Statistical comparison of satellite-retrieved precipitation products with rain gauge observations over Bangladesh

Md. Atiqul Islam

Assistant Professor, Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna-9203, Bangladesh

Corresponding author:

Name: Md. Atiqul Islam

Email: atiqul@ce.kuet.ac.bd

Address: Assistant Professor, Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna-9203, Bangladesh

Statistical comparison of satellite-retrieved precipitation products with rain gauge observations over Bangladesh

In this investigation, six satellite-derived precipitation products namely Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN), Climate Prediction Centre (CPC) Morphing Technique (CMORPH), Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG) final run both non gauge-calibrated (IMERG) and gauge-calibrated (IMERG-GC), and Global Satellite Mapping of Precipitation (GSMaP) for GPM both non gauge-calibrated (GSMaP) and gauge-calibrated (GSMaP-GC) are evaluated over Bangladesh, using ground-based rain gauge observations as reference over a 3 years period from January 2014 to December 2016. Nine widely used categorical and volumetric statistical matrices such as bias, probability of detection (POD), volumetric hit index (VHI), false alarm ratio (FAR), volumetric false alarm ratio (VFAR), critical success index (CSI), volumetric critical success index (VCSI), miss index (MI), and volumetric miss index (VMI) are employed to exploit the performance of the precipitation products in detecting extremes above different quantile thresholds (i.e., 50%, 75%, and 90% quantiles) for various temporal window (i.e., 3 hours, 6 hours, 12 hours, and 24 hours). The bias values show that none of the satellite rainfall data sets are ideal for detecting extreme rainfall accumulations. In fact, all products lose their detection skills consistently as the extreme precipitation thresholds (50%, 75%, and 90% quantiles) increase. The results indicate that PERSIANN shows the worst performance over the study region. Overall, GSMaP-GC performs better than the other precipitation products. However, the FAR values of GSMaP are also higher over monsoon and post-monsoon months. The categorical and volumetric scores reveal that the detection skill increases remarkably for all rainfall data sets throughout the year with the increase of extreme quantile thresholds. At higher temporal accumulations, the detection capability of the products also improve considerably, and this improvement is more significant during monsoon period. The performance is relatively poor for all precipitation data sets over the cold months. In general, all six satellite precipitation products are doing well in detecting the occurrence of rainfall, but are not so good in estimating the amount of rainfall.

Keywords: Rain gauge observation, satellite-based precipitation, Bangladesh, extreme

1. Introduction

Extensive and reliable observation of precipitation is the prerequisite for their application in early warning systems, disaster mitigation policies, and water resources allocation problems. The availability of high-quality precipitation data is very limited in most of the developing countries such as Bangladesh. The rain gauge network is very sparse in Bangladesh, there are only 35 rain gauges for 147,570 km² of area. Heavy (extreme) rainfall events during monsoon season often bring severe damages due to inadequate ground-based rainfall measurement facilities (Yilmaz et al. 2005). In such situation, satellite-retrieved precipitation estimates with near global coverage and real time availability can be used as a potential alternative source of rainfall information over un-gauged or sparsely gauged basins. In recent years, the popularity of satellite-based rainfall estimates has increased due to their finer spatial ($\leq 0.25^\circ$) and temporal (≤ 3 hours) resolution, long period of records (more than 10 years), and public availability (Zulkafli et al. 2014; Cattani, Merino, and Levizzani 2016).

At present, several high resolution satellite precipitation products are available, e.g. Global Precipitation Measurement (GPM) mission (Hou et al. 2014; Huffman et al. 2015; Huffman, Bolvin, and Nelkin 2015; Huffman 2016), the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA; Huffman et al. 2007), Global Satellite Mapping of Precipitation (GSMaP) for GPM (Okamoto et al. 2005; Kubota et al. 2007, Ushio et al. 2009; Aonashi et al. 2009), Climate Prediction Centre (CPC) Morphing technique (CMORPH; Joyce et al. 2004), the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Sorooshian et al. 2000), the PERSIANN Cloud Classification System

estimation (PERSIANN-CCS; Hong et al. 2004), and so on. Satellite precipitation estimates are based on active and passive microwave (PMW) observations from low Earth orbiting (LEO) satellites and visible infrared (Vis-IR) measurements from Polar orbiting geostationary (GEO) satellites (Sapiano and Arkin 2009). Precipitation estimation from microwave sensors are more accurate than the infrared observations as the former observation method is directly related to the size and phase of the hydrometeors present in the atmosphere (Mei et al. 2014). Again, the sampling frequency of microwave sensors is lower than that of infrared sensors. Therefore, information from both microwave and infrared sensors are combined to increase the spatial and temporal coverage of the satellite precipitation products. Consequently, this approach takes the strength of the individual sensor and improves the accuracy and consistency of the precipitation estimation algorithm. However, due to the indirect nature of the sensors, errors and uncertainty of satellite precipitation products remain as an intrinsic limitation (Yang et al. 2015; Gebregiorgis and Hossain 2014). The biases and errors also vary significantly in terms of locations, seasons, topography, and climate conditions (Dinku, Ceccato, and Connor 2011; Sorooshian et al. 2011).

Thus, it is essential to evaluate the satellite-retrieved precipitation products against ground-based measurements before applying the data sets in hydrological applications. So far, numerous studies have been performed to validate and verify satellite-based rainfall estimates at different spatiotemporal scales and climate regimes using various methods and techniques (Dinku et al. 2007; Hughes 2006; AghaKouchak et al. 2011; Behrangi et al. 2011; Zhou et al. 2014). Thiemeig et al. (2012) conducted a comprehensive evaluation of six satellite-derived precipitation products i.e., TMPA 3B42-V6, PERSIANN, CMORPH, GSMaP-MVK, Rainfall Estimation Algorithm, version 2.0 (RFE 2.0), and Goddard Profiling Algorithm, version 6.0 (GPROF 6.0) over

four sparsely-gauged basins in continental Africa. They concluded that the satellite products underestimated the heavy rainfall accumulations, particularly over semiarid regions. Meanwhile, the highest performance was observed over the tropical wet and dry zones. Vernimmen et al. (2012) investigated the performance of three satellite rainfall products (TMPA 3B42-RT, PERSIANN, and CMORPH) by comparison with ground-based observations for real time drought monitoring in Indonesia. The study pointed out that a single bias correction equation for satellite data sets was adequate in terms of accuracy, coverage, and cost effectiveness for drought monitoring in comparison with only in-situ measurements. It is worth mentioning here that there is very limited literature related to the validation and verification of satellite-derived precipitation products over Bangladesh, which is the main focus of this work. For example, Rahman et al. (2012) compared rain gauge measurements with TMPA 3B42 and the European Centre for Medium-Range Weather Forecasts (ECMWF) satellite rainfall estimates at only three locations (Dinajpur, Rangpur, and Sylhet) for the years of 2004 and 2006. The findings showed that ECMWF data sets were more reliable than TMPA 3B42 for flood forecasting related purposes.

Clearly, there is a lack of validation works of satellite-retrieved precipitation estimates over Bangladesh for floods and drought monitoring, landslides modelling, data assimilation into weather models, and verification of regional climate models. In this study, six satellite rainfall products are evaluated against rain gauge observations over the country at different temporal scales. Bangladesh receives extreme rainfall during monsoon period and monsoon rainfall shows significant spatial and temporal variability over the study region. This spatiotemporal variability of rainfall are associated with topography, slope, and other geographic factors. Hence, the study area is expected to be a good platform for testing the detection skills of different satellite

precipitation products in the sub-tropical monsoon climate. Moreover, it is not only necessary to evaluate the satellite precipitation data over Bangladesh but also to identify which precipitation products show better performance in detecting heavy rainfall accumulations as the products exhibit space and seasonal dependency. Therefore, the main objective of the present study is to assess the capability of several satellite precipitation estimates to detect the extreme rainfall accumulations at different temporal scales. The findings will surely help to understand at what temporal scale satellite data sets show better performance in detecting heavy rainfall accumulations over the study area. In hydrologic modelling, one may want to know at what temporal scale the input precipitation data sets are associated with less uncertainty as the biases and errors of input data affect the model predictions.

