METEOROLOGICAL APPLICATIONS
Meteorol. Appl. (2012)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/met.1327

RMetS
Royal Meteorological Society

Development of daily gridded rainfall dataset over the Ganga, Brahmaputra and Meghna river basins

Venkatraman Prasanna,[a]* Juvy Subere,[b] Dwijendra K. Das,[c] Srinivasan Govindarajan[c] and Tetsuzo Yasunari[d]
[a] *APEC Climate Center (APCC), Busan, South Korea*
[b] *School of Bio-agricultural Sciences, Nagoya University, Japan*
[c] *Regional Integrated Multi-Hazard Early Warning System (RIMES), Bangkok, Thailand*
[d] *Hydrospheric Atmospheric Research Center, Nagoya University, Japan*

ABSTRACT: The India Meteorological Department (IMD) gridded rainfall dataset, the 47 Bangladesh gauge rainfall observations and the Tropical Rainfall Measuring Mission (TRMM) 3B42V6 satellite data are used in the present analysis. The nearest neighbour interpolation scheme is used, wherein the interpolated values are computed from a weighted sum of observations. The Bangladesh daily gauge measured rainfall is interpolated into regular grids of $0.5^\circ \times 0.5^\circ$ resolution every day from January 1988 to December 2007 and appended with the daily gridded dataset of the IMD over the Indian region. A similar resolution dataset of $0.5^\circ \times 0.5^\circ$ for the TRMM-3B42V6 data from January 1998 to December 2007 is created from the original data of $0.25^\circ \times 0.25^\circ$ resolution. To produce a merged rainfall product, all the gridded datasets are merged. The merging of datasets is done in such a way as to include the highest rainfall at each grid point from the three products. Based on the three available sets of daily observations (IMD dataset ($1^\circ \times 1^\circ$), TRMM-3B42 ($0.25^\circ \times 0.25^\circ$) and 46 daily station observations over Bangladesh), a dataset of $0.5^\circ \times 0.5^\circ$ resolution on a daily scale is generated. The focus of this study is to compare the TRMM-3B42V6 rainfall data over the Ganga, Brahmaputra and Meghna (GBM) domain with observed point gauge data, and assess the possibility of using them for application in real time flood forecasting as well as to serve as a comparison tool for the baseline simulation of high resolution atmospheric models aimed at flood forecasting and climate change projections. Copyright © 2012 Royal Meteorological Society

KEY WORDS gauge rainfall dataset; Bangladesh floods; TRMM 3B42V6; gridded rainfall dataset

Received 26 December 2010; Revised 14 March 2012; Accepted 4 April 2012

1. Introduction

Bangladesh is among the most natural disaster prone (flood prone) countries in the world. It has suffered 170 large-scale floods between 1970 and 1998. The frequency of flooding episodes is growing, with catastrophic 'once in a generation' floods occurring more regularly (Chowdhury, 1988, 1989). This includes eight major floods between 1974 and 2004, many of which are considered by hydrologists to be at a size expected only once in every 20 years. During the 2007 monsoon, Bangladesh experienced severe floods with serious consequences for the national economy and the livelihoods of millions of people. A major cause of flooding in Bangladesh is due to the fact that it is a low lying country, with 70% of its land area being less than 1 m above sea level and 80% of it floodplain. Moreover, Bangladesh receives large amounts of water passing through it with two major rivers (Ganges and Brahmaputra) converging and forming a huge delta as they enter the sea. Both rivers also have large drainage basins, which increases flood risk (Brammer, 1990a, 1990b). The occurrence of a monsoon climate and the annual torrential rains in Bangladesh often results in the rivers exceeding their capacity and flooding. Moreover, melting snow from the Himalayas further increases the flood risks as torrents of melt water enter the rivers at their source in the spring (Pant and Kumar, 1997).

Each year, short-lived flooding occurs throughout the summer and early autumn with sufficient irregularity to produce adverse social and agricultural impacts. Even in the normal years, about one-fifth of the country can be inundated for days. The timing of the flooding is critical for the harvesting of the winter rice crop, planting of the summer rice, and viability of other crops planted during the monsoons. As an example, during the summer of 1998 over 60% of Bangladesh is inundated for nearly 3 months and in July 2004, widespread flooding occurred as the Brahmaputra exceeded dangerous flood levels over a 2 week period (Ali, 2007). In July and September 2007 the Brahmaputra flooded severely, with the resulting inundation lasting for weeks (Saiful *et al*., 2010).

