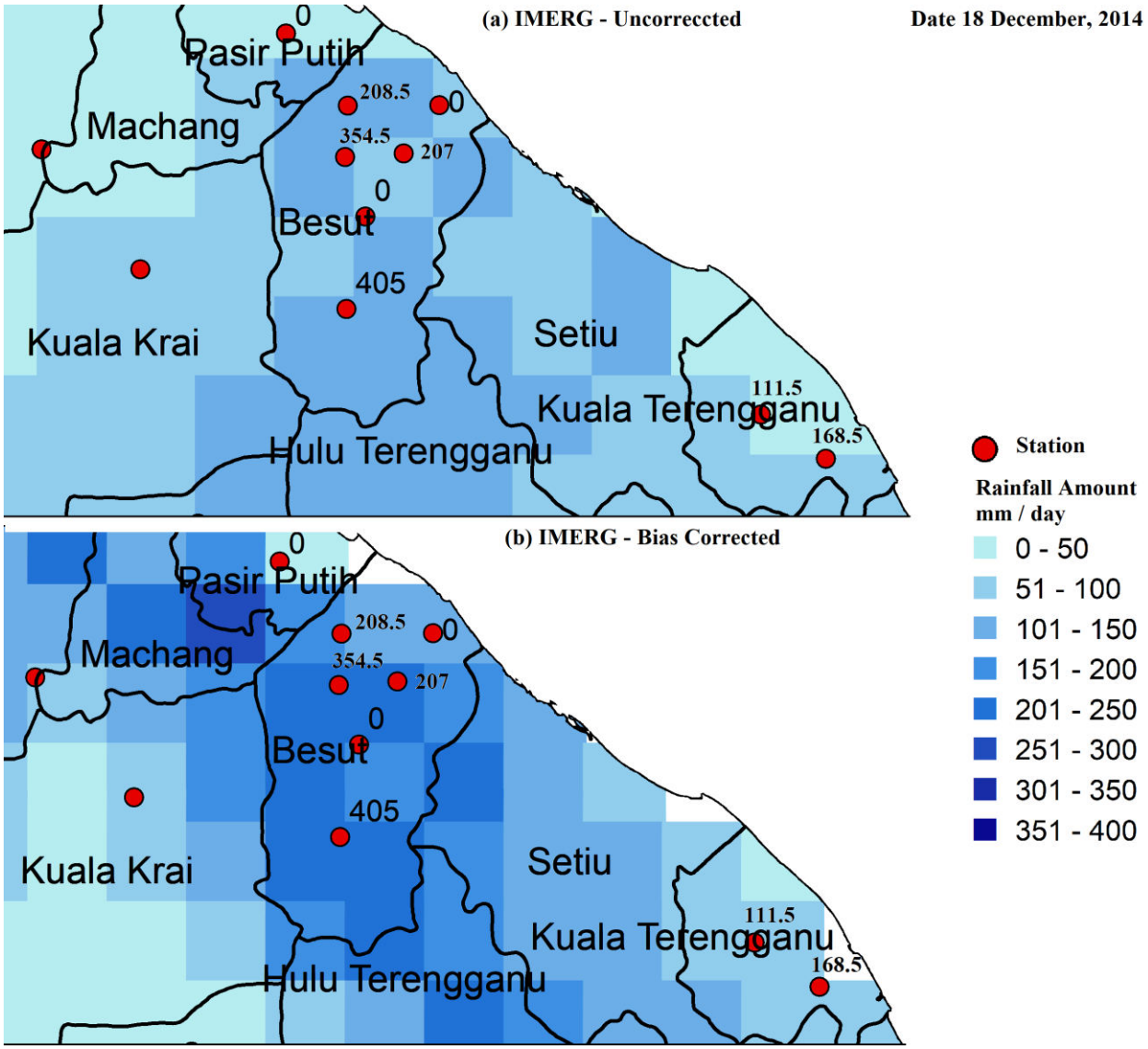

Fig 11. Performance of IMERG and BIMERG in reconstructing heavy rainfall event on December 18, 2014.

