Evaluation of climate reanalysis and space-borne precipitation products over Bangladesh

Md. Atiqul Islam* and Nick Cartwright

School of Engineering and Built Environment and Griffith Centre for Coastal Management, Gold Coast Campus, Griffith University, Queensland 4222, Australia

*Corresponding author:

Name: Md. Atiqul Islam

Email: atiqul.islam@griffithuni.edu.au, atiqul@ce.kuet.ac.bd

Address: School of Engineering and Built Environment, Griffith University (Gold Coast Campus), Parklands Drive, Southport 4222, Queensland, Australia, Telephone: +61755529142

Abstract This study aims to quantify the spatial distribution of errors in two climate reanalysis (ERA5 and CFSR) and two satellite (TMPA-RT and TMPA-V7) precipitation products over Bangladesh. The datasets are assessed against ground-based rain gauge observations to capture the extreme rainfall accumulations at daily temporal scale over a 5-year period (January 2010–December 2014). The bias ratio scores indicate that CFSR and TMPA-RT seriously overestimate the rainfall values over much of the study area. Whilst TMPA-V7 performs better than the other precipitation products, all datasets lose their detection skills substantially for higher quantile thresholds (i.e. above 50th and 75th percentiles). With respect to rainfall detection metrics – probability of detection (POD) and volumetric hit index (VHI) – both ERA5 and CFSR show superior performance (in the range 0.9–1.0 for all the analysis grid boxes). All rainfall datasets are equally good in terms of false alarm ration (FAR) and volumetric FAR (VFAR), even though the lowest values are associated with ERA5 for higher quantiles. All products demonstrate a decrease in skill to capture the amount of rainfall but show satisfactory results to detect the rainfall events when using higher quantile thresholds (i.e. rainfall above the 50th and 75th percentiles) to sample the data before computing product skill.
Keywords Bangladesh; monsoon; TRMM, CFSR, ECMWF, ERA5

1 Introduction

Precipitation is an important resource for various socioeconomic activities globally, such as irrigation, domestic water supply, hydroelectric power generation, fish farming, navigation and so on. To better manage and mitigate issues arising from the temporal and spatial variability in rainfall (e.g. prediction of hydrological extremes such as floods and droughts and efficient water resources management), high quality rainfall data is required. However, in certain countries, access to high quality rain gauge data with fine spatial and temporal resolution is severely limited and can be almost absent in some countries e.g. Burkina Faso (Dembélé and Zwart 2016), Laos (Liu et al. 2017) and Peru (Mantas et al. 2015). Consequently, these regions rely on the use of satellite derived or climate reanalysis products for rainfall estimation where surface rainfall measurement facilities (rain gauge and/or ground-based radar) are sparse or non-existent (Maidment et al. 2013).

One such country lacking in ground-based rain data is Bangladesh. Landslides are very common in the hilly and mountainous south-eastern parts of Bangladesh. On 12 June 2017, intense and prolonged heavy monsoon rainfall triggered a series of floods and landslides in this region and killed at least 152 people including property damages and economic losses. Unfortunately, rain gauges are absent in this part of the country (see Figs 1 and 2). Therefore, this paper aims to evaluate a range of satellite-derived and reanalysis rainfall products for Bangladesh.

In the past few decades, the application of reanalysis and satellite rainfall estimates has increased remarkably for their higher spatial coverage, uninterrupted time series and near real time availability. These precipitation data products are currently being used in hydrological modelling and river flow routing for streamflow simulation
and flood prediction, landslide modelling for nowcasting any potential landslide activity, disaster preparedness in areas likely to be affected due to extreme precipitation events, drought warning and monitoring, irrigation scheduling, data assimilation into land surface models and validation of climate model precipitation outputs (Kirschbaum et al. 2017). Although, model-derived and satellite-based datasets provide a promising substitute for ground-based rainfall estimates, undetermined errors and biases associated with the observations, sampling and estimation algorithms of the precipitation products may hamper their use in practical applications.

The accuracy of reanalysis and satellite rainfall estimates may vary over different parts of the world with respect to climate, season and topography (Ebert et al. 2007, Sorooshian et al. 2011, Dinku et al. 2011). Therefore, it is very important to assess these precipitation datasets at different geographic locations and various climatic conditions. These assessments will surely help to identify the strengths and limitations of the rainfall products under diverse conditions. To date, numerous studies have been carried out to validate and verify satellite-retrieved precipitation products. Literature related to the evaluation of these datasets can be classified with respect to different point of views, such as, most of the studies are performed on the scale basis (Prakash et al. 2016b), whereas very few research works are carried out on the event basis (Mei et al. 2014). A large number of studies have been devoted to assessing satellite rainfall datasets through streamflow simulation (Xue et al. 2013). Again, some studies reported in the literature have been involved to find the suitability of satellite precipitation for drought monitoring (Vernimmen et al. 2012). A significant number of articles in the literature have also contributed to validate satellite-based rainfall products over complex topography and diverse climate (Gao and Liu 2013, Thiemig et al. 2012). Although the majority of the works is carried out over sparsely-gauged basins (Liu et al. 2017),
several studies are performed over the areas with sophisticated surface rainfall measurement facilities (Sohn et al. 2010). Both gridded reference data (Prakash et al. 2016a) and point rain gauge observations (Yang et al. 2016) are applied to verify satellite rainfall estimates. Reanalysis rainfall datasets are also evaluated over different parts of the world (Leewu et al. 2015, Tan et al. 2018, Nkiaka et al. 2017, Derin et al. 2016).