The current research validates and intercompares two groups of satellite rainfall data sets. The first group has relatively coarser spatial (0.25°) and temporal (1 hour and 3 hours) resolution. This group consists of the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) and the Climate Prediction Centre (CPC) Morphing technique (CMORPH) satellite precipitation products. The second group includes precipitation estimates with finer spatial (0.10°) and temporal resolution (half-hourly and 1 hour). This group includes two variables i.e., precipitation uncalibrated (not gauge-adjusted) and precipitation calibrated (gauge-adjusted) of the Integrated Multi-satellitE Retrievals for GPM (IMERG) half-hourly final run version 04A and two variables i.e., hourly precipitation rate (not gauge-adjusted) and gauge-calibrated hourly precipitation rate (gauge-adjusted) of the Global Satellite Mapping of Precipitation (GSMaP) for GPM version 04B. The six satellite rainfall products are selected for investigation because they are most widely studied and used. Both IMERG and GSMaP are reputed as the state-of-the-art space-borne

precipitation products and combine rainfall estimates from both PMW and IR sensors. On the other hand, PERSIANN and CMORPH are mostly based on IR and PMW estimates, respectively. This combination of products will enable to explore the capabilities of PMW-only, IR-only, and combined PMW-IR algorithms over Bangladesh. Again, IMERG rainfall estimates include information from other precipitation sources such as PERSIANN and CMORPH. Therefore, comparisons are made between a product (i.e., IMERG) derived from other data sources and the data sources (i.e., PERSIANN and CMORPH) themselves. Moreover, IMERG and GSMaP are adjusted with rain gauge observations while PERSIANN and CMORPH do not incorporate rain gauge data. GSMaP algorithm also integrates topographic data in the orographic and nonorographic rainfall classification scheme (Yamamoto and Shige 2015). These differences in satellite rainfall estimation algorithms will allow a more comprehensive evaluation of the precipitation products over the study region.

2. Study area and data sets

2.1. Study area

The research focuses over Bangladesh which is known as the largest delta (the Ganges delta) in the world. The country is situated in the north-eastern part of Indian sub-continent and covers from 20° N to 27° N and from 88° E to 93° E. Approximately 80% area of the country is flood plain (Dewan 2015). The north-western parts of the country consists of highlands. There are also some hilly regions in the south-eastern (Chittagong Hill Tracts) and north-eastern (Sylhet) parts of Bangladesh (see Figure 1(a)). Four distinct seasons can be recognized in Bangladesh from the climatic point of view: (i) the dry winter season from December to February, (ii) the pre-monsoon hot summer season from March to May, (iii) the rainy monsoon season from June to September, and (iv)

the post-monsoon autumn season which lasts from October to November (Shahid 2010). With a sub-tropical monsoon climate, the study domain receives the highest rainfall during monsoon in the hilly and coastal areas (see Figure 1(b)-1(e) and Figure 2). Minimum rainfall occurs in the drier western part of the country. Although, almost no rainfall occurs during winter months (from December to February), most parts of the country receive at least 1000 mm rainfall per year.

2.2. Rainfall data sets

2.2.1. Rain gauge data

Bangladesh is covered by 35 fully automatic remotely operated rain gauge stations (see Figure 1(a)). The gauges are operated by Bangladesh Meteorological Department (BMD) and provide 3 hourly rainfall accumulations at UTC (Coordinated Universal Time) time scale. Beginning hour of the accumulation periods are 0, 3, 6, 9, 12, 15, 18, and 21 UTC. For example, precipitation accumulation at 6 UTC represents the accumulated rainfall between 03:00:00 and 06:00:00 UTC. The six satellite rainfall estimates are accumulated to 3-hourly totals for comparison with rain gauge observations.

The analysis is based on direct comparison of satellite pixel values with point measurements of rain gauge stations. This means comparing a point observation against an average value within an area of $11\text{ km} \times 11\text{ km}$ (0.10°) or $27\text{ km} \times 27\text{ km}$ (0.25°) grid box and this would affect the validation results. As a relatively long study period (3 years) is applied in this comparison, the spatial representativeness errors of the rain gauge data sets should be minimized. However, several studies employed point-area comparison having at least one rain gauge per 0.25° and 0.10° pixel such as Romilly and Gebremichael (2011), Moazami et al. (2013), Yang et al. (2015), Cohen Liechti et

al. (2012), and Thiemi et al. (2012). No interpolation method is applied in the present study to predict the rain gauge values inside each satellite pixel because the sparse rain gauge density may seriously affect the accuracy of the interpolation. Furthermore, this attempt intends to evaluate the overall performance of the satellite precipitation products over Bangladesh rather than quantify the errors precisely. Therefore, this data set is considered as the most complete, accurate, and representative giving preference to the general data availability over the study region.

2.2.2. *Satellite data*

Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) precipitation data set is based on the IR brightness temperature (T_b) images from global geosynchronous satellites (Janowiak, Joyce, and Yarosh 2001; GOES-8, GOES-10, GMS-5, Metsat-6, and Metsat-7) provided by the Climate Prediction Centre (CPC), National Oceanic and Atmospheric Administration (NOAA). The algorithm generates mean rainfall rate from the gridded IR images using artificial neural network function approximation/classification procedures (Sorooshian et al. 2000). Since the relationships between cloud-top brightness temperatures and rainfall rates have a high degree of uncertainties, the system updates the neural network parameters when independent microwave (MW) precipitation estimates are available from the LEO satellites (e.g. AMSU-B on NOAA satellites and SSM/I aboard DMSP satellite). Geographical coverage of the rainfall product is 60° S-60° N and 180° W-180° E. PERSIANN precipitation products with $0.25^\circ \times 0.25^\circ$ latitude/longitude spatial and 3 hours temporal resolution are employed for investigation in this effort. The main features of the different satellite rainfall products are provided in Table 1.

CPC MORPHing technique (CMORPH) precipitation estimates are also used in this study with 3 hours temporal and $0.25^\circ \times 0.25^\circ$ spatial resolution. The CMORPH

rainfall product is derived from the LEO satellites measured PMW observations exclusively (Joyce et al. 2004). When PMW data are unavailable, motion vector matrices are computed from GEO satellites IR data to transport only the features not the value of PMW observations. The intensity and shape of the precipitation system between two successive PMW scans are determined by applying time-weighted interpolation technique. The system transports rainfall features forward from the previous PMW scan and backward from the next PMW scan. The algorithm is so flexible that it can incorporate PMW observations from any LEO satellite.

After successful completion of Tropical Rainfall Measuring Mission (TRMM; PR data ended on 7 October 2014 and TMI switched off on 8 April 2015; Huffman 2016), Global Precipitation Measurement (GPM) mission started to provide precipitation data from March 2014 to onward. However, the mission plans to provide continuous record of precipitation estimates from the beginning time of TRMM data (i.e., from 1998). The Integrated Multi-satellitE Retrievals for GPM (IMERG) half-hourly final run version 04A precipitation estimates are employed here for evaluation over Bangladesh. Both non gauge-calibrated (IMERG) and gauge-calibrated (IMERG-GC) products are investigated in this study. Spatial resolution of the data set is $0.10^{\circ} \times 0.10^{\circ}$ and covers 60° S- 60° N and 180° W- 180° E. These half-hourly data sets are aggregated in 3 hourly time frame to synchronize with rain gauge accumulations. IMERG integrates precipitation products from two other sources i.e., CMORPH and PERSIANN-CCS. At first, passive microwave (PWM) related precipitation estimates are collected from various LEO satellites. These products are gridded and inter-calibrated to generate half-hourly satellite only precipitation. The zenith-angle-corrected merged infrared (IR) estimates are also collected from various GEO satellites. The microwave-calibrated IR products are then applied to compute displacement vectors.

Now, the gaps between the PMW precipitation estimates are filled with IR estimates in two ways. First, LEO-PMW precipitation products are morphed using displacement vectors when sufficient PMW estimates are available. Second, the GEO-IR data are included using CMORPH-Kalman filter Lagrangian time interpolation scheme if PMW data are too sparse. Finally, these multi-satellite PMW-IR merged half-hourly precipitation products are summed over a calendar month. The monthly gridded rain gauge data sets (from Global Precipitation Climatology Centre, GPCC) are combined with monthly multi-satellite rainfall fields to generate research-grade monthly IMERG products. Furthermore, the ratios of original multi-satellite and satellite-gauge combined precipitation estimates are calculated and incorporated to each half-hourly rainfall fields to create half-hourly research-grade products.

Global Satellite Mapping of Precipitation (GSMaP) for GPM algorithm integrates PMW rainfall estimates from two types of sensors i.e., conical scanning passive microwave imagers and cross-tracking passive microwave sounders from various satellites. When PMW estimates are absent, globally-merged full-resolution IR rainfall data are applied to derive moving vectors using a Kalman filter framework (Ushio et al. 2009). These moving vectors are used to propagate PMW estimates (both forward and backward as like CMORPH algorithm) in time and space, but it also incorporates IR estimates along with the propagated PMW rainfall estimates. This is one of the basic differences between GSMaP and CMORPH precipitation estimation algorithms. Dual polarization radar (DPR) of GPM core observatory satellite is utilized to develop data base and physical models of precipitation and included in the estimation algorithm indirectly. GSMaP algorithm also uses various ancillary data i.e., atmospheric condition derived from the Japan Meteorological Agency (JMA) Global Analysis (GANAL) and Forecast data set, the JMA Merged Satellite and in-situ Global Daily Sea

Surface Temperatures (MGDSST), 30-arc second topographic data from United States Geological survey (USGS), and high resolution ocean, land, and coast data base in the coast detection scheme. Finally, NOAA CPC Unified Gauge-Based Analysis of Global Daily Precipitation data set is used to generate gauge-combined GSMaP precipitation estimates with $0.10^\circ \times 0.10^\circ$ spatial and 1 hour temporal resolution. Geographical coverage of the data set extends between 60° S- 60° N and 180° W- 180° E. This mission started to release precipitation data from March 2014 to present. Both non gauge-calibrated (GSMaP) and gauge-calibrated (GSMaP-GC) versions of the product are investigated here.