The comparative performance of IMERG and BIMERG in reconstructing Event-2 is shown in Fig 11. The figure shows that IMERG underestimated the event at most of the observed stations, whereas BIMERG estimated the pertinent event reliably. An extremely high rainfall amount of 405 mm was underestimated by BIMERG as 350 mm. However, it was able to estimate the rainfall amount and distribution reliably. The results indicate the potential of BIMERG for hydro-climatic studies like flood monitoring in Peninsular Malaysia.

5. Conclusion

The performance of SM2RAIN-ASCAT, CHIRPS, GsMap, PERSIANN-CDR, PERSIANN-CSS and IMERG in replicating observed daily rainfall data over Peninsular Malaysia is evaluated, and then the bias of the most suitable SBP is corrected using a novel two-stage bias-correction method. Comparing the 6 SBP with the observed data showed lower errors and a higher correlation of IMERG with observed rainfall than all other data sets. However, the results also revealed that IMERG still inherits significant biases. A novel two-stage bias correction method based on ML methods is proposed to correct bias in IMERG. A significant enhancement in the capability of bias-corrected IMERG indicates the effectiveness of the method. Using a classifier before correcting the amount of rainfall through regression has made the bias-correction method highly efficient. However, the method's performance largely depends on the classification of rainfall/no rainfall days by the classifier. Therefore, efficient classification and regression algorithms should be employed for the better performance of the method. In this study, the performance of two ML-based classifications and two regression algorithms were employed.

In the future, the potential of other ML algorithms for classification and regression should be explored to improve the bias-correction method. Besides, the classifier can be used to classify rainfall events that belong to different classes of rainfall intensity. The regression model can be employed to estimate rainfall amount within the intensity class to explore the possibility of improving model performance.

Acknowledgement

The authors would like to acknowledge the Belt and Road Special Foundation of the State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering (No. 2019491311, 2020491011), Young Top-Notch Talent Support Program of National High-level Talents Special Support Plan and the Ministry of Higher Education Malaysia (FRGS) (No. R.J130000.78515F092) for providing financial support to conduct this research. We also acknowledge the Department of Irrigation and Drainage Malaysia for providing the rainfall data of entire Peninsular Malaysia. Authors are also grateful to the European Space Agency (ESA), NASA data and Information Services Centre (DISC), the University of California, Irvine (UCI), and Climate Hazards Group

InfraRed Precipitation with Station data (CHIRPS) for providing satellite precipitation data through their web portals.

Statements & Declarations

Declaration

The authors declare that this paper is the result of our analysis and compilation. The paper is solely based on freely available data. All the authors contributed to the work.

Conflict of Interest

The authors have no relevant financial or non-financial interests to disclose

Funding Statement

Belt and Road Special Foundation of the State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering (No. 2019491311, 2020491011), Young Top-Notch Talent Support Program of National High-level Talents Special Support Plan and the Ministry of Higher Education Malaysia (FRGS) (No. R.J130000.78515F092)

Author's Contribution

Zafar Iqbal: Conceptualization, Methodology, Software, writing original draft. **Shamsuddin Shahid**: Formal analysis, Conceptualization, Validation, Software, Project administration: **Kamal Ahmed**: Draft Preparation, writing review and editing. **Tarmizi Ismail**: Resources, Technical Proofreading, Funding Acquisition, **Hamza Farooq Gabriel:** Data Curation. **Xiaojun Wang**: Writing - Review & Editing

Availability of data and material

The observed rainfall data is not available to be shared with a third party as per instruction from the Department of Irrigation and Drainage Malaysia. However, the GCM and Satellite data sets are freely available on the website/references given in the article**.**

Code availability

The codes used in the current study are available from the corresponding author on request.

Ethics approval

All authors provided ethical approval to submit the manuscript.

Consent to participate

The author has the consent to participate in the review process of this manuscript

Consent for publication

All authors provided consent to the publication of the article in TAAC.