However, evaluation of model reanalysis and satellite rainfall products is very limited over Bangladesh. Rahman et al. (2012) validated ERA40 (European Centre for Medium-Range Weather Forecasts, ECMWF Reanalysis 40 years) and TMPA (Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis) 3B42 at daily time scale. Islam and Uyeda (2007) analysed TMPA 3B42 and TMPA 3B43 to understand the rainfall characteristics of Bangladesh and Islam and Uyeda (2008) studied the variation of vertical rainfall intensity in and around Bangladesh using various TRMM products i.e. 2A25, 3B42 and 3B43. More recently, Islam (2018) compared six satellite precipitation products with rain gauge measurements at various temporal scales i.e. 3, 6, 12 and 24 h over Bangladesh. The study concluded that all products indicate reasonably good skill in detecting the occurrence of rainfall events, however, they show very poor performance in estimating the amount of rainfall especially the extreme (heavy) rainfall accumulations. However, all the previous studies evaluated older version of reanalysis and TMPA rainfall products using a very limited number of rain gauges ($N = 3$–$35$). In addition, none of the literature shows spatial distribution of errors of the precipitation datasets to capture the heavy rainfall accumulations.

The distribution of rain gauges in Bangladesh is very sparse and uneven throughout the country. There is almost no rain gauges in the remote areas such as
wetlands (e.g. haors in the northeastern parts of Bangladesh), forests (e.g. the largest mangrove forest Sundarban in the southwestern parts of the country), hilly and mountainous regions (see Figs 1 and 2). Figure 2 shows that rain gauges are almost absent at higher elevations (i.e. more than 100 m above mean sea level). Lack of in situ observations as well as access to the available observations are the major challenges to validate reanalysis and satellite rainfall estimates over the study region.

This study aims to build on the limited number of rainfall product evaluations in Bangladesh by evaluating two model-based climate reanalyses and two satellite precipitation products against daily rain gauge observations for a period of five years (January 2010–December 2014). The main objective of the present research is to explore the performance of the reanalysis and satellite datasets to detect the extreme daily rainfall accumulations. Specific objectives include: (a) to evaluate the skills of the precipitation products in estimating the amount of rainfall; (b) to assess the capabilities in detecting the rainfall occurrences; (c) to determine the temporal detection skills of the rainfall datasets; and (d) to quantify the seasonal performance of the precipitation estimates. Furthermore, this work highlights the comparative performance of reanalysis and satellite-derived rainfall estimates over the monsoon dominated sub-tropical humid climate of Bangladesh.

The study domain and datasets are described in Section 2, while Section 3 presents the evaluation statistics and results. Finally, section 4 summarizes the findings and draws some conclusions.
2 Validation area and data

2.1 Validation area

The precipitation datasets are evaluated over Bangladesh, which is situated in the south-central part of the Asian continent (see Fig. 1). With an area of about 147,570 km², the country is located between 20°N–27°N and 88°E–93°E. Bangladesh has almost flat topography with the exceptions of the mountainous regions in the southeastern and northeastern parts of the country. There are also high grounds in the northwestern parts of Bangladesh. The study area experiences a sub-tropical monsoon climate, which sustains from June to September, with about 80% of the total annual rainfall occurring during the monsoon period which are accompanied by severe floods. The areal distribution of rainfall is not uniform and varies significantly over the study region with the mountainous and coastal regions experiencing the maximum rainfall while western parts are the driest.

2.2 Precipitation data

2.2.1 Rain gauge data

The rain gauge data used to evaluate the rainfall products in this study were collected from Bangladesh Meteorological Department (BMD) and Bangladesh Water Development Board (BWDB). The BWDB gauge network consists of 261 rain gauges with daily temporal resolution. After a rigorous screening process, 234 stations were selected for the analysis (see Fig. 1). Some gauges were excluded from analysis for not having required meta data (i.e. station names, latitudes and longitudes) and for missing rainfall records during the study period. An additional 35 rain gauges from BMD were also included in the analysis. The temporal resolution of BMD gauge is 3 h so daily accumulations were calculated from these 3-hourly accumulations. Therefore, a total of
269 rain gauges are applied for the evaluation of the reanalysis and satellite rainfall estimates. This is a significantly more robust evaluation dataset compared to previous studies who used a limited number of gauges ($N = 3–35$, e.g. Rahman et al. 2012, Islam and Uyeda 2007, Islam and Uyeda 2008, Islam 2018).

To facilitate the comparison of satellite and reanalysis products against the gauge data, a regular $0.25^\circ \times 0.25^\circ$ analysis grid was established across the study region. If a grid cell contained a single rain gauge, the gauge data was used as the reference data for that cell. If there were multiple gauges within a cell, then the arithmetic average of the gauges was used as the reference data. As the rain gauge network is very sparse over the study area, interpolation based on in-situ data would result serious error in interpolated estimates (Lu et al. 2018, Moazami et al. 2013, Gao and Liu 2013). The arithmetic mean is both widely used and acceptable in the context of evaluating satellite and reanalysis precipitation products (Palomino-Ángel et al. 2019, Tan and Santo 2018, Tan et al. 2018, Hobouchian et al. 2017, Beck et al. 2017, Tan et al. 2016, Moazami et al. 2016, Yang et al. 2016 and Pipunic et al. 2015). If there were no gauges in a cell, the cell was not considered in the computation of the skill metrics.

In accordance with data availability, the analysis was conducted over the 5-year time slice from 2010 to 2014 using daily accumulations, which equates to 1826 rain gauge data pairs at each grid cell location. As the analysis is performed for a relatively long time period (5 years), it is assumed that the spatial representativeness errors of the rain gauge observations will be minimized (Hirpa et al. 2010, Moazami et al. 2013, Romilly and Gebremichael 2011). It is noted that the daily accumulations of TMPA products start from 1.5 UTC and finish at 1.5 UTC of the next day (Sapiano and Arkin 2009), as compared to ERA5, CFSR and rain gauge data, which have the common time window from 0 UTC to 24 UTC. Since this work is carried out at the daily (24 h)
temporal scale, it is considered that 1.5 h offset time of TMPA datasets will not affect the results significantly.