3. Methodology

3.1. Preliminary assessment of rainfall data sets

In this study, three years (January 2014-December 2016) of satellite rainfall data are compared against ground-based rain gauge observations to explore the performance of the products. IMERG, IMERG-GC, GSMaP, and GSMaP-GC data start from March 2014. Thus, these data sets are compared with rain gauge data from March 2014 to December 2016. In order to evaluate the satellite data sets at different temporal window, 6 hours, 12 hours, and 24 hours (daily) rainfall accumulations are derived from the 3 hours data. The empirical cumulative distribution functions (ECDFS) are computed for different temporal accumulations using rain gauge and satellite data equal to or greater than 1 mm (see Figures 3 and 4). As noted from Figures 3 and 4, PERSIANN shows the lowest agreement with the rain gauge distributions. Comparing the ECDFS, IMERG, IMERG-GC, GSMaP, and GSMaP-GC are relatively closer to the gauge distributions above 50th percentile. Overall, for higher quantiles (> 50 th percentile) PERSIANN and CMORPH precipitation products exhibit underestimation for all accumulation periods

over the study region. Above 75th percentile, IMERG, IMERG-GC, and GSMaP overestimates the rainfall values whereas GSMaP-GC underestimates the magnitudes for 3 hours, 6 hours, and 12 hours and overestimates for 24 hours accumulation intervals (Figures 3 and 4). The accumulations are lower (< 10 mm; see Table 2) below the 50th percentile and are not considered in the subsequent analysis.

Table 2 displays the summary of the statistics for the entire rain gauge and satellite data. In this table, mean ($\bar{P}$), standard deviation (SD), and bias are defined as follows:

$$\bar{P} = \frac{\sum_{i=1}^N P_i}{N} \quad (1)$$

where P_i is the rain gauge/satellite estimates in mm and N is the number of rain gauge/satellite data.

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

$$\text{Bias} = \frac{\sum_{i=1}^N P_{S_i}}{\sum_{i=1}^N P_{G_i}} \quad (3)$$

where $P_S (= P_{S_i})$ are satellite measurements in mm and $P_G (= P_{G_i})$ are rain gauge observations in mm. Bias of 1 indicates no bias in the estimated satellite data compared with rain gauge observations. Bias greater than 1 represents overestimation while less than 1 shows an underestimation of the satellite measurements. As shown in Table 2, the bias values of PERSIANN and CMORPH indicate underestimation during the study period. The exception is the overestimation by IMERG, IMERG-GC, and GSMaP rainfall estimates. However, the overestimation decreases with the

accumulation time. In terms of bias, GSMaP-GC shows superior performances for all temporal accumulations compared to the other precipitation products. As the accumulation period increases, underestimation becomes more prominent for CMORPH whereas underestimation decreases considerably for PERSIANN satellite data sets. From Table 2 one can also see that PERSIANN and CMORPH satellite accumulations associated with 50th, 75th, 90th, and 95th percentiles show underestimation for all accumulation periods. The exceptions are the overestimations by IMERG, IMERG-GC, and GSMaP rainfall products. However, 50th and 75th percentile values of IMERG data sets for 12 hours and 24 hours intervals are lower than those of rain gauge values. Again, GSMaP-GC overestimates the rainfall for 50th percentile and underestimates for 75th, 90th, and 95th percentiles. Furthermore, as the quantile threshold (extreme precipitation threshold) increases, the difference between rain gauge and satellite measurements increase significantly above 50th percentile for all precipitation products (see the Q75, Q90, and Q95 columns in Table 2). Comparing CMORPH and PERSIANN for 12 hours accumulation time, CMORPH performs better with 5%, 15%, and 25% decreases in underestimation for 75th, 90th, and 95th percentiles, respectively. Comparing GSMaP-GC with CMORPH, GSMaP-GC shows a significant improvement with 17%, 10%, and 7% decreases in underestimation for 75th, 90th, and 95th percentiles, respectively. From Table 2 it is also evident that GSMaP exhibits higher percentages of overestimation than IMERG and IMERG-GC rainfall products in comparison with rain gauge observations i.e., 6%, 14%, and 36% for 75th, 90th, and 95th percentiles, respectively. Moreover, IMERG and IMERG-GC show almost identical percentages of overestimation for 12 hours accumulation period.

3.2. Statistical indices to evaluate satellite products

In-depth understanding of errors and biases of satellite precipitation data with respect to

the reference (rain gauge) observations is inevitable for their application in decision-making systems. In addition to bias four categorical skill indices i.e., probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), and miss index (MI) are introduced to investigate the performance of PERSIANN, CMORPH, IMERG, IMERG-GC, GSMaP and GSMaP-GC satellite precipitation estimates to detect the heavy rainfall events (Wilks 2006; AghaKouchak and Mehran 2013; AghaKouchak et al. 2011). Additionally, four volumetric skill scores i.e., volumetric hit index (VHI), volumetric false alarm ratio (VFAR), volumetric critical success index (VCSI), and volumetric miss index (VMI) are also applied to detect the volumetric skills of the precipitation products for different extreme precipitation thresholds i.e., 50%, 75%, and 90% quantiles of reference data.

The quantile bias (QBias) indicates the ratio of the total satellite precipitation estimates over the total rain gauge observations above a quantile threshold, t (e.g. t is equal to 75th percentile of reference data). When t is equal to zero, QBias is simply bias. The fraction of the number of correct identifications of precipitation to the total number of reference precipitation occurrences is termed as probability of detection (POD). The volumetric hit index (VHI) refers the proportion of the correctly identified volume of precipitation to the total volume of reference (rain gauge) precipitation. The false alarm ratio (FAR) denotes the fraction of the number of falsely identified precipitation to the total number of satellite precipitation occurrences. The volumetric false alarm ratio (VFAR) represents the ratio of the volume of falsely detected precipitation to the total volume of precipitation identified by satellite data. The critical success index (CSI) refers the proportion of the number of correctly identified precipitation by satellite to the total number of correct, miss, and false observations of precipitation. The volumetric critical success index (VCSI) displays the ratio of the

correctly detected volume of precipitation to the total volume of correct, miss, and false observations of satellite precipitation data. The miss index (MI) refers the fraction of the number of observations missed by satellite to the total number of rain gauge estimates. The volumetric miss index depicts the portion of volume missed by satellite data relative to the volume of precipitation observed by reference data. POD, VHI, FAR, VFAR, CSI, VCSI, MI, and VMI above a quantile threshold (t) is known as quantile POD (QPOD), quantile VHI (QVHI), quantile FAR (QFAR), quantile VFAR (QVFAR), quantile CSI (QCSI), quantile VCSI (QVCSI), quantile MI (QMI), and quantile VMI (QVMI) respectively.

$$QBias = \frac{\sum_{i=1}^N P_S((P_S \geq t | P_G \geq t))}{\sum_{i=1}^N P_G((P_S \geq t | P_G \geq t))} \quad (4)$$

$$QPOD = \frac{N_H}{N_H + N_M} \quad (5)$$

$$QVHI = \frac{V_H}{V_H + V_M} \quad (6)$$

$$QFAR = \frac{N_F}{N_H + N_F} \quad (7)$$

$$QVFAR = \frac{V_F}{V_H + V_F} \quad (8)$$

$$QCSI = \frac{N_H}{N_H + N_M + N_F} \quad (9)$$

$$QVCSI = \frac{V_H}{V_H + V_M + V_F} \quad (10)$$

$$QMI = \frac{N_M}{N_H + N_M} \quad (11)$$

$$QVMI = \frac{V_M}{V_H + V_M} \quad (12)$$

In the above Equations N_H , N_M , N_F , V_H , V_M , and V_F are defined as follows:

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

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

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

$$V_H = \sum_{i=1}^N P_S((P_S > 0 \& P_G > 0) \& (P_S \geq t \mid P_G \geq t))$$

$$V_M = \sum_{i=1}^N P_G((P_S = 0 \& P_G > 0) \& (P_S \geq t \mid P_G \geq t))$$

$$V_F = \sum_{i=1}^N P_S((P_S > 0 \& P_G = 0) \& (P_S \geq t \mid P_G \geq t))$$

where P_S and P_G are as defined in Equation (3). I represents the indicator function. The values of QPOD, QVHI, QFAR, QVFAR, QCSI, QVCSI, QMI, and QVMI vary between 0 and 1. 0 represents a perfect score for QFAR, QVFAR, QMI, and QVMI whereas 1 denotes perfect score for QPOD, QVHI, QCSI, and QVCSI. Figures 5-16 are prepared using Equations (4)-(12). In Figures 5-14, horizontal and vertical axes represent the months and the values of the relevant statistical indices, respectively.