References

Abera, W., Brocca, L. & Rigon, R. 2016. Comparative evaluation of different satellite rainfall estimation products and bias correction in the Upper Blue Nile (UBN) basin. *Atmospheric research,* 178**,** 471-483.

Ahmed, F. & Schumacher, C. 2015. Convective and stratiform components of the precipitation-moisture relationship. *Geophysical Research Letters,* 42**,** 10,453-10,462.

Ahmed, K., Shahid, S., Chung, E.-S., Ismail, T. & Wang, X.-J. 2017. Spatial distribution of secular trends in annual and seasonal precipitation over Pakistan. *Climate Research,* 74**,** 95-107.

Alharbi, R. 2019. *Bias Adjustment of Satellite-Based Precipitation Estimation Using Limited Gauge Measurements and Its Implementation on Hydrologic Modeling.* UC Irvine.

Badron, K., Ismail, A. F., Asnawi, A., Nordin, M. a. W., Alam, A. Z. & Khan, S. 2015. Classification of precipitation types detected in Malaysia. *Theory and Applications of Applied Electromagnetics.* Springer.

Beck, H. E., Wood, E. F., Mcvicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O. M., Sheffield, J. & Karger, D. N. 2020. Bias Correction of Global High-Resolution Precipitation Climatologies Using Streamflow Observations from 9372 Catchments. *Journal of Climate,* 33**,** 1299-1315.

Bhatti, H. A., Rientjes, T., Haile, A. T., Habib, E. & Verhoef, W. 2016. Evaluation of Bias Correction Method for Satellite-Based Rainfall Data. *Sensors (Basel, Switzerland)* [Online], 16. Available: https://doi.org/10.3390/s16060884 [Accessed 2016/06//].

Bitew, M. M. & Gebremichael, M. 2011. Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. *Water Resources Research,* 47.

Breiman, L. 2001. Random Forests, Vol. 45. *Mach Learn,* 1.

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W. & Levizzani, V. 2014. Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. *Journal of Geophysical Research: Atmospheres,* 119**,** 5128-5141.

Brocca, L., Filippucci, P., Hahn, S., Ciabatta, L., Massari, C., Camici, S., Schüller, L., Bojkov, B. & Wagner, W. 2019. SM2RAIN-ASCAT (2007–2018): Global daily satellite rainfall from ASCAT soil moisture. *Earth Syst. Sci. Data Discuss***,** 1-31.

Chaudhary, S. & Dhanya, C. Investigating the performance of bias correction algorithms on satellite-based precipitation estimates. Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI, 2019. International Society for Optics and Photonics, 111490Z.

Ciabatta, L., Brocca, L., Massari, C., Moramarco, T., Puca, S., Rinollo, A., Gabellani, S. & Wagner, W. 2015. Integration of satellite soil moisture and rainfall observations over the Italian territory. *Journal of Hydrometeorology,* 16**,** 1341-1355.

Ciabatta, L., Marra, A. C., Panegrossi, G., Casella, D., Sanò, P., Dietrich, S., Massari, C. & Brocca, L. 2017. Daily precipitation estimation through different microwave sensors: Verification study over Italy. *Journal of hydrology,* 545**,** 436-450.

Ciabatta, L., Massari, C., Brocca, L., Gruber, A., Reimer, C., Hahn, S., Paulik, C., Dorigo, W., Kidd, R. & Wagner, W. 2018. SM2RAIN-CCI: a new global long-term rainfall data set derived from ESA CCI soil moisture. *Earth System Science Data,* 10**,** 267.

Dewan, A., Hu, K., Kamruzzaman, M. & Uddin, M. R. 2019. Evaluating the spatiotemporal pattern of concentration, aggressiveness and seasonality of precipitation over Bangladesh with time–series Tropical Rainfall Measuring Mission data. *Extreme Hydroclimatic Events and Multivariate Hazards in a Changing Environment.* Elsevier.