2.2.2 Climate reanalysis data

The main objective of climate reanalysis projects is to create a multiyear, quasi-global gridded representation of the atmosphere from the Earth’s surface to the stratosphere by implementing a constant numerical weather prediction (NWP) model and data assimilation technique. The most recent climate reanalysis is ERA5 produced by ECMWF which is an update of the widely used ERA-Interim (Dee et al. 2011). ERA5 has a spatial resolution of ~0.28° (nearly 31 km) and 137 vertical levels and is based on the Integrated Forecast System (IFS) Cycle 41r2 4D-Var forecast model developed by the ECMWF. The output analysis and forecast fields are produced at hourly temporal resolution worldwide. This hourly time resolution is an important improvement of ERA5 with respect to ERA-Interim. For example, precipitation estimation by ERA5 is better than that of ERA-Interim over land in the deep tropics (Hennermann 2018). More information about ERA5 can be found online\(^1\). ERA5 precipitation estimates with 0.25° × 0.25° latitude/longitude spatial and hourly temporal resolution were downloaded from the Copernicus website\(^2\). Table 1 shows some basic information of different precipitation products used in this work.

\[\text{The Climate Forecast System Reanalysis (CFSR) climate reanalysis product is also widely used and is produced by the National Centers for Environmental Prediction (NCEP) (Saha et al. 2010). CFSR precipitation data with 6 h temporal and 0.50°} \times \text{0.50°}\]

\(^1\) https://confluence.ecmwf.int/display/CKB/ERA5+data+documentation

\(^2\) http://cds.climate.copernicus.eu/
spatial resolution were collected from the Rainfall Data Archive of the US National Center for Atmospheric Research (NCAR, Boulder, CO, USA)\(^3\). Version 1 (CFSv1) is the first global fully coupled atmosphere–ocean–land surface–sea ice model used at NCEP (Saha et al. 2014) and covers the time slice from January 1979 to March 2011, with a native grid resolution of 0.3°. Version 2 (CFSv2) improves the consistency between the model and the initial conditions developed by the data assimilation systems and has been implemented from April 2011 to present on a native grid of 0.2°. Both CFSv1 and CFSv2 (hereafter CFSR) have been re-gridded onto the 0.25° × 0.25° analysis grid via bilinear interpolation.

2.2.3 Satellite derived rainfall data

Satellite precipitation estimation algorithms combine information mostly from passive microwave (PMW) sensors board on low Earth orbiting (LEO) satellites and visible infrared (Vis-IR) sensors board on Polar orbiting geostationary (GEO) satellites (Sapiano and Arkin 2009). Vis-IR sensors indirectly infer precipitation rates from outgoing long wave radiation and cloud-top brightness temperature. PMW sensors conform a more direct relationship of rainfall rates with physical properties of cloud masses (i.e. size and phase of the hydrometeors) but the sampling frequency is extremely low (a full scan of globe is fewer than once per day). Each satellite product has its own estimation algorithm that differs from others with respect to spatial and temporal scales, sensors of observation, merging technique of input datasets etc.

In this study, TMPA 3B42-real time (3B42-RT) and TMPA 3B42 (version 7) precipitation products are validated (Huffman et al. 2007). The 3B42 product is chosen

\(^3\) http://rda.ucar.edu/
for the analysis as it is one of the most widely used multi-satellite precipitation estimates (Darand et al. 2017). The two satellite rainfall datasets are selected for evaluation to investigate the competence of Vis-IR and PMW-based (i.e. TMPA 3B42-RT) and Vis-IR, PMW and gauge-adjusted (i.e. TMPA 3B42) precipitation estimation algorithms across the study domain. Both datasets are available at daily time scale and 0.25° spatial resolution. In the creation of both products, the microwave and infrared precipitation observations are combined, taking microwave estimates just as they are, when and where available, and the gaps between the microwave estimates are filled with the microwave calibrated infrared estimates. Then, 3-hourly microwave-infrared merged precipitation products are aggregated over a calendar month to produce monthly gridded multi-satellite (MS) precipitation products.

The research grade product (TMPA 3B42) is then obtained by combining the monthly gridded multi-satellite (MS) precipitation products with monthly gridded rain gauge data according to Huffman et al. (1997) to produce a post real time research-grade precipitation product (SG). The ratios of gauge-adjusted satellite and satellite only precipitation (SG/MS) are determined for all the 0.25° × 0.25° grid boxes to scale each 3-hourly satellite only rainfall field of the month. 3-hourly 3B42 estimates are accumulated to produce the daily TMPA 3B42 product (hereafter TMPA-V7).

The real time product (TMPA 3B42-RT, hereafter TMPA-RT) does not include the gauge data correction step. It is worth mentioning that routine production of TRMM precipitation estimates ended on October 2014 (PR data ended on 7 October 2014 and TMI switched off on 8 April 2015, Bolvin and Huffman 2015, Huffman 2016). To keep

---

4 https://ssw.gsfc.nasa.gov/SSW/
research-grade products usable, TMPA-V7 datasets are produced using climatological calibration as like TMPA-RT from October 2014.

3 Validation statistics and results

3.1 General assessment

Figure 3 presents the empirical cumulative distribution functions (ecdfs) of wet days (daily rainfall > 1mm) for rain gauge, climate reanalysis and satellite rainfall estimates over all of Bangladesh (i.e. all cells in the analysis grid). In Fig. 3, the ratio of the number of rainfall data less than or equal to a rainfall accumulation ($x$) to the total number of rainfall data, $F(x)$, is plotted against the rainfall accumulation ($x$). It is clear that TMPA-RT and CFSR overestimate at all rainfall magnitudes, while TMPA-V7 shows a reasonably good correspondence with rain gauge observations below 25 mm/d accumulation (see Fig. 3(a)). There is almost no difference between ERA5 and reference measurements below 20 mm accumulation, but it seriously underestimates higher rainfall values. For higher quantiles (see Fig. 3(b)), the ECDFs of CFSR and TMPA-V7 are relatively closer to the rain gauge distribution whereas ERA5 and TMPA-RT exhibit the poorest agreement with the rain gauge ECDF. To further quantify this comparison, Table 2 compares some error statistics for sample means and select quantiles. Here, mean ($\bar{P}$), standard deviation (SD) and bias ratio (BR) are determined by applying Equations (1)–(3):