4. Results

4.1. Bias

Figures 5 and 6 represent Bias and QBias scores for PERSIANN, CMORPH, IMERG, IMERG-GC, GSMaP, and GSMaP-GC satellite precipitation estimates. Figure 5(a) displays biases when entire data are included in the analysis. Figures 5(b) and 5(c) show

QBias with respect to 75th and 90th percentiles of rain gauge data respectively. It is shown that PERSIANN and CMORPH satellite data sets underestimate the rainfall over the monsoon months June, July, and August for 3 hours temporal accumulation. Moreover, the QBias values of PERSIANN and CMORPH decrease significantly with the increase of extreme precipitation threshold during the rainy monsoon season (i.e., June, July, August, and September; compare Figure 5(a) with Figures 5(b) and 5(c)). However, the magnitudes of these two products increase considerably for each quantile threshold as the accumulation time increases (compare Figures 5(b) and 5(c) with Figures 6(b) and 6(c) respectively). On the other hand, the bias scores of IMERG, IMERG-GC, GSMaP and GSMaP-GC distribute around 1 during monsoon period. As the quantile threshold increases, QBias scores of GSMaP-GC decrease remarkably, while they noticeably increase for IMERG, IMERG-GC, and GSMaP satellite data. For 12 hours accumulations, QBias values of GSMaP do not change significantly with the extreme quantile threshold over the monsoon months (see Figures 6(a) and 6(b)).

Although the six satellite data sets indicate either underestimation or overestimation over cold months (i.e., December, January, and February), the QBias scores of the data sets drop consistently as the extreme precipitation threshold increases. One can also see that all precipitation products exhibit serious overestimation during post-monsoon period (i.e., October and November; see Figure 5(a)). While PERSIANN, CMORPH, IMERG-GC, and GSMaP-GC indicate relatively good agreement with the reference data during pre-monsoon hot summer period (i.e., March, April, and May), IMERG and GSMaP denote overestimation (see Figure 5(a)). For higher quantile thresholds, the results of IMERG-GC and GSMaP-GC are also very promising during this season. It is shown that some points are missing for higher quantiles during cold months. During winter, rainfall accumulations are very low over the study region (see Figure 2).

Therefore, the missing points indicate either both rain gauge and satellite estimates are lower than the extreme precipitation threshold or the values of the statistical indices are equal to zero. If rain gauge observations are below a given quantile threshold (e.g. 90th quantile) and satellite data are above the same threshold, then QBias would be greater than zero.

4.2. *POD and VHI*

Figures 7 and 8 display POD (QPOD) and VHI (QVHI) scores, respectively, for PERSIANN, CMORPH, IMERG, IMERG-GC, GSMaP, and GSMaP-GC satellite precipitation products. As shown in Figure 7, all products display higher POD during pre-monsoon, monsoon, and post-monsoon periods. For higher accumulation periods, POD increases considerably all over the year for all precipitation products and remains higher throughout the pre-monsoon, monsoon, and post-monsoon months (compare Figure 7(a) with Figure 7(c)). Furthermore, as the quantile thresholds increase, QPOD scores also increase remarkably for all rainfall products (compare Figures 7(a) and 7(c) with Figures 7(b) and 7(d) respectively). It is worth pointing out that AghaKouchak et al. (2011) and Mehran and AghaKouchak (2014) observed opposite results for CMORPH and PERSIANN products over the United States of America (USA). Figure 7 also highlight that PERSIANN data sets show rather low POD and QPOD over cold months (December-February). Overall, GSMaP-GC estimates are subject to higher POD than all other precipitation data sets.

It is worthwhile to mention that QPOD (or POD) demonstrates the ratio of the correct number of identification to the total number of reference precipitation occurrences above a threshold. A low QPOD does not necessarily lead to a low volume of rainfall, as a low QPOD may represent a large fraction of the total volume of detected

rainfall. Again, a low QVHI does not really indicate a low number of occurrences, as the observed occurrences by a low QVHI may constitute a large fraction of the total number of detected rainfall events. For this reason, one should consider both metrics to evaluate precipitation data.

Similar results are observed with respect to POD (QPOD) and VHI (QVHI) values for all precipitation products (compare Figure 7 with Figure 8 for the respective temporal accumulations and quantile thresholds). Figure 8 represents that VHI scores are slightly higher than the corresponding POD values particularly over pre-monsoon, monsoon, and post-monsoon periods. It reveals that volumetric skills of the six products are higher than categorical skills during pre-monsoon, monsoon, and post-monsoon seasons (March-November). Figure 8 also shows that PERSIANN satellite data sets exhibit comparatively lower QVHI with respect to the corresponding QPOD scores throughout the year for higher quantile threshold (compare Figure 7(b) with Figure 8(b)). The findings point out that both categorical and volumetric indices are necessary to evaluate satellite precipitation products.

4.3. FAR and VFAR

Figure 9 presents the FAR and quantile FAR (QFAR) for PERSIANN, CMORPH, IMERG, IMERG-GC, GSMaP, and GSMaP-GC rainfall products, while Figure 10 shows the VFAR and QVFAR. It is worth reminding that QFAR values only indicate the ratio of false identification of precipitation to the total number of satellite rainfall occurrences above a threshold (e.g. 50th quantile). It is noted that all products denote higher FAR during post-monsoon, winter, and pre-monsoon periods for 6 hours temporal accumulation (i.e., from September to April) and show least change as the time interval increases (compare Figure 9(a) with Figure 9(c)). However, the FAR

values decrease significantly from May to August for all rainfall data sets with the increase of accumulation period. As shown in Figure 9(a), GSMaP exhibits relatively lower FAR during monsoon period (i.e., June, July, August, and September) whereas GSMaP-GC shows higher magnitudes. Again, IMERG and IMERG-GC show almost identical FAR throughout the year and remains the same for higher accumulation time (see Figure 9(a) and 9(c)). Furthermore, IMERG and IMERG-GC display lower false alarms during pre-monsoon season (i.e., March, April, and May). It is also shown that CMORPH comparatively indicates higher FAR out of the six products over the cold and pre-monsoon months (i.e., from December to April). Although, the QFAR scores of all precipitation data sets almost diminish for extreme quantile threshold, all data sets indicate noticeable QFAR from August to October (see Figure 9(b)). On the other hand, the VFAR values of Figure 10 are lower than the respective FAR values of the satellite products. Among the precipitation data sets, PERSIANN denotes the highest VFAR over the pre-monsoon and monsoon months, whereas the values are relatively lower for CMORPH, IMERG, IMERG-GC, GSMaP, and GSMaP-GC during these two seasons (see Figure 10(a)). As the quantile threshold increases, all precipitation data sets display very little QVFAR for 50% quantile, while QVFAR almost diminishes for 90% quantiles (see Figures 10(b) and 10(c)). Figures 10(b) and 10(c) also indicate that all precipitation estimates show considerable QVFAR during August-October with respect to 50% and 90% quantiles.

4.4. *CSI and VCSI*

As shown in Figure 11, the critical success index (CSI) scores are higher over pre-monsoon and monsoon months (from March to September) for the six satellite products. It is also observed that the CSI values increase considerably during the pre-monsoon and monsoon periods as the accumulation time increases (compare Figure 11(a) with

Figure 11(c)). It implies that with respect to higher temporal accumulations, the magnitudes of CSI are higher over pre-monsoon and monsoon seasons. While IMERG and IMERG-GC denotes higher CSI during pre-monsoon period, GSMaP-GC reveals lower CSI scores out of the six rainfall products during monsoon for 3 hours accumulation. One can see that the QCSI scores improve remarkably throughout the year with the increase of extreme precipitation thresholds (compare Figures 11(a) and 11(c) with Figures 11(b) and 11(d) respectively), however, this improvement is more significant over the post-monsoon and winter months (compare Figure 11(a) with Figure 11(b)). Although the VCSI values of all products are higher than the corresponding CSI scores (compare Figures 11(a) and 11(c) with Figures 12(a) and 12(c) respectively), the QVCSI values are lower than the respective QCSI scores over post-monsoon and cold months (compare Figure 11(b) with Figure 12(b)). Comparing the VCSI and QVCSI scores of all products show that PERSIANN exhibits relatively lower skill throughout the year.