Fawagreh, K., Gaber, M. M. & Elyan, E. 2014. Random forests: from early developments to recent advancements. *Systems Science & Control Engineering: An Open Access Journal,* 2**,** 602-609.

Fernández-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. 2014. Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research,* 15**,** 3133-3181.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. & Michaelsen, J. 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data,* 2**,** 150066.

Gadelha, A. N., Coelho, V. H. R., Xavier, A. C., Barbosa, L. R., Melo, D. C., Xuan, Y., Huffman, G. J., Petersen, W. A. & Almeida, C. D. N. 2019. Grid box-level evaluation of IMERG over Brazil at various space and time scales. *Atmospheric Research,* 218**,** 231-244.

Gerrity Jr, J. P. 1992. A note on Gandin and Murphy's equitable skill score. *Monthly weather review,* 120**,** 2709-2712.

Ghaffari, A., Abdollahi, H., Khoshayand, M., Bozchalooi, I. S., Dadgar, A. & Rafiee-Tehrani, M. 2006. Performance comparison of neural network training algorithms in modeling of bimodal drug delivery. *International journal of pharmaceutics,* 327**,** 126-138.

Gupta, V., Jain, M. K., Singh, P. K. & Singh, V. 2019. An assessment of global satellite-based precipitation datasets in capturing precipitation extremes: A comparison with observed precipitation dataset in India. *International Journal of Climatology*.

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**,** 2491-2509.

Heidke, P. 1926. Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geografiska Annaler,* 8**,** 301-349.

Heung, B., Bulmer, C. E. & Schmidt, M. G. 2014. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. *Geoderma,* 214-215**,** 141-154.

Hsu, K.-L., Gao, X., Sorooshian, S. & Gupta, H. V. 1997. Precipitation estimation from remotely sensed information using artificial neural networks. *Journal of applied meteorology,* 36**,** 1176-1190.

Huang, M., Lin, R., Huang, S. & Xing, T. 2017. A novel approach for precipitation forecast via improved K-nearest neighbor algorithm. *Advanced Engineering Informatics,* 33**,** 89-95.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K.-L., Joyce, R. J., Kidd, C., Nelkin, E. J., Sorooshian, S., Stocker, E. F. & Tan, J. 2020. Integrated multi-satellite retrievals for the Global Precipitation Measurement (GPM) mission (IMERG). *Satellite precipitation measurement.* Springer, Cham.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P. & Yoo, S.-H. 2015. NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). *Algorithm Theoretical Basis Document (ATBD) Version,* 4**,** 26.

Iqbal, Z., Shahid, S., Ahmed, K., Ismail, T., Khan, N., Virk, Z. T. & Johar, W. 2020. Evaluation of global climate models for precipitation projection in sub-Himalaya region of Pakistan. *Atmospheric Research***,** 105061.

Katiraie-Boroujerdy, P.-S., Rahnamay Naeini, M., Akbari Asanjan, A., Chavoshian, A., Hsu, K.-L. & Sorooshian, S. 2020. Bias Correction of Satellite-Based Precipitation Estimations Using Quantile Mapping Approach in Different Climate Regions of Iran. *Remote Sensing,* 12**,** 2102.

King, F., Erler, A. R., Frey, S. K. & Fletcher, C. G. 2020. Application of machine learning techniques for regional bias correction of SWE estimates in Ontario, Canada. *Hydrology and Earth System Sciences Discussions***,** 1-26.

Kummerow, C. D., Randel, D. L., Kulie, M., Wang, N.-Y., Ferraro, R., Joseph Munchak, S. & Petkovic, V. 2015. The Evolution of the Goddard Profiling Algorithm to a Fully Parametric Scheme. *Journal of Atmospheric and Oceanic Technology,* 32**,** 2265-2280.

Le, X.-H., Lee, G., Jung, K., An, H.-U., Lee, S. & Jung, Y. 2020. Application of Convolutional Neural Network for Spatiotemporal Bias Correction of Daily Satellite-Based Precipitation. *Remote Sensing,* 12**,** 2731.