$$\bar{P} = \frac{\sum_{i=1}^{N} P_i}{N}$$  \hspace{1cm} (1)
\[ SD = \sqrt{\frac{\sum_{i=1}^{N} (P_i - \bar{P})^2}{N - 1}} \] (2)

\[ BR = \frac{\sum_{i=1}^{N} P_{S_i}}{\sum_{i=1}^{N} P_{G_i}} \] (3)

where \( P_i \) is the reference or evaluated precipitation data in mm and \( N \) is the number of reference or evaluated reanalysis/satellite data. The terms \( P_S (= P_{S_i}) \) and \( P_G (= P_{G_i}) \) are the satellite (or reanalysis) and rain gauge measurements (in mm), respectively. A bias ratio of less than 1 denotes underestimation, while that greater than 1 exhibits overestimation.

In terms of the bias ratio, ERA5 underestimates the rainfall mean by 20%, whereas CFSR, TMPA-RT and TMPA-V7 overestimate it by 18%, 30% and 10%, respectively. From Table 2 it is also evident that the differences between the reanalysis/satellite product and gauge accumulations increase for the larger quantiles, except for the case of CFSR, which indicates a decreasing overestimation as the quantile precipitation threshold increases. The percentage overestimation also decreases with the quantile threshold for TMPA-RT, but it increases for TMPA-V7. On the other hand, only ERA5 shows underestimation and it starts to increase after the 50th percentile.

### 3.2 Spatial distribution of product skill: Process

To examine the spatial distribution of the performance of the climate reanalysis and satellite precipitation products with respect to the 50% and 75% quantiles of the rain gauge data, the following quantile skill metrics were used:
• bias ratio (BR) which presents the ratio of the total reanalysis or satellite measurements to the total reference observations;

• mean absolute error (MAE) indicates mean error between model or satellite and gauge data;

• probability of detection (POD; Wilks 2006) refers to the ratio of the number of correctly detected rainfall events to the total number of rainfall events detected by rain gauge data;

• volumetric hit index (VHI) shows the ratio of the volume of correctly identified precipitation to the total volume of reference observations;

• false alarm ratio (FAR) represents the ratio of the number of falsely detected precipitation events to the total number of precipitation events detected by reanalysis or satellite estimates; and

• volumetric false alarm ratio (VFAR) which denotes the ratio of the volume of falsely identified rainfall to the total volume of model or satellite rainfall data.

The equations for each of the skill metrics are as follows (where \( t \) is the quantile threshold, e.g. 50% quantile) (e.g. AghaKouchak and Mehran 2013, Islam 2018):

\[
QBR = \frac{\sum_{i=1}^{N} P_S(P_S \geq t \mid P_G \geq t)}{\sum_{i=1}^{N} P_G(P_S \geq t \mid P_G \geq t)}
\]

(4)

\[
QMAE = \frac{\sum_{i=1}^{N} \text{abs}(P_S(P_S \geq t \mid P_G \geq t) - P_G(P_S \geq t \mid P_G \geq t))}{N}
\]

(5)

\[
QPOD = \frac{N_H}{N_H + N_M}
\]

(6)
\[ QVHI = \frac{V_H}{V_H + V_M} \]  
(7) 

\[ QFAR = \frac{N_F}{N_H + N_F} \]  
(8) 

\[ QVFAR = \frac{V_F}{V_H + V_F} \]  
(9) 

with 

\[ N_H = \sum_{i=1}^{N} I((P_S > 0 \land P_G > 0) \land (P_S \geq t \mid P_G \geq t)) \] 

\[ N_M = \sum_{i=1}^{N} I((P_S = 0 \land P_G > 0) \land (P_S \geq t \mid P_G \geq t)) \] 

\[ N_F = \sum_{i=1}^{N} I((P_S > 0 \land P_G = 0) \land (P_S \geq t \mid P_G \geq t)) \] 

\[ V_H = \sum_{i=1}^{N} P_S((P_S > 0 \land P_G > 0) \land (P_S \geq t \mid P_G \geq t)) \] 

\[ V_M = \sum_{i=1}^{N} P_G((P_S = 0 \land P_G > 0) \land (P_S \geq t \mid P_G \geq t)) \] 

\[ V_F = \sum_{i=1}^{N} P_S((P_S > 0 \land P_G = 0) \land (P_S \geq t \mid P_G \geq t)) \] 

Here, \( I \) denotes the indicator function and abs and | are used as absolute and logical OR operator respectively. A value of 1 exhibits perfect value for POD and VHI, while 0 is the perfect value for FAR and VFAR. 

The correlation coefficient (CC) represents the scale of agreement between reanalysis or satellite estimates and rain gauge observations:
\[ CC = \frac{\sum_{i=1}^{N}(P_s - \bar{P}_s)(P_G - \bar{P}_G)}{\sqrt{\sum_{i=1}^{N}(P_s - \bar{P}_s)^2 \sum_{i=1}^{N}(P_G - \bar{P}_G)^2}} \]  

where \( \bar{P}_s \) and \( \bar{P}_G \) show mean satellite or reanalysis and rain gauge observations, respectively. A CC value of +1 denotes perfect agreement between reanalysis or satellite and rain gauge estimates, whereas −1 indicates perfect negative agreement. 0 exhibits no agreement between satellite or reanalysis and gauge measurements. It is worth mentioning here that CC values are estimated considering all the rainfall data greater than or equal to zero.

3.3 Spatial distribution of product skill: Results

3.3.1 Bias ratios

Figure 4 shows the spatial distribution of BR and QBR scores for ERA5, CFSR, TMPA-RT and TMPA-V7 precipitation datasets over the analysis grid boxes in the study area. The first row of Fig. 4 presents the BR for all the rainfall data and the second and third rows show QBR values for rainfall quantile thresholds of 50% and 75%, respectively.