4.5. MI and VMI

Figure 13 shows that PERSIANN exhibits the poorest skill with respect to miss index (MI) compared to the other satellite products (see Figures 13(a) and 13(c)). It is also shown that GSMaP-GC indicates the lowest MI scores throughout the year. One can see that MI decreases considerably over the monsoon months for all precipitation products with the increase of time interval (compare Figure 13(a) with Figure 13(c)). However, all products indicate noticeable magnitudes of MI over post-monsoon, winter, and pre-monsoon months, regardless of the choice of accumulation period. As the quantile threshold increases, the QMI values of all rainfall data sets decrease significantly throughout the year (compare Figure 13(a) with Figure 13(b)), while PERSIANN display noticeable QMI scores particularly over the cold months. Figure 14 presents the

volumetric skills of the six satellite precipitation products with respect to volumetric miss index (VMI) and quantile volumetric miss index (QVMI). Among the precipitation products, PERSIANN exhibits the highest VMI scores throughout the year (see Figures 14(a) and 14(c)). Moreover, the QVMI scores of PERSIANN remain higher for extreme quantile threshold (see Figure 14(b)). GSMaP-GC misses the least amount of rainfall with respect to VMI during monsoon whereas GSMaP performs better during pre-monsoon and post-monsoon seasons (see Figure 14(a)).

4.6. Seasonal variation of the statistical indices

Figures 15 and 16 summarize the skills of the six satellite precipitation products with respect to different statistical metrics during monsoon (left) and winter (right) months. The figure is generated by estimating the statistical indices on the basis of all monsoon and cold months data of the study period from January 2014 to December 2016. Clearly, Figures 15 and 16 highlight the performances of the satellite precipitation estimation algorithms in different climate conditions over Bangladesh. Overall, all rainfall data sets show better performance during monsoon despite the higher FAR and VFAR values during this season.

5. Discussion and conclusions

This study evaluates six well-known satellite-based precipitation estimates (PERSIANN, CMORPH, IMERG, IMERG-GC, GSMaP and GSMaP-GC) with respect to available rain gauge observations over Bangladesh. It is noteworthy that infrared (IR) sensors relate surface rainfall to the brightness and temperature of the cloud tops whereas microwave (PMW) sensors estimate rainfall from the emission and scattering of microwave radiation. According to the findings of this study, the microwave-based precipitation products (GSMaP-GC, GSMaP, IMERG-GC, IMERG, and CMORPH)

outperform the infrared-based products (PERSIANN). Moreover, non gauge-adjusted combined PMW-IR based products, GSMaP and IMERG, show a remarkable improvement over the unadjusted PMW-only CMORPH product. The results show that CMORPH suffers from serious underestimation while GSMaP and IMERG overestimates heavy rainfall amounts. The incorporation of rain gauge observations significantly reduces the overestimation of higher rainfall values by GSMaP. However, the integration of rain gauge data does not decrease the overestimation by IMERG. In fact, the biases for IMERG is better than that of IMERG-GC. Comparison of GSMaP-GC and IMERG-GC indicates that GSMaP-GC has better correspondence with the rain gauge measurements. This may be ascribed to two major factors: first, the algorithm for GSMaP-GC incorporates CPC gauge-based daily precipitation data for its bias correction, which performs better than the monthly gridded rain gauge data set of GPCC adopted in IMERG-GC algorithm; and second, GSMaP-GC uses some auxiliary data sources such as topographic data, GANAL, and MGDSSST (described in section 2). These results are consistent with the findings of Dinku et al. (2010), Romilly and Gebremichael (2011), Ning et al. (2016), and Mehran and AghaKouchak (2014) at different spatial and temporal scales over other parts of the world.

The first group of the precipitation products includes PERSIANN and CMORPH, while the second group consists of IMERG, IMERG-GC, GSMaP, and GSMaP-GC. The spatial resolution is 0.25° for PERSIANN and CMORPH and is 0.10° for IMERG, IMERG-GC, GSMaP, and GSMaP-GC. All the products were evaluated at 3 hours, 6 hours, 12 hours, and 24 hours time scales. The findings of the present investigation over the study region can be explained as follows:

The computed bias scores indicate that PERSIANN and CMORPH tend to underestimate the rainfall during the rainy monsoon season (i.e., June, July, August, and

September). These findings are in agreement with the results from studies by Gao and Liu (2013) and Sohn, Han, and Sue (2010). On the other hand, the bias values are around 1 for IMERG, IMERG-GC, GSMaP and GSMaP-GC during monsoon. A product wise comparison reveals that none of the four products has distinctly good or bad performance by means of bias compared to the others. As the extreme quantile threshold of rain gauge data increases, the QBias scores of PERSIANN, CMORPH, and GSMaP-GC decrease consistently over monsoon months, while IMERG, IMERG-GC, and GSMaP show serious overestimation. It implies that satellite rainfall-retrieval algorithms lose their detection skills as the extreme precipitation threshold increases. It also points out that the current bias adjustment techniques do not improve the accuracy of IMERG-GC and GSMaP-GC satellite products in detecting extreme rainfall accumulations at least over the current test area. During monsoon, most of the precipitation systems are distributed in the northern, north-eastern, and southern parts of Bangladesh. The relatively drier western part has limited precipitation systems during this period (Rafiuddin, Uyeda, and Islam 2010). The underestimation by satellite data sets may partly be associated with warm orographic rain over the study region. Rafiuddin, Uyeda, and Islam (2010) argued that the orography of north-eastern and eastern sides of Bangladesh helps to uplift the warm moist air carried by low level monsoonal wind blowing from the Bay of Bengal. This orographic lifting of warm moist air will lead to the development of cloud. Rainfall may start while the cloud top still remains warm. IR sensors have difficulty to detect rainfall from the warm clouds as the cloud tops are too warm to detect rainfall for the temperature thresholds of the sensors. Again, warm orographic rain will not produce much ice aloft to be identified by PMW sensors. As the PMW sensors estimate rainfall depending upon the scattering by

ice particles, this would obviously lead to underestimation for PMW-based precipitation estimation algorithms.

Conversely, all six precipitation products show serious overestimation during post-monsoon period (i.e., October and November). This overestimation becomes more significant for each quantile threshold as the accumulation time increases. During post monsoon months, the climate is comparatively dry and hot over Bangladesh. Therefore, the overestimation may be attributed to the fact that raindrops detected by the IR and PMW sensors have to pass through a dry and hot atmosphere and may partly or fully evaporate before reaching the ground. PERSIANN also showed extremely high bias over Ethiopia and Zimbabwe (Dinku et al. 2008). The neural networks in PERSIANN algorithm were calibrated using the data of Japan and Florida, where the air masses are mostly maritime type. Thus, the overestimation by PERSIANN may also be associated with the monsoonal type air masses over Bangladesh. The overestimation by CMORPH may be ascribed to the fact that the deep convective systems tend to increase the ice aloft, which is misidentified as precipitation by the PMW sensors (Romilly and Gebremichael 2011).

During pre-monsoon hot summer period (i.e., March, April, and May), PERSIANN, CMORPH, IMERG-GC, and GSMaP-GC satellite rainfall data sets display reasonably good agreement with the reference data. Considering the performances of IMERG-GC and GSMaP-GC in detecting the extreme rainfall accumulations, these results are also very encouraging. However, the overestimation of high rainfall values by IMERG and GSMaP is very prominent during this season. It emphasizes the importance of incorporating rain gauge observations into satellite estimates. Precipitation systems are mostly found over the north-eastern part of the country during summer. Pre-monsoon systems generally do not reach in the south-

eastern part of Bangladesh. Precipitation systems are sparsely distributed over other parts of the study area during this season. During pre-monsoon, the low level warm moist air masses from the Bay of Bengal move through the north-western flat region of Bangladesh toward India and interact with the hot dry air masses coming from India (Rafiuddin, Uyeda, and Islam 2010). This interaction leads to create unstable conditions favourable for the development of convection. Moreover, the mountains of north-eastern side of Bangladesh cause orographic uplifting of the low level warm moist air masses. This orographic uplifting also helps to develop deep convections. Islam and Uyeda (2008) inferred that pre-monsoon deep convective cloud systems are characterized by high intensity rainfall, whereas monsoon and post-monsoon rainfall events are less intense over Bangladesh. Moazami et al. (2013) concluded that in heavy rainfall events, the correlation between IR brightness temperature and rainfall is strong. Again, Ebert, Janowiak, and Kidd (2007) showed that PMW sensors perform better in summer convective conditions over most parts of the USA except the northwest and the far northwest parts. During cold months (i.e., December, January, and February), all satellite algorithms exhibit (either underestimation or overestimation) very limited detection skills, particularly above high quantile thresholds. These results are consistent with previous studies over other regions, e.g. over northwest Europe (Kidd et al. 2012) and over China (Ning et al. 2016).