Liu, Y. Y., Dorigo, W. A., Parinussa, R., De Jeu, R. A., Wagner, W., Mccabe, M. F., Evans, J. & Van Dijk, A. 2012. Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sensing of Environment,* 123**,** 280-297.

Liu, Y. Y., Parinussa, R., Dorigo, W. A., De Jeu, R. A., Wagner, W., Van Dijk, A., Mccabe, M. F. & Evans, J. 2011. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals.

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.

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**,** 1213-1233.

Mega, T., Ushio, T., Takahiro, M., Kubota, T., Kachi, M. & Oki, R. 2019. Gauge-Adjusted Global Satellite Mapping of Precipitation. *IEEE Transactions on Geoscience and Remote Sensing,* 57**,** 1928-1935.

Nashwan, M. & Shahid, S. 2019a. Symmetrical uncertainty and random forest for the evaluation of gridded precipitation and temperature data. *Atmospheric Research,* 230**,** 104632.

Nashwan, M., Shahid, S., Dewan, A., Ismail, T. & Alias, N. 2019b. Performance of five high resolution satellite-based precipitation products in arid region of Egypt: An evaluation. *Atmospheric Research,* 236**,** 104809.

Nashwan, M. S., Shahid, S. & Chung, E.-S. 2020. High-Resolution Climate Projections for a Densely Populated Mediterranean Region. *Sustainability,* 12**,** 3684.

Nguyen, P., Ombadi, M., Sorooshian, S., Hsu, K., Aghakouchak, A., Braithwaite, D., Ashouri, H. & Thorstensen, A. R. 2018. The PERSIANN family of global satellite precipitation data: A review and evaluation of products. *Hydrology and Earth System Sciences,* 22**,** 5801-5816.

Noor, M., Bin Ismail, T., Ullah, S., Iqbal, Z., Nawaz, N. & Ahmed, K. 2019a. A non-local model output statistics approach for the downscaling of CMIP5 GCMs for the projection of rainfall in Peninsular Malaysia. *Journal of Water and Climate Change*.

Noor, M., Ismail, T. B., Shahid, S., Ahmed, K., Chung, E.-S. & Nawaz, N. 2019b. Selection of CMIP5 multi-model ensemble for the projection of spatial and temporal variability of rainfall in peninsular Malaysia. *Theoretical and Applied Climatology,* 138**,** 999-1012.

Paredes-Trejo, F., Barbosa, H. A. & Rossato Spatafora, L. 2018. Assessment of SM2RAIN-derived and state-of-the-art satellite rainfall products over Northeastern Brazil. *Remote Sensing,* 10**,** 1093.

Peña-Arancibia, J. L., Van Dijk, A. I. J. M., Renzullo, L. J. & Mulligan, M. 2013. Evaluation of Precipitation Estimation Accuracy in Reanalyses, Satellite Products, and an Ensemble Method for Regions in Australia and South and East Asia. *Journal of Hydrometeorology,* 14**,** 1323-1333.

Pour, S. H., Abd Wahab, A. K., Shahid, S. & Ismail, Z. B. 2020. Changes in reference evapotranspiration and its driving factors in peninsular Malaysia. *Atmospheric Research***,** 105096.

Pratama, A. W., Buono, A., Hidayat, R. & Harsa, H. 2018. Estimating parameter of nonlinear bias correction method using nsga-ii in daily precipitation data. *Telkomnika,* 16**,** 241-249.

Rauniyar, S., Protat, A. & Kanamori, H. 2017. Uncertainties in TRMM-Era multisatellite-based tropical rainfall estimates over the Maritime Continent. *Earth and Space Science,* 4**,** 275-302.

Sa'adi, Z., Shahid, S., Chung, E.-S. & Ismail, T. B. 2017. Projection of spatial and temporal changes of rainfall in Sarawak of Borneo Island using statistical downscaling of CMIP5 models. *Atmospheric Research,* 197**,** 446-460.