As seen in Fig. 4(a), ERA5 exhibits overestimation for around 54% of the analysis grid pixels while it underestimates for some locations over the hilly, mountainous and coastal regions of the country. Noting that BR values in the range 0.9 to 1.1 indicate reasonable performance of a product, it is seen that the bias ratio of ERA5 ranges between 0.9 and 1.1 across the country for about 25% of analysis grid boxes. According to Fig. 4(d), CFSR overestimates rainfall amounts for almost all the analysis grid boxes with BR > 1.2.
The BR value of TMPA-RT indicates serious overestimation for approximately 81% of analysis grid pixels (see Fig. 4(g)), whereas TMPA-V7 shows relatively better performance throughout the country (BR between 0.9 and 1.1 for nearly 36% analysis grid boxes). The gauge-based correction in the TMPA-V7 research grade product brings a remarkable improvement over TMPA-RT real time product. Again, both TMPA products demonstrate reasonably accurate estimations for some analysis grid boxes over the hilly, mountainous and coastal areas.

The QBR values of ERA5 reduce substantially as the extreme rainfall threshold increases (compare Fig. 4(b) and (c) with Fig. 4(a)). ERA5 shows underestimation for only 21% of analysis grid pixels for all data analysis, while it underestimates for around 49% and 85% of analysis grid pixels above the 50th and 75th percentiles, respectively. The overestimation by CFSR also decreases with the quantile threshold, i.e. 94%, 91% and 76% of analysis grid pixels show overestimation for all data, the 50th percentile and the 75th percentile, respectively.

It is evident that with an increase in quantile threshold, QBR scores of TMPA-RT and TMPA-V7 increase persistently and this increment is more prominent for TMPA-V7 (compare Fig. 4(j) with Fig. 4(l)). The overestimation of high rainfall values by TMPA-V7 consists of 63% of analysis grid boxes for the 75% quantile, whereas this value is 51% when all data are included in the analysis.

3.3.2 Mean absolute error

Figure 5 displays mean absolute error (MAE) and quantile mean absolute error (QMAE) for the model-based reanalysis and satellite precipitation products. Among the rainfall datasets, ERA5 and TMPA-RT have the lowest and highest MAE, respectively (see the first row of Fig. 5). It should be noted that ERA5, CFSR, TMPA-V7 are gauge-adjusted products, while TMPA-RT are not gauge-adjusted. Although the gauge
adjustment of TMPA-V7 reduces MAE for some analysis grid pixels in the north western part, it does not change the errors considerably over other parts of the study domain (compare Fig. 5(j) with Fig. 5(g)). For all precipitation datasets, the QMAE increases with the increase of extreme quantile threshold. The results reveal that around 21% analysis grid pixels are above 40 mm QMAE for ERA5, while the corresponding figures for CFSR, TMPA-RT and TMPA-V7 are 49%, 77% and 62%, respectively (see the third row of Fig. 5). Figure 5 also indicates that quantile mean absolute error is comparatively higher over the hilly, mountainous and highland areas for the rainfall products, irrespective of the choice of extreme precipitation threshold.

3.3.3 Rainfall detection metrics

Figure 6 presents POD and QPOD scores for ERA5, CFSR, TMPA-RT and TMPA-V7 rainfall products. The first row of Fig. 6 shows that ERA5 and CFSR demonstrate superior performance in detecting the occurrence of rainfall (POD lies between 0.9 and 1.0 for all the analysis grid boxes). On the other hand, TMPA-RT and TMPA-V7 exhibit lower probability of detection around the central part of the country. Comparing the POD values of all products indicates that TMPA-RT has the lowest score (around 53% analysis grid pixels have POD in the range of 0.4–0.6), while TMPA-V7 is subject to higher POD than that of TMPA-RT (nearly 71% analysis grid pixels are above 0.6). As the quantile threshold increases, the QPOD values increase consistently for both TMPA products throughout the study area. Above 75% quantile, TMPA-RT and TMPA-V7 show QPOD values over 0.8 for approximately 77% and 80% analysis grid boxes, respectively. However, the model-based reanalysis datasets ERA5 and CFSR outperform the TMPA products at higher quantile thresholds. The results are in agreement with the findings of Islam (2018) for other satellite precipitation estimates over Bangladesh.
The three rows in Fig. 7 represent VHI, QVHI above the 50th percentile and QVHI above the 75th percentile, respectively. The QVHI scores display the volume of precipitation identified correctly by reanalysis and satellite rainfall products above a threshold ($t$). One can see that ERA5 and CFSR show similar results in terms of VHI (QVHI) and POD (QPOD). However, the VHI values of TMPA-RT and TMPA-V7 products are higher than the corresponding POD values and the increase is more significant for both products over the central western region (compare the first rows of Figs 6 and 7). As for the QPOD plots, QVHI of TMPA datasets increase as the extreme precipitation threshold increases. This indicates that the detection skills of the satellite products increase at higher rainfall thresholds. While TMPA-RT has higher QVHI (51% of analysis grid pixels are above 0.80) than the corresponding QPOD (38% of analysis grid pixels are above 0.80) for the 50th percentile, TMPA-V7 indicates lower QVHI (68% of analysis grid pixels are above 0.80) than the QPOD (80% of analysis grid pixels are above 0.80) for the 75th percentile. The findings verify that both POD and VHI are necessary for the evaluation of precipitation products.

3.3.4 False alarm metrics

The four columns of Fig. 8 display the false alarm metrics (FAR and QFAR) of ERA5, CFSR, TMPA-RT and TMPA-V7 precipitation datasets, respectively. FAR determines how often the satellite estimates identify rainfall when in reality there is no rain. Among the rainfall products, TMPA-V7 has the lowest FAR values (57% of analysis grid pixels are below 0.40), whereas CFSR has the highest scores (44% of analysis grid pixels are below 0.40). However, the areal distributions of FAR are almost similar for all precipitation datasets. As the quantile thresholds increase, QFAR values of the reanalysis and satellite precipitation estimates reduce. Further, the reduction is more prominent for ERA5 with respect to other rainfall datasets. Above the 50th percentile,
ERA5, CFSR, TMPA-RT and TMPA-V7 have QFAR scores below 0.20 for around 63%, 40%, 37% and 42% of analysis grid pixels, respectively. The corresponding figures of QFAR for the 75th percentile are 94%, 64%, 46% and 56%.