While the bias is used to assess the performance of satellite rainfall products in estimating the amount of rainfall, categorical skill matrices (i.e., POD, FAR, CSI, and MI) are applied to evaluate the capabilities in detecting the occurrence of rainfall events. On the other hand, VHI, VFAR, VCSI, and VMI are used to measure the volumetric skills of the products. The estimated POD and VHI values for the entire distribution of the satellite data sets reveal that the detection skill of all products are

higher during pre-monsoon, monsoon, and post-monsoon periods (March-November). Again, all satellite estimates show better performance with respect to CSI and VCSI over pre-monsoon and monsoon months (March-September). POD represents the ratio of the number of correctly identified events to the total number of correct and miss observations by satellite data sets, while CSI shows the ratio of the number of correctly detected occurrences to the total number of correct, miss, and false observations (see Equations (5) and (9)). Therefore, it is reasonable that the CSI scores will be lower than the corresponding POD values. However, the comparison of POD (VHI) with CSI (VCSI) indicates that the decrease of CSI (VCSI) scores of all precipitation products is more significant over post-monsoon than those over pre-monsoon and monsoon period. The computed FAR and VFAR values of the products are also higher during this season. Furthermore, for higher quantile thresholds, all six satellite rainfall data sets represent considerable QFAR and QVFAR over the post monsoon months. This problem might be attributed to two major factors: the first one is sub-cloud evaporation (as described above). Below the cloud systems, the atmosphere is mostly hot and dry over Bangladesh during post-monsoon. As a result, although satellite sensors may detect rainfall aloft, it may not reach the ground surface resulting in huge false alarms in satellite estimates; the second reason may be the coarse spatial resolution ($0.25^{\circ}/0.10^{\circ} \sim 27 \text{ km}/11 \text{ km}$ in latitude and longitude) of satellite rainfall data sets. The coarse satellite pixel may cover both rain and no-rain areas whose average may be incorrectly identified as a non-raining pixel (Dinku et al. 2011).

In terms of MI and VMI scores, GSMaP-GC shows better performance during monsoon whereas the performance of PERSIANN is poor throughout the year. Again, PERSIANN exhibits higher VFAR over pre-monsoon and monsoon months and lower VCSI all over the year. GSMaP has a relatively better skill with respect to FAR during

monsoon. In addition, GSMaP displays lower VMI scores over pre-monsoon and post-monsoon seasons. Surprisingly, IMERG and IMERG-GC indicate almost similar FAR values throughout the year. Moreover, these two products reveal the lowest false alarms during pre-monsoon. Both categorical and volumetric indices demonstrate that the detection skill of the products increase remarkably as the extreme precipitation threshold increases. The detection capability of all six satellite data sets also increases with the increase of accumulation time. Considering the categorical and volumetric matrices it is shown that with respect to higher temporal accumulations, the performance of the precipitation products are comparatively better both in discriminating rain/no-rain events and estimating the corresponding rainfall amounts during monsoon period. However, all the products denote considerable magnitude of FAR and VFAR values during this season.

With respect to POD, VHI, FAR, VFAR, CSI, VCSI, MI, and VMI scores, all six precipitation products exhibit relatively poor performance over the cold months, regardless of the preference of accumulation times. However, the values of QPOD, QVHI, QCSI, and QVCSI of CMORPH, IMERG, IMERG-GC, GSMaP, and GSMaP-GC increase during this season as the extreme quantile threshold increases. The corresponding QFAR and QVFAR values of these products also decrease with the increase of extreme precipitation threshold.

Comparison of different satellite precipitation products reveal that all products are doing well in detecting the occurrence of rainfall, but are not so good in estimating the amount of rainfall. PERSIANN exhibits the worst performance over Bangladesh. Overall, GSMaP-GC performs better than the other products, although it shows the highest false alarms during monsoon and post-monsoon periods over the study region. Moreover, GSMaP-GC shows serious underestimation in estimating monsoon rainfall

whereas it overestimates the rainfall over post-monsoon months. Someone would have still expected an improved results for IMERG-GC because it uses more advanced algorithm and superior input data sets. However, the performance of IMERG-GC is also very promising over Bangladesh.

It is worth pointing out here that there is only one rain gauge station inside each satellite pixel. For comparing satellite measurements with rain gauge observations, especially for extreme rainfall analysis at least three or more rain gauges are necessary inside each pixel grid (Moazami et al. 2013). The present study considers relatively a long time period (3 years), which should minimize the errors associated with the spatial representativeness of the rain gauge data. Spatial and temporal uncertainty may also exist while comparing rain gauge data with satellite estimates (AghaKouchak et al. 2011). In future, it is planned to explore the spatial distribution of the statistical indices over the study region, which will certainly provide more detailed understanding of the biases and errors associated with the satellite precipitation products.

Acknowledgements

The author would like to thank Bangladesh Meteorological Department (BMD) for providing rain gauge data of Bangladesh. The satellite precipitation data sets were downloaded from the respective mission websites and are sincerely acknowledged. The author also thanks the anonymous reviewers for the constructive comments and suggestions which brought significant improvements in the revised version of the manuscript.

References

- AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai. 2011. "Evaluation of Satellite-Retrieved Extreme Precipitation Rates Across the Central United States." *Journal of Geophysical Research* 116: D02115. doi:10.1029/2010JD014741.
- AghaKouchak, A., and A. Mehran. 2013. "Extended Contingency Table: Performance Metrics for Satellite Observations and Climate Model Simulations." *Water Resources Research* 49: 7144-7149. doi:10.1002/wrcr.20498.

- Aonashi, K., J. Awaka, M. Hirose, T. Kozu, T. Kubota, G. Liu, S. Shige, et al. 2009. "GSMaP Passive Microwave Precipitation Retrieval Algorithm: Algorithm Validation and Description." *Journal of the Meteorological Society of Japan* 87A: 119-136. doi:10.2151/jmsj.87A.119.
- Behrangi, A., B. Khakbaz, T. C. Jaw, A. AghaKouchak, K. Hsu, and S. Sorooshian. 2011. "Hydrologic Evaluation of Satellite Precipitation Products over a Mid-Size Basin." *Journal of Hydrology* 397: 225-237. doi:10.1016/j.jhydrol.2010.11.043.
- Cattani, E., A. Merino, and V. Levizzani. 2016. "Evaluation of Monthly Satellite-Derived Precipitation Products over East Africa." *Journal of Hydrometeorology* 17: 2555-2573. doi:10.1175/JHM-D-15-0042.1.
- Cohen Liechti, T., J. P. Matos, J.-L. Boillat, and A. J. Schleiss. 2012. "Comparison and Evaluation of Satellite Derived Precipitation Products for Hydrological Modeling of the Zambezi River Basin." *Hydrology and Earth System Sciences* 16: 489-500. doi.org/10.5194/hess-16-489-2012, 2012.
- Dewan, T. H. 2015. "Societal Impacts and Vulnerability to Floods in Bangladesh and Nepal." *Weather and Climate Extremes* 7: 36-42. <http://dx.doi.org/10.1016/j.wace.2014.11.001>.
- Dinku, T., P. Ceccato, and S. J. Connor. 2011. "Challenges of Satellite Rainfall Estimation over Mountainous and Arid Parts of East Africa." *International Journal of Remote Sensing* 32(21): 5965-5979. <http://dx.doi.org/10.1080/01431161.2010.499381>.
- Dinku, T., P. Ceccato, E. Grover-Kopec, M. Lemma, S. J. Connor, and C. F. Ropelewski. 2007. "Validation of Satellite Rainfall Products over East Africa's Complex Topography." *International Journal of Remote Sensing* 28(7): 1503-1526. doi: 10.1080/01431160600954688.
- Dinku, T., S. Chidzambwa, P. Ceccato, S. J. Connor, and C. F. Ropelewski. 2008. "Validation of High-Resolution Satellite Rainfall Products over Complex Terrain." *International Journal of Remote Sensing* 29(14): 4097-4110. doi:10.1080/01431160701772526.
- Dinku, T., F. Ruiz, S. J. Connor, and P. Ceccato. 2010. "Validation and Intercomparison of Satellite Rainfall Estimates over Colombia." *Journal of Applied Meteorology and Climatology* 49: 1004-1014. <https://doi.org/10.1175/2009JAMC2260.1>.