Sa'adi, Z., Shiru, M. S., Shahid, S. & Ismail, T. 2020. Selection of general circulation models for the projections of spatio-temporal changes in temperature of Borneo Island based on CMIP5. *Theoretical and Applied Climatology,* 139**,** 351-371.

Schroeer, K., Kirchengast, G. & O, S. 2018. Strong dependence of extreme convective precipitation intensities on gauge network density. *Geophysical Research Letters,* 45**,** 8253-8263.

Serrat-Capdevila, A., Merino, M., Valdes, J. B. & Durcik, M. 2016. Evaluation of the performance of three satellite precipitation products over Africa. *Remote Sensing,* 8**,** 836.
Shankar, T. & Bandyopadhyay, S. 2007. Prediction of extrudate properties using artificial neural networks. *Food and Bioproducts Processing,* 85**,** 29-33.
Sivapragasam, C., Arun, V. M. & Giridhar, D. 2010. A simple approach for improving spatial interpolation of rainfall using ANN. *Meteorology and Atmospheric Physics,* 109**,** 1-7.
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***,** 105007.
Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S. & Hsu, K. L. 2018. A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics,* 56**,** 79-107.
Tan, M. L. & Duan, Z. 2017. Assessment of GPM and TRMM precipitation products over Singapore. *Remote Sensing,* 9**,** 720.
Tan, M. L., Ficklin, D. L., Ibrahim, A. L. & Yusop, Z. 2014. Impacts and uncertainties of climate change on streamflow of the Johor River Basin, Malaysia using a CMIP5 General Circulation Model ensemble. *Journal of Water and Climate Change,* 5**,** 676-695.
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 Sensing,* 7**,** 1504-1528.
Tan, M. L. & Santo, H. 2018. Comparison of GPM IMERG, TMPA 3B42 and PERSIANN-CDR satellite precipitation products over Malaysia. *Atmospheric Research,* 202**,** 63-76.
Urbanczik, R. 1996. Learning in a large committee machine: worst case and average case. *EPL (Europhysics Letters),* 35**,** 553.
Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K., Aonashi, K., Inoue, T., Takahashi, N., Iguchi, T., Kachi, M., Oki, R., Morimoto, T. & Kawasaki, Z. I. 2009. A kalman filter approach to the global satellite mapping of precipitation (GSMaP) from combined passive microwave and infrared radiometric data. *Journal of the Meteorological Society of Japan,* 87 A**,** 137-151.
Webster, P. J., Magana, V. O., Palmer, T., Shukla, J., Tomas, R., Yanai, M. & Yasunari, T. 1998. Monsoons: Processes, predictability, and the prospects for prediction. *Journal of Geophysical Research: Oceans,* 103**,** 14451-14510.
Wmo 1994. GUIDE TO HYDROLOGICAL PRACTICES. *DATA ACQUISITION AND PROCESSING, ANALYSIS, FORECASTING AND OTHER APPLICATIONS*.
Xie, P. & Xiong, A. Y. 2011. A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *Journal of Geophysical Research: Atmospheres,* 116.
Yang, Z., Hsu, K., Sorooshian, S., Xu, X., Braithwaite, D. & Verbist, K. M. 2016. Bias adjustment of satellite-based precipitation estimation using gauge observations: A case study in Chile. *Journal of Geophysical Research: Atmospheres,* 121**,** 3790-3806.
Yong, B., Ren, L. L., Hong, Y., Wang, J. H., Gourley, J. J., Jiang, S. H., Chen, X. & Wang, W. 2010. Hydrologic evaluation of Multisatellite Precipitation Analysis standard precipitation products in basins beyond its inclined latitude band: A case study in Laohahe basin, China. *Water Resources Research,* 46.
Zafar, B. J. & Chandrasekar, V. Classification of precipitation type from space borne precipitation radar data and 2D wavelet analysis. IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, 2004. IEEE, 3570-3573.