Figure 9 presents the volumetric false alarm metrics (VFAR and QVFAR) for the selected precipitation products. QVFAR indicates the volume of incorrectly detected rainfall by reanalysis or satellite datasets above a threshold ($t$). The VFAR values of the rainfall datasets are almost identical (see the first row of Fig. 9). The VFAR scores of ERA5, CFSR, TMPA-RT and TMPA-V7 are below 0.40 for about 71%, 67%, 70% and 72% of analysis grid boxes, respectively. It is evident that QVFAR reduces noticeably when considering high quantile thresholds (97%, 89%, 83% and 86% analysis grid pixels of ERA5, CFSR, TMPA-RT and TMPA-V7, respectively, are below 0.40 for the 75% quantile).

One can also see that VFAR data are lower than the respective FAR scores (see the first rows of Figs 8 and 9). This denotes that the volumetric skills of the precipitation products are higher than the relevant categorical skills. However, all precipitation estimates show higher QVFAR than the corresponding QFAR values for higher rainfall accumulations (see the second and third rows of Figs 8 and 9).

3.3.5 Temporal detection skills

To investigate the temporal detection skills, we estimate the temporal correlation coefficients (CC) at the analysis grid pixels over the study region (Fig. 10). With respect to the CC values, ERA5 shows relatively better skill throughout the country. However, all the rainfall datasets indicate comparatively higher CC values over the hilly, mountainous and coastal areas. It should be noted that CC values are computed considering all the rainfall data greater than or equal to zero. The results show that almost 27% of analysis grid pixels of ERA5 have CC values above 0.50, while the
corresponding percent of pixels for CFSR, TMPA-RT and TMPA-V7 are 12%, 20% and 21%, respectively.

3.3.6 Discussion of seasonal performance

Figure 11 highlights the performance of the model-based and satellite-derived rainfall estimates in terms of various statistical indices during pre-monsoon (March–May), monsoon (June–September) and post-monsoon (October–November) seasons for the 5-year study period. The values of the statistics are plotted against different precipitation thresholds (i.e. all data, 10th, 25th, 50th, 75th, 90th and 95th percentiles).

As seen in Fig. 11(a), all rainfall datasets overestimate the rainfall values for the pre-monsoon summer season. For TMPA-RT and TMPA-V7, the overestimation increases as the quantile threshold increases, while ERA5 exhibits underestimation for higher rainfall amounts above the 75th percentile. CFSR shows reasonably good correspondence with the gauge observations for lower rainfall accumulations but overestimates the higher rainfall amounts above the 75th percentile.

During the monsoon season, TMPA-V7 implies better estimates than those of ERA5, CFSR and TMPA-RT (see Fig. 11(b)). ERA5 shows serious underestimation, whereas CFSR and TMPA-RT indicate severe overestimation over the monsoon months. Moreover, the underestimation of ERA5 and overestimation of TMPA-RT become more significant with the increase of extreme quantile threshold.

According to the evaluated results during the post-monsoon season, both TMPA products are in better agreement with the rain gauge observations (see Fig. 11(c)). In fact, the best scores of BR and QBR are associated with TMPA-V7 over the post-monsoon months. However, TMPA-RT is as good as TMPA-V7 for this season. Furthermore, the worst values of BR and QBR are related to ERA5. Although, CFSR overestimates low rainfall amounts, it exhibits relatively better performance in capturing
the extreme rainfall values above the 75th percentile. As shown in Fig. 11(d)–(f), all rainfall products show almost identical MAE and QMAE values for the three seasons. One can see that, for higher quantiles, the QMAE values are also higher.

Model-based ERA5 and CFSR and satellite-derived TMPA-RT and TMPA-V7 indicate two different patterns with respect to POD and QPOD over pre-monsoon, monsoon and post-monsoon months, although the scores for ERA5 and CFSR are much higher than those of TMPA-RT and TMPA-V7, particularly for low rainfall accumulations. In terms of FAR and QFAR, ERA5 gives slightly better performance during the monsoon and post-monsoon seasons. Again, with the increase of extreme thresholds, the QFAR values of all products decrease over all seasons.

4 Discussion and conclusions

Precipitation is one of the most important climate variables and accurate estimation of this variable is indispensable for forecasting and monitoring extreme weather. Unfortunately, surface rainfall measurement facilities (either rain gauge or radar) are very scant in Bangladesh. Precipitation products based on global model reanalysis and satellites may provide viable source of information in Bangladesh. However, it is necessary to validate these rainfall datasets against existing reference observations before incorporating the data into decision-making systems. Therefore, the aim of this study was to assess the performance of two model reanalysis (i.e. ERA5 and CFSR) and two satellite (i.e. TMPA-RT and TMPA-V7) precipitation estimates over Bangladesh.

The distribution of rain gauge, reanalysis and satellite data shows that CFSR and TMPA-RT overestimate at all rainfall accumulations. However, CFSR shows comparatively better performance to detect the high rainfall values, while ERA5 severely underestimates. For higher thresholds, ERA5 and TMPA-RT have the worst correspondence with the rain gauge distribution compared to the other products.
Overall, TMPA-V7 performs better than other precipitation estimates for all ranges of accumulation. As expected, the gauge-adjusted TMPA-V7 exhibits noticeable improvements over TMPA-RT. When all data are included in the computation, the estimated bias ratio scores display underestimation for ERA5 and overestimation for CFSR, TMPA-RT and TMPA-V7. Yang et al. (2016) observed that TMPA-RT and TMPA-V7 overestimated rainfall values over a humid basin in China at a daily time scale. Furthermore, the discrepancies between reanalysis/satellite and reference observations increased with the increase in quantile. As the quantile threshold increases, the percentage overestimation for TMPA-V7 and the percentage underestimation for ERA5 increased, while the percentage overestimations for CFSR and TMPA-RT decreased.