- Ebert, E. E., J. E. Janowiak, and C. Kidd. 2007. "Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models." *Bulletin of the American Meteorological Society* 88: 47-64. doi:10.1175/bams-88-1-47.
- Gao, Y. C., and M. F. Liu. 2013. "Evaluation of High-Resolution Satellite Precipitation Products Using Rain Gauge Observations over the Tibetan Plateau." *Hydrology and Earth System Sciences* 17: 837-849. doi:10.5194/hess-17-837-2013.
- Gebregiorgis, A. S., and F. Hossain. 2014: Estimation of Satellite Rainfall Error Variance Using Readily Available Geophysical Features. *IEEE Transactions on Geoscience and Remote Sensing* 52(1): 288-304. doi:10.1109/TGRS.2013.2238636.
- Hong, Y., K. Hsu, S. Sorooshian, and X. Gao. 2004. "Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System." *Journal of Applied Meteorology* 43: 1834-1853. <http://dx.doi.org/10.1175/JAM2173.1>.
- Hou, A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi. 2014. "The Global Precipitation Measurement Mission." *Bulletin of the American Meteorological Society* 95: 701-722. <https://doi.org/10.1175/BAMS-D-13-00164.1>.
- Huffman, G. J. 2016. "The Transition in Multi-Satellite Products from TRMM to GPM (TMPA to IMERG)." Version 161025, 5 pp. https://pmm.nasa.gov/sites/default/files/document_files/TMPA-to-IMERG_transition_161025.pdf.
- Huffman, G. J., D. T. Bolvin, and E. J. Nelkin. 2015. "Integrated Multi-satellite Retrievals for GPM (IMERG) Technical Documentation." IMERG_doc, 48 pp. https://pmm.nasa.gov/sites/default/files/document_files/IMERG_doc.pdf.
- Huffman, G. J., D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, and P. Xie. 2015. "NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG)." Algorithm Theoretical Basis Doc. (ATBD), version 4.5, 30 pp. https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.5_0.pdf.
- Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker. 2007. "The TRMM Multisatellite Precipitation

- Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales.” *Journal of Hydrometeorology* 8: 38-55.
<http://dx.doi.org/10.1175/JHM560.1>.
- Hughes, D. A. 2006. “Comparison of Satellite Rainfall Data with Observations from Gauging Station Networks.” *Journal of Hydrology* 327(3-4): 399-410.
<http://dx.doi.org/10.1016/j.jhydrol.2005.11.041>.
- Islam, M. N., and H. Uyeda. 2008. “Vertical Variations of Rain Intensity in Different Rainy Periods in and around Bangladesh Derived from TRMM Observations.” *International Journal of Climatology* 28: 273-279. doi:10.1002/joc.1585.
- Janowiak, J., R. Joyce, and Y. Yarosh. 2001. “A Real-Time Global Half-Hourly Pixel-Resolution Infrared Data Set and Its Applications.” *Bulletin of the American Meteorological Society* 82: 205-217. [http://dx.doi.org/10.1175/1520-0477\(2001\)082%3C0205:ARTGHH%3E2.3.CO;2](http://dx.doi.org/10.1175/1520-0477(2001)082%3C0205:ARTGHH%3E2.3.CO;2).
- Joyce, R., J. Janowiak, P. Arkin, and P. Xie. 2004. “CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution.” *Journal of Hydrometeorology* 5: 487-503. [http://dx.doi.org/10.1175/1525-541\(2004\)005%3C0487:CAMTPG%3E2.0.CO;2](http://dx.doi.org/10.1175/1525-541(2004)005%3C0487:CAMTPG%3E2.0.CO;2).
- Kidd, C., P. Bauer, J. Turk, G. J. Huffman, R. Joyce, K. Hsu, and D. Braithwaite. 2012. “Intercomparison of High-Resolution Precipitation Products over Northwest Europe.” *Journal of Hydrometeorology* 13: 67-83. doi:10.1175/JHM-D-11-042.1.
- Kubota, T., S. Shige, H. Hashizume, K. Aonashi, N. Takahashi, S. Seto, M. Hirose, et al. 2007. “Global Precipitation Map Using Satellite-borne Microwave Radiometers by the GSMaP Project : Production and Validation.” *IEEE Transactions on Geoscience and Remote Sensing* 45(7): 2259-2275.
doi:10.1109/TGRS.2007.895337.
- Mehran, A., and A. AghaKouchak. 2014. “Capabilities of Satellite Precipitation Data sets to Estimate Heavy Precipitation Rates at Different Temporal Accumulations.” *Hydrological Processes* 28(4): 2262-2270.
doi:10.1002/hyp.9779.
- Mei, Y., E. N. Anagnostou, E. I. Nikolopoulos, and M. Borga. 2014. “Error Analysis of Satellite Precipitation Products in Mountainous Basins.” *Journal of Hydrometeorology* 15: 1778-1793. doi:10.1175/JHM-D-13-0194.1.

- Moazami, S., S. Golian, M. R. Kavianpour, and Y. Hong. 2013. "Comparison of PERSIANN and V7 TRMM Multi-Satellite Precipitation Analysis (TMPA) Products with Rain Gauge Data over Iran." *International Journal of Remote Sensing* 34(22): 8156-8171. doi:10.1080/01431161.2013.833360.
- Ning, S., J. Wang, J. Jin, and H. Ishidaira. 2016. "Assessment of the Latest GPM-Era High-Resolution Satellite Precipitation Products by Comparison with Observation Gauge Data over the Chinese Mainland." *Water* 8(11): 481. doi:10.3390/w8110481.
- Okamoto, K., T. Ushio, T. Iguchi, N. Takahashi, and K. Iwanami. 2005. "The Global Satellite Mapping of Precipitation (GSMaP) Project." *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 2005*: 3414-3416. doi:10.1109/IGARSS.2005.1526575.
- Rafiuddin, M., H. Uyeda, and M. N. Islam. 2010. "Characteristics of Monsoon Precipitation Systems in and around Bangladesh." *International Journal of Climatology* 30: 1042-1055. doi:10.1002/joc.1949.
- Rahman, M. M., D. S. Arya, N. K. Goelb, and A. K. Mitra. 2012. "Rainfall Statistics Evaluation of ECMWF Model and TRMM Data over Bangladesh for Flood Related Studies." *Meteorological Applications* 19: 501-512. doi:10.1002/met.293.
- Romilly, T. G., and M. Gebremichael. 2011. "Evaluation of Satellite Rainfall Estimates over Ethiopian River Basins." *Hydrology and Earth System Sciences* 15: 1505–1514. doi.org/10.5194/hess-15-1505-2011.
- Sapiano, M. R. P., and P. A. Arkin. 2009. "An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data." *Journal of Hydrometeorology* 10: 149-166. doi:10.1175/2008jhm1052.1.
- Shahid, S. 2010. "Rainfall Variability and the Trends of Wet and Dry Periods in Bangladesh." *International Journal of Climatology* 30: 2299-2313. doi: 10.1002/joc.2053.
- Sohn, B. J., H.-J. Han, and E.-K. Seo. 2010. "Validation of Satellite-Based High-Resolution Rainfall Products over the Korean Peninsula Using Data from a Dense Rain Gauge Network." *Journal of Applied Meteorology and Climatology* 49: 701-714. doi:10.1175/2009jamc2266.1.
- Sorooshian, S., A. AghaKouchak, P. Arkin, J. Eylander, E. Foufoula-Georgiou, R. Harmon, J. M. H. Hendrickx, et al. 2011. "Advanced Concepts on Remote

- Sensing of Precipitation at Multiple Scales.” *Bulletin of the American Meteorological Society* 92: 1353-1357.
<http://dx.doi.org/10.1175/2011BAMS3158.1>.
- Sorooshian, S., K. Hsu, X. Gao, H. Gupta, B. Imam, and D. Braithwaite. 2000.
 “Evaluation of PERSIANN System Satellite-Based Estimates of Tropical Rainfall.” *Bulletin of the American Meteorological Society* 81: 2035-2046.
[http://dx.doi.org/10.1175/1520-477\(2000\)081%3C2035:EOPSSE%3E2.3.CO;2](http://dx.doi.org/10.1175/1520-477(2000)081%3C2035:EOPSSE%3E2.3.CO;2).
- Thiemig, V., R. Rojas, M. Zambrano-Bigiarini, V. Levizzani, and A. De Roo. 2012.
 “Validation of Satellite-Based Precipitation Products over Sparsely Gauged African River Basins.” *Journal of Hydrometeorology* 13: 1760-1783.
<http://dx.doi.org/10.1175/JHM-D-12-032.1>.
- Ushio, T., K. Sasashige, T. Kubota, S. Shige, K. Okamoto, K. Aonashi, T. Inoue, et al. 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* 87A: 137–151. doi:10.2151/jmsj.87A.137.
- Vernimmen, R. R. E., A. Hooijer, Mamenun, E. Aldrian, and A. I. J. M. van Dijk. 2012.
 “Evaluation and Bias Correction of Satellite Rainfall Data for Drought Monitoring in Indonesia.” *Hydrology and Earth System Sciences* 16: 133-146.
 doi:10.5194/hess-16-133-2012.
- Wilks, D. S. 2006. *Statistical Methods in the Atmospheric Science*. 2nd ed. Burlington, MA, USA: Academic Press.
- Yamamoto, M. K., and S. Shige. 2015. “Implementation of An Orographic/Nonorographic Rainfall Classification Scheme in the GSMaP Algorithm for Microwave Radiometers.” *Atmospheric Research* 163: 36-47.
<https://doi.org/10.1016/j.atmosres.2014.07.024>.
- Yang, X., B. Yong, Y. Hong, S. Chen, and X. Zhang. 2015. “Error Analysis of Multi-Satellite Precipitation Estimates with An Independent Rain Gauge Observation Network over A Medium-Sized Humid Basin.” *Hydrological Sciences Journal* 61(10): 1813-1830. doi:10.1080/02626667.2015.1040020.
- Yilmaz, K., T. Hogue, K. Hsu, S. Sorooshian, H. Gupta, and T. Wagener. 2005.
 “Intercomparison of Rain Gauge, Radar and Satellite-Based Precipitation Estimates with Emphasis on Hydrologic Forecasting.” *Journal of Hydrometeorology* 6: 497-517. <http://dx.doi.org/10.1175/JHM431.1>.