There is a considerable requirement for rainfall related information for timely action to avert the river flood calamities in Bangladesh. Such timely information can impact decision-making, specifically when to evacuate the flood-prone regions, ensure protection of food and water supplies and the safety of livestock. If pre-emptive actions are taken, then these can significantly lower the cost of rehabilitation following a period of flooding. Providing rainfall information for the upstream areas in India in advance would provide better warnings of severe flooding in the low lying areas near the Ganges and Brahmaputra catchments of the India-Bangladesh border. The dataset would provide an opportunity to map low-lying areas prone to flooding and inundation during the

* Correspondence to: V. Prasanna, APEC Climate Center, 1463 U-Dong, Haeundae-gu, Busan 612-020, South Korea.
E-mail: Prasa_arnala@yahoo.com

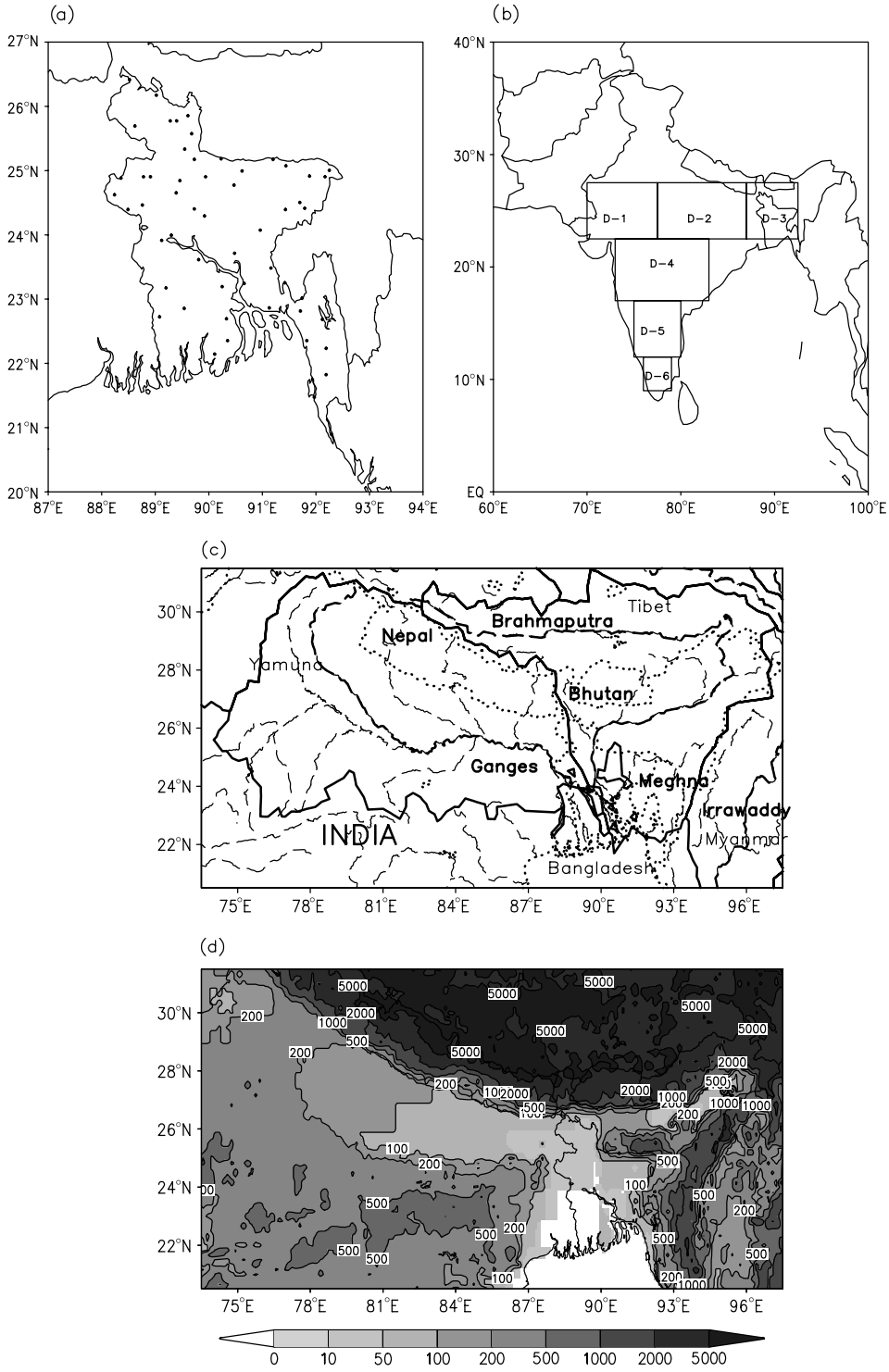

Figure 1. (a) Map of Bangladesh Water Development Board Monitored Rainfall Stations. (b) Selected land domains over India and Bangladesh for comparison of rainfall representation by gauge and TRMM 3B42V6. Domains are abbreviated as D. (c) River systems over and around GBM domain and river basin map of GBM domain. Thick lines are GBM basin boundary, dotted lines are political boundary and dashed lines are rivers; thick dashed lines are GBM river system. (d) Topography over the GBM domain.

monsoon season. The dataset also serves as an observational reference to compare with climate model outputs aimed at flood forecasting. Efforts are in hand to validate global operational climate forecast model results over this domain. It is envisaged in this study to create a useful dataset for flood related studies.

The paper is organized into five sections. Section 2 describes the different datasets used for constructing the precipitation products, and the methodologies adopted for gridding and merging the precipitation products are also described. Section 3 discusses the results and provides details of some specific flooding cases over Bangladesh. Lastly, Section 4 concludes,