With regard to bias ratio, ERA5 overestimates rainfall for several analysis grid boxes throughout the country. However, this dataset performs reasonably well for nearly a quarter of the grids. CFSR also exhibits severe overestimation for nearly all analysis grid pixels over the country.

The TMPA-RT significantly overestimates precipitation values over the flat regions and indicates comparatively better performance over the hilly, mountainous and coastal areas. Romilly and Gebremichael (2011) pointed out that TMPA-RT overestimates precipitation in the low elevation areas of the Blue Nile basin, whereas it displays improved results in high-elevation areas. Khan et al. (2014) showed that TMPA-RT overestimates rainfall values in the lower flat region of Pakistan. PMW sensors infer rainfall rates from ice scattering at the upper surface of the clouds. The huge overestimation by TMPA-RT in the lower flat regions may be attributed to the following factors. Firstly, this part of the country is relatively drier and has lower humidity. As a consequence, although rainfall is detected aloft by PMW or IR sensors,
it may fully or partially evaporate before reaching the ground, leading to severe overestimation of surface rainfall. Secondly, the deep convection may increase ice aloft, which is considered as precipitation by PMW sensors.

TMPA-V7 performs better throughout the country compared to the other precipitation datasets. The reasonably accurate estimation of TMPA-V7 is attributed to the inclusion of rain gauge observations in the product. CFSR exhibits significant overestimation for higher quantiles over the study area, even though the detection skill increases for several analysis grid pixels. Overall, all the products lose their detection skills with the increase of extreme rainfall thresholds. The results indicate that none of these products is good for detecting extremes over the current study area.

According to the MAE values, ERA5 display good performance across the study domain, followed by CFSR, while TMPA-RT has the worst performance. As the extreme precipitation threshold increases, the QMAE rises continuously for all rainfall products. However, all the rainfall datasets indicate relatively higher MAE and QMAE over hilly, mountainous and highland regions, while these statistics are lower over other parts of the country.

The obtained values of POD show that climate reanalyses, ERA5 and CFSR, have good skill, whereas TMPA-RT has the poorest. For both TMPA precipitation estimates, QPOD increases significantly as the quantile threshold increases. At higher rainfall accumulations, TMPA precipitation products also show good performance in detecting the occurrence of rainfall. Reanalysis datasets still outperform the satellite-based estimates. On the other hand, the evaluated VHI scores of the TMPA rainfall estimates are higher than the respective POD values. The noticeable increments are observed for TMPA-RT and TMPA-V7 over the central-western parts of the country. For both TMPA products, QVHI also increases with the precipitation threshold similar
to QPOD. ERA5 and CFSR exhibit identical results with respect to VHI (QVHI) and POD (QPOD).

In terms of FAR, TMPA-V7 performs better over the study area, while CFSR is the product with a poorer performance. For higher quantile thresholds, QFAR values of all rainfall products reduce substantially. Among the rainfall datasets, ERA5 shows the lowest QFAR values, whereas TMPA-RT denotes the highest scores. Although, almost similar results are observed for all precipitation datasets with respect to VFAR (QVFAR) and FAR (QFAR), the VFAR values are considerably lower than those of FAR for all rainfall estimates. Again, QVFAR scores are higher than the corresponding QFAR values for all precipitation datasets.

With respect to CC values, ERA5 shows better skill in detecting temporal variation of rainfall accumulations. However, each product indicates higher CC values over the hilly, mountainous and coastal regions of the country.

The presented statistical results underline the seasonal variation of the products’ skill over the study area. The estimated BR and QBR values of CFSR, TMPA-RT and TMPA-V7 display overestimation during pre-monsoon and monsoon periods. ERA5 overestimates low rainfall amounts during the pre-monsoon season, whereas the product underestimates the high rainfall values. ERA5 also shows serious underestimation over monsoon months. Considering the BR and QBR scores, TMPA-V7 exhibits the best performance during the post-monsoon period, although TMPA-RT is as good as TMPA-V7. CFSR also shows better skill to detect higher rainfall accumulations during this season. The estimated MAE and QMAE of the rainfall datasets indicate almost identical performance during the three seasons. It is also noted that MAE and QMAE increase with the increase of extreme precipitation thresholds. The reported POD and QPOD values show better performance of the model-based reanalysis precipitation
products than those of satellite-derived datasets. Among the rainfall products, ERA5 indicates relatively better skill with respect to FAR and QFAR during monsoon and post-monsoon periods. For all products, the QFAR values decrease consistently as the extreme quantile thresholds increase over all three seasons.

Bias ratio and MAE are used to assess the performance of the reanalysis and satellite datasets in estimating the amount of rainfall, whereas POD, VHI, FAR and VFAR are applied to evaluate the rainfall event detection skills. According to the obtained results, the gauge-adjusted TMPA-V7 shows better performance than the three other precipitation data products, while TMPA-RT exhibits the worst performance over Bangladesh. However, TMPA-V7 shows serious overestimation for higher rainfall amounts. The MAE and QMAE values of TMPA-V7 are also higher than those of ERA5 and CFSR. Again, TMPA-V7 indicates relatively lower POD (QPOD) and VHI (QVHI) values than those of ERA5 and CFSR. Furthermore, ERA5 shows lower QFAR and QVFAR and higher CC values than TMPA-V7.

The findings of this research indicate that accurate estimation of high to extreme rainfall amounts remains the main challenge for all the investigated precipitation products. One possible approach to reduce biases and errors of the precipitation products can be to merge these rainfall estimates with available rain gauge observations at the local scale (e.g. Dinku et al. 2007). Therefore, applying appropriate bias adjustment techniques, these rainfall datasets can be used in flood forecasting, landslide warning, drought monitoring, demand optimization of water resources, irrigation purposes and verification of climate model precipitation outputs over the study area.