- Zhou, T., B. Nijssen, G. J. Huffman, and D. P. Lettenmaier. 2014. "Evaluation of Real-Time Satellite Precipitation Data for Global Drought Monitoring." *Journal of Hydrometeorology* 15: 651-660. <http://dx.doi.org/10.1175/JHM-D-13-0128.1>.
- Zulkafli, Z., W. Buytaert, C. Onof, B. Manz, E. Tarnavsky, W. Lavado, and J. Guyot. 2014. "A Comparative Performance Analysis of TRMM 3B42 (TMPA) Versions 6 and 7 for Hydrological Applications over Andean–Amazon River Basins." *Journal of Hydrometeorology* 15: 581-592. <http://dx.doi.org/10.1175/JHM-D-13-094.1>.

Table 1. Summary of the satellite precipitation products evaluated here. The PMW and Gauge columns show whether the products integrate passive microwave estimates or rain gauge observations (Y and N stands for yes and no respectively).

Product	Temporal resolution	Spatial resolution	Existence	Spatial coverage	PMW	Gauge
PERSIANN	3-hourly	0.25°	2000-present	60° S-60° N	N	N
CMORPH	3-hourly	0.25°	2003-present	60° S-60° N	Y	N
IMERG version 04A	Half-hourly	0.10°	2014-present	60° S-60° N	Y	N
IMERG-GC version 04A	Half-hourly	0.10°	2014-present	60° S-60° N	Y	Y
GSMaP version 04B	Hourly	0.10°	2014-present	60° S-60° N	Y	N
GSMaP-GC version 04B	Hourly	0.10°	2014-present	60° S-60° N	Y	Y

Table 2. Summary statistics of rain gauge and satellite precipitation products for different temporal accumulations during 2014-2016 (Q is quantile and SD is standard deviation).

Accumulation period	Product	Mean (mm)	SD (mm)	Bias	Q10 (mm)	Q25 (mm)	Q50 (mm)	Q75 (mm)	Q90 (mm)	Q95 (mm)
3 hours	Rain gauge	14.4	15.5	--	1.7	3.4	8.9	19.8	35.0	47.2
	PERSIANN	10.2	8.5	0.71	2.0	3.9	7.7	14.1	21.5	27.0
	CMORPH	12.7	12.8	0.89	2.2	4.1	8.6	16.5	29.0	38.7
	IMERG	17.1	20.1	1.19	2.5	4.6	9.6	20.8	41.9	62.0
	IMERG-GC	17.8	20.1	1.24	2.8	5.3	10.8	22.0	41.5	59.6
	GSMaP	19.6	28.1	1.36	3.5	6.3	11.5	21.1	39.5	61.5
	GSMaP-GC	14.3	13.0	0.99	3.3	6.1	10.6	18.6	29.2	37.4
6 hours	Rain gauge	19.4	20.9	--	1.9	4.5	12.0	26.6	46.8	61.9
	PERSIANN	14.4	13.2	0.74	2.5	4.9	10.4	19.6	31.7	39.7
	CMORPH	16.8	18.3	0.87	2.5	4.9	10.5	21.5	38.9	53.4
	IMERG	22.0	27.5	1.13	2.9	5.6	12.0	26.3	54.1	79.6
	IMERG-GC	23.0	27.6	1.19	3.2	6.3	13.5	27.8	54.8	78.1
	GSMaP	25.5	35.7	1.31	4.0	7.6	14.6	28.1	53.9	81.9
	GSMaP-GC	19.4	18.7	1.00	3.9	7.3	13.9	25.3	41.2	55.0
12 hours	Rain gauge	24.6	26.7	--	2.1	5.9	15.8	33.8	58.9	77.5
	PERSIANN	18.8	18.1	0.76	2.8	5.9	13.3	25.8	41.9	53.0
	CMORPH	20.6	22.8	0.84	2.8	5.7	12.6	27.0	48.0	66.0
	IMERG	26.8	33.3	1.10	3.2	6.6	14.6	33.1	66.6	96.3
	IMERG-GC	28.1	33.9	1.14	3.6	7.3	16.1	35.5	66.2	96.4
	GSMaP	30.8	41.9	1.25	4.2	8.3	17.3	35.9	67.4	105.2
	GSMaP-GC	24.5	25.0	1.00	4.4	8.7	17.1	31.5	52.8	70.9
24 hours	Rain gauge	30.9	35.1	--	3.0	7.4	19.1	41.6	73.7	97.6
	PERSIANN	24.0	24.4	0.78	3.2	7.3	16.5	32.4	54.4	70.7
	CMORPH	24.8	28.2	0.80	3.1	6.4	14.8	32.4	59.9	80.5
	IMERG	32.0	40.5	1.04	3.7	7.6	17.3	39.2	79.2	113.6
	IMERG-GC	33.8	41.3	1.10	4.0	8.5	19.1	43.1	81.6	115.5
	GSMaP	36.4	49.5	1.18	4.3	9.2	19.8	43.3	82.8	127.3
	GSMaP-GC	31.1	33.5	1.01	5.5	10.6	20.8	39.7	67.5	91.0

Figure 1. (a) Topography and rain gauge distribution of the study area. Annual rainfall accumulation (mm year⁻¹) during 2014-2016 for (b) PERSIANN, (c) CMORPH, (d) IMERG, and (e) GSMaP satellite precipitation products.

Figure 2. Mean monthly rainfall during 2014-2016 derived from the rain gauge observations. Value at the top of each bar represents percentage contribution of the respective month with respect to the mean annual rainfall over Bangladesh.

Figure 3. Empirical cumulative distribution functions of rain gauge and satellite precipitation products for (a) 3 hours and (b) 6 hours temporal accumulations.

Figure 4. Empirical cumulative distribution functions of rain gauge and satellite precipitation products for (a) 12 hours and (b) 24 hours (daily) temporal accumulations.

Figure 5. Bias and quantile bias (QBias) for 3 hours temporal accumulation with respect to (a) all data and (b) Q75 and (c) Q90 thresholds of reference data (Q, quantiles).

Figure 6. Quantile bias (QBias) for 12 hours and 24 hours temporal accumulations with respect to (a) all data, (b) Q75, and (c) Q90 thresholds of reference data (Q, quantiles).

Figure 7. Probability of detection (POD) for (a) 3 hours and (c) 24 hours accumulations with respect to all data and quantile probability of detection (QPOD) for (b) 3 hours and (d) 24 hours accumulations with respect to Q90 threshold (Q, quantiles). 0 at the top of the bar represents zero value of the respective bar.

Figure 8. Volumetric hit index (VHI) for (a) 3 hours and (c) 24 hours accumulations with respect to all data and quantile volumetric hit index (QVHI) for (b) 3 hours and (d) 24 hours accumulations with respect to Q90 threshold (Q, quantiles). 0 at the top of the bar represents zero value of the respective bar.

Figure 9. False alarm ratio (FAR) for (a) 6 hours and (c) 24 hours accumulations with respect to all data and quantile false alarm ratio (QFAR) for (b) 6 hours accumulation with respect to Q90 threshold (Q, quantiles).

Figure 10. Volumetric false alarm ratio (VFAR) and quantile volumetric false alarm ratio (QVFAR) for 6 hours accumulation with respect to (a) all data and (b) Q50 and (c) Q90 thresholds (Q, quantiles).

Figure 11. Critical success index (CSI) for (a) 3 hours and (c) 12 hours accumulations with respect to all data and quantile critical success index (QCSI) for (b) 3 hours and (d) 12 hours accumulations with respect to Q50 and Q90 threshold, respectively (Q, quantiles). 0 at the top of the bar represents zero value of the respective bar.

Figure 12. Volumetric critical success index (VCSI) for (a) 3 hours and (c) 12 hours accumulations with respect to all data and quantile volumetric critical success index (QVCSI) for (b) 3 hours and (d) 12 hours accumulations with respect to Q50 and Q90 threshold, respectively (Q, quantiles). 0 at the top of the bar represents zero value of the respective bar.

Figure 13. Miss index (MI) for (a) 3 hours and (c) 24 hours accumulations with respect to all data and quantile miss index (QMI) for (b) 3 hours accumulation with respect to Q75 threshold (Q, quantiles).

Figure 14. Volumetric miss index (VMI) for (a) 3 hours and (c) 24 hours accumulations with respect to all data and quantile volumetric miss index (QVMI) for (b) 3 hours accumulation with respect to Q75 threshold (Q, quantiles).

Figure 15. Seasonal variation of statistical indices for 3 hours temporal accumulations (Q, quantiles).

Figure 16. Seasonal variation of statistical indices for 6 hours temporal accumulations (Q, quantiles).