Table 1. Data obtained from Flood Forecasting and Warning Center of Bangladesh Water Development Board Monitored Rainfall Stations.

Sl.no	Station name	Latitude (°)	Longitude (°)	No. of available observation days (days)	Missing no. of observation days (days)	Percentage of available data (%)
1	Bandarban	22.22	92.19	6662	635	91.20
2	Barguna	22.13	90.11	6800	497	93.09
3	Barisal	22.68	90.33	6655	642	91.10
4	Bhairab Bazar	24.06	90.96	7072	225	96.81
5	Bogra	24.83	89.46	6854	443	93.83
6	Chandpur	23.23	90.66	6820	477	93.36
7	Chapai-Nawabganj	24.61	88.24	6548	749	89.64
8	Chilmari	25.56	89.68	5608	1689	76.77
9	Chittagong	22.34	91.83	5967	1330	81.68
10	Comilla	23.47	91.16	7119	178	97.45
11	Dalia	26.16	89.02	5579	1718	76.37
12	Dewanganj	25.16	89.73	6252	1045	85.59
13	Dhaka	23.70	90.48	6665	632	91.24
14	Dinajpur	25.68	88.62	6978	319	95.52
15	Durgapur	24.45	88.76	5613	1684	76.84
16	Faridpur	23.60	89.81	7007	290	95.92
17	Gaibandha	25.32	89.55	5154	2143	70.55
18	Habiganj	24.38	91.43	6934	363	94.92
19	Jessore	23.16	89.20	6886	411	94.26
20	Kanairghat	24.99	92.25	5570	1727	76.25
21	Kaunia	25.76	89.40	6856	441	93.85
22	Khulna	22.84	89.54	6612	685	90.51
23	Kurigram	25.84	89.61	7056	241	96.59
24	Kushtia	23.90	89.12	7012	285	95.99
25	Lama	21.81	92.19	6949	348	95.13
26	Manu Rly Bridge	24.40	91.79	5681	1616	77.77
27	Moulvi Bazar	24.49	91.70	6942	355	95.03
28	Mymensingh	24.76	90.47	7074	223	96.84
29	Nakuagaon	25.17	90.23	6496	801	88.93
30	Naogaon	24.89	88.91	6912	385	94.62
31	Narayanhat	22.80	91.71	6966	331	95.36
32	Noakhali	22.85	91.13	6461	836	88.45
33	Pabna	23.98	89.30	5182	2115	70.94
34	Panchagarh	26.40	88.52	6561	736	89.82
35	Parshuram	23.24	91.42	7004	293	95.88
36	Patuakhali	22.34	90.35	5836	1461	79.89
37	Rajshahi	24.38	88.49	6615	682	90.55
38	Ramgarh	23.00	91.74	6673	624	91.35
39	Rangamati	22.67	92.12	5789	1508	79.25
40	Rangpur	25.76	89.28	5747	1550	78.67
41	Satkhira	22.71	89.08	6792	505	92.98
42	Sheola	24.89	92.17	6989	308	95.67
43	Sirajganj	24.39	89.73	6986	311	95.63
44	Sunamganj	25.06	91.44	6669	628	91.29
45	Sylhet	24.9	91.88	7110	187	97.33
46	Tangail	24.28	89.92	6599	698	90.34