Acknowledgements

Rain gauge data were collected from Bangladesh Meteorological Department and Bangladesh Water Development Board and are thankfully acknowledged. The authors appreciate the
anonymous reviewers for their valuable comments that improved the quality of the paper. The first author also acknowledges Griffith University, Australia for providing financial support through GUIPRS and GUPRS scholarships.

References


Prakash, S., et al., 2016a. From TRMM to GPM: How well can heavy rainfall be detected from space? *Advances in Water Resources*, 88 (Supplement C), 1-7. doi:10.1016/j.advwatres.2015.11.008


**Table 1.** Summary information of various precipitation products studied here. The PMW/IR and Gauge columns denote whether the product includes passive microwave/infrared or rain gauge estimates (Y and N indicate yes and no, respectively).

<table>
<thead>
<tr>
<th>Product</th>
<th>Temporal resolution</th>
<th>Spatial resolution</th>
<th>Time slice</th>
<th>Spatial coverage</th>
<th>PMW/IR</th>
<th>Gauge</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERA5</td>
<td>Hourly</td>
<td>0.25°</td>
<td>2000 to present</td>
<td>90°S–90°N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CFSR</td>
<td>6-hourly</td>
<td>0.50°</td>
<td>1979 to present</td>
<td>90°S–90°N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>TMPA-RT</td>
<td>Daily</td>
<td>0.25°</td>
<td>2000 to present</td>
<td>50°S–50°N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>TMPA-V7</td>
<td>Daily</td>
<td>0.25°</td>
<td>1998 to present</td>
<td>50°S–50°N</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

**Table 2.** Different statistics of rain gauge, reanalysis, and satellite rainfall data for daily temporal accumulation during January 2010–December 2014. SD: standard deviation; Q: quantile. Values in parentheses show percentage difference between satellite and rain gauge measurements for the respective quantile. Negative and positive signs represent under- and overestimation, respectively.

<table>
<thead>
<tr>
<th>Product</th>
<th>Mean (mm)</th>
<th>SD (mm)</th>
<th>Bias ratio</th>
<th>Q10 (mm)</th>
<th>Q25 (mm)</th>
<th>Q50 (mm)</th>
<th>Q75 (mm)</th>
<th>Q90 (mm)</th>
<th>Q95 (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain gauge</td>
<td>23.1</td>
<td>28.3</td>
<td>--</td>
<td>3.0</td>
<td>6.3</td>
<td>14.3</td>
<td>29.7</td>
<td>53.0</td>
<td>75.2</td>
</tr>
<tr>
<td>ERA5</td>
<td>18.4</td>
<td>17.7</td>
<td>0.80</td>
<td>(±27%)</td>
<td>(+19%)</td>
<td>13.5</td>
<td>23.0</td>
<td>37.7</td>
<td>51.4</td>
</tr>
<tr>
<td>CFSR</td>
<td>27.3</td>
<td>28.2</td>
<td>1.18</td>
<td>(+87%)</td>
<td>(+71%)</td>
<td>(+35%)</td>
<td>(+10%)</td>
<td>(+8%)</td>
<td>(+4%)</td>
</tr>
<tr>
<td>TMPA-RT</td>
<td>30.0</td>
<td>32.9</td>
<td>1.30</td>
<td>(+50%)</td>
<td>(+29%)</td>
<td>(+27%)</td>
<td>(+35%)</td>
<td>(+35%)</td>
<td>(+27%)</td>
</tr>
<tr>
<td>TMPA-V7</td>
<td>25.4</td>
<td>30.3</td>
<td>1.10</td>
<td>(+7%)</td>
<td>(+2%)</td>
<td>(+3%)</td>
<td>(+11%)</td>
<td>(+16%)</td>
<td>(+13%)</td>
</tr>
</tbody>
</table>
Figure 1. Elevation map of the study domain and locations of rain gauges used in the study. Elevation values are in m above MSL. Data source: 1 arc-second digital elevation model (DEM) from https://earthexplorer.usgs.gov/.
Figure 2. Number of rain gauges for each 5-m elevation bin is plotted against the centre elevation value of the bin. Elevation (m) values were interpolated from the 1 arc-s DEM data shown in Fig. 1.
Figure 3. Empirical cumulative distribution functions of rain gauge, reanalysis and satellite datasets considering days with rainfall >1 mm for: (a) 0–250 mm/d and (b) zoomed into the accumulation range 100–250 mm/d. In the x-axis, rainfall rate is overlapped with different scales in the y-axis.
Figure 4. Bias ratio (BR) and quantile bias ratio (QBR) of reanalysis and satellite data (here, Q means quantile) relative to gauge data. White cells represent no gauges in the pixels.
Figure 5. Mean absolute error, MAE (mm), and quantile mean absolute error, QMAE (mm), of reanalysis and satellite data (here, Q means quantile) relative to gauge data. White cells represent no gauges in the pixels.
Figure 6. Probability of detection (POD) and quantile probability of detection (QPOD) of reanalysis and satellite data (here, Q means quantile). White cells represent no gauges in the pixels.
Figure 7. Volumetric hit index (VHI) and quantile volumetric hit index (QVHI) of reanalysis and satellite data (here, Q means quantile). White cells represent no gauges in the pixels.
Figure 8. False alarm ratio (FAR) and quantile false alarm ratio (QFAR) of reanalysis and satellite data (here, Q means quantile). White cells represent no gauges in the pixels.
Figure 9. Volumetric false alarm ratio (VFAR) and quantile volumetric false alarm ratio (QVFAR) of reanalysis and satellite data (here, Q means quantile). White cells represent no gauges in the pixels.

Figure 10. Correlation coefficient (CC) values of reanalysis and satellite products using all data greater than or equal to zero. White cells represent no gauges in the pixels.
Figure 11. Distribution of statistical matrices for different quantile thresholds during pre-monsoon, monsoon and post-monsoon periods.