Period of available station data for the calendar dates, starting date: 1 January 1988; ending date: 31 December 2007.

noting the advantages of using precipitation products and highlights some of their limitations.

2. Data and methods

The six different datasets used in the present study are given below, of which three are primary datasets (Bangladesh rain gauge dataset, India Meteorological Department (IMD) gridded dataset and Tropical Rainfall Measuring Mission (TRMM) 3B42V6 dataset) used for constructing the merged product, whereas the secondary datasets (Global Precipitation and Climatology Project (GPCP) data, Global Precipitation and Climatology Centre (GPCC) data and Aphrodite data) are used for comparison with the merged product. A brief description of each of these datasets and their respective sources are described below

2.1. Datasets

2.1.1. Bangladesh rain gauge data

Daily precipitation fields over Bangladesh are derived from a dense network of 46 (individual station observations) rain gauges. Rain gauge data are obtained from the Flood Forecasting and Warning Center (FFWC) of Bangladesh Water Development Board (BWDB) from 1988 to 2008. The spatial distribution of stations used for the merged dataset over the

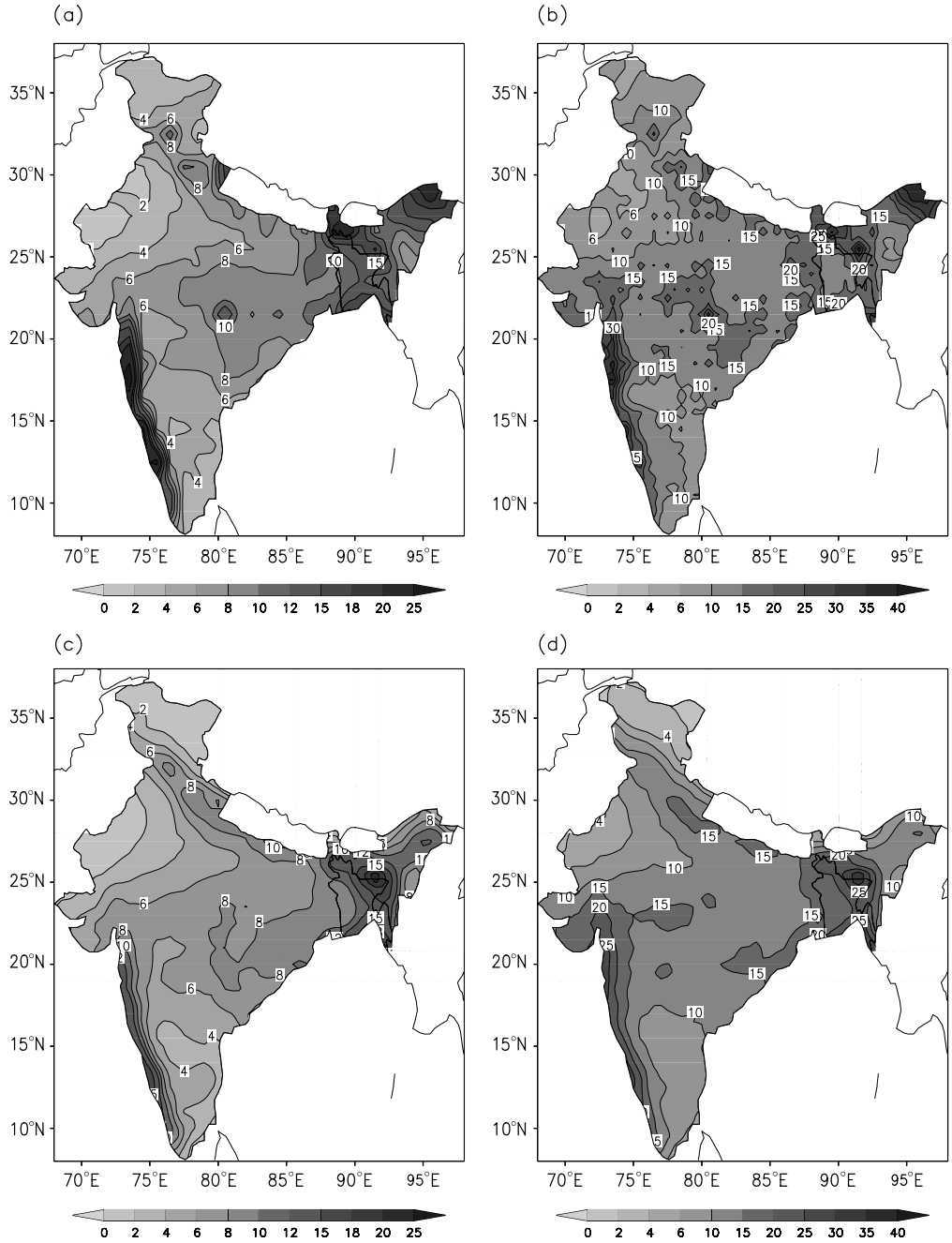

Figure 2. (a) Monsoon seasonal mean rainfall and (b) standard deviation (June, July, August, September) from daily observations of IMD and BWDB gridded data (IMD + BWDB) and (c) monsoon season mean rainfall and (d) standard deviation (June, July, August, September) from daily observations of TRMM.

study area is shown in Figure 1(a). A simple statistical analysis on the data distribution is shown in Table 1. The quality controlled gauge observations are then gridded to a $0.5^\circ \times 0.5^\circ$ grid box template using an inverse distance weighted interpolation. The observations are carried out once a day at 0600 LST (LST = UTC + 6 h).

2.1.2. *IMD gridded rainfall dataset*

The daily gridded rainfall data with a resolution of $1^\circ \times 1^\circ$ from the IMD (Rajeevan *et al.*, 2006) are used in the study. There exist some differences in the preparation of different datasets such as IMD, GPCC and TRMM merged analysis 3B42V6 daily rainfall products over India. The IMD uses the Shepard (1968) interpolation technique for gridding data from individual stations, while GPCC uses the Willmott *et al.* (1985) method for interpolation. The IMD product uses gauge data from 1803 stations to estimate the accumulated rainfall in 24 h ending 0830 IST (IST = UTC + 5.5 h).

2.1.3. *Tropical Rainfall Measuring Mission (TRMM) precipitation product*

The TRMM is a joint space mission between NASA and the Japan Aerospace Exploration Agency (JAXA) designed to monitor and study tropical rainfall. The TRMM is part of NASA's Mission to Planet Earth, a long-term, co-ordinated research effort to study the Earth as a global system. The

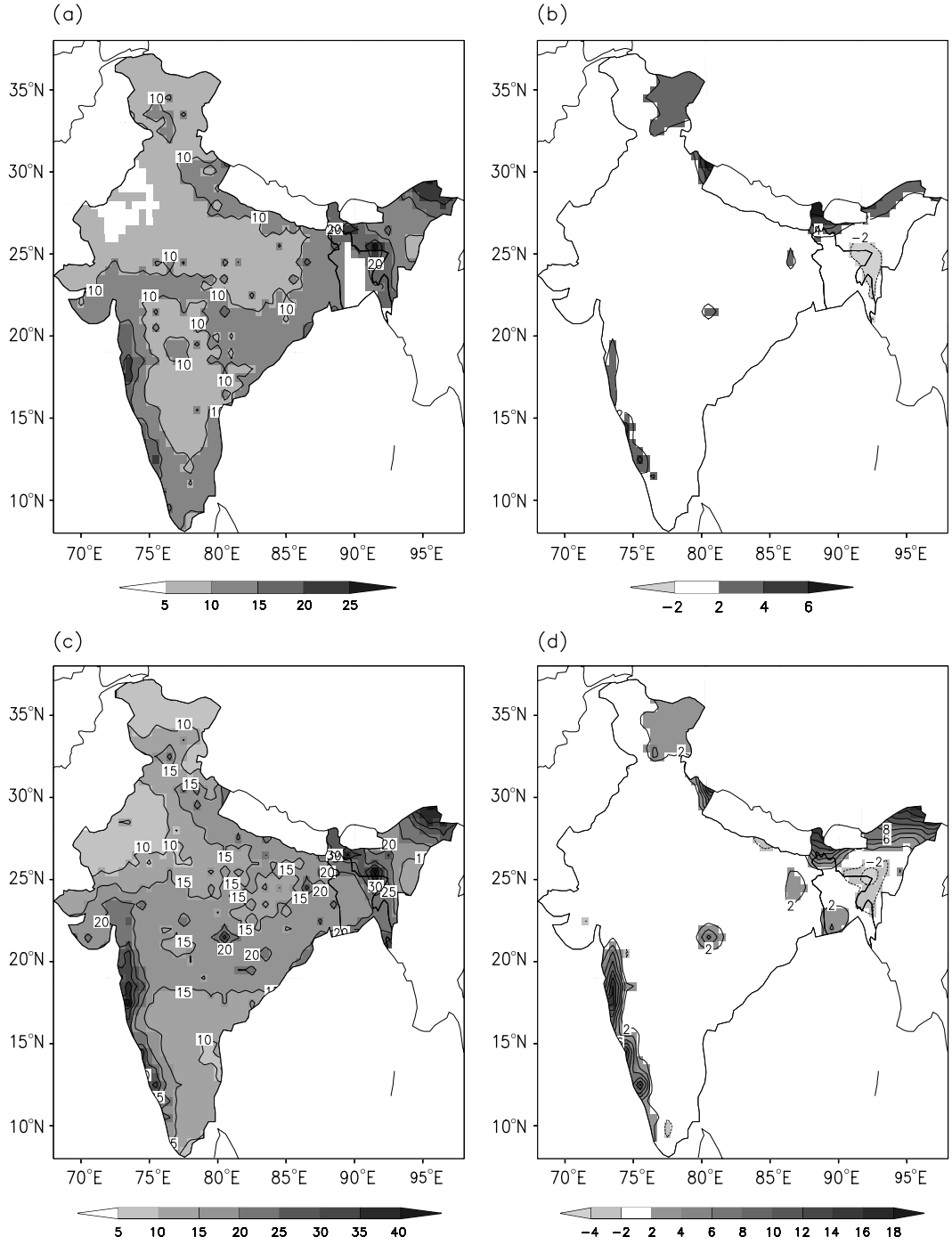

Figure 3. (a) RMSE (1998–2007) and (b) BIAS (1998–2007) for entire period between daily observations of IMD and BWDB Gridded data (IMD + BWDB) and TRMM. (c) RMSE (June, July, August, September) and (d) BIAS (June, July, August, September) for monsoon season between daily observations of IMD and BWDB gridded data (IMD + BWDB) and TRMM.

algorithm used in the TRMM 3B42V6 precipitation product provides a combination of TRMM's merged passive microwave (HQ) and microwave calibrated IR (VAR). The current scheme is a simple replacement of each grid box with the merged passive microwave (HQ) value if available; otherwise the calibrated IR (VAR) value is used. Thereafter, in each grid box all of the multi-satellite (SSM/I, AMSR, AMSU) values in the month are summed and combined with monthly gauge data in order to produce the Version 6 TRMM Data product. Finally, for each grid box, all the 3 h values are scaled to approximately sum to the 3B43 monthly value. This system is developed to produce the TRMM and other data estimates by applying new concepts in merging quasi-global precipitation estimates and taking advantage of the increasing availability of fine-scale input datasets. The overall system is referred to as Version-6 TRMM Multi-Satellite Precipitation Analysis (TMPA) (Huffman *et al*., 2007).

2.1.4. GPCP rainfall dataset

The GPCP precipitation product is a widely used product for climate research studies (Huffman *et al*., 1997, 2001; Adler *et al*., 2000, 2003). The GPCP Combined Product Version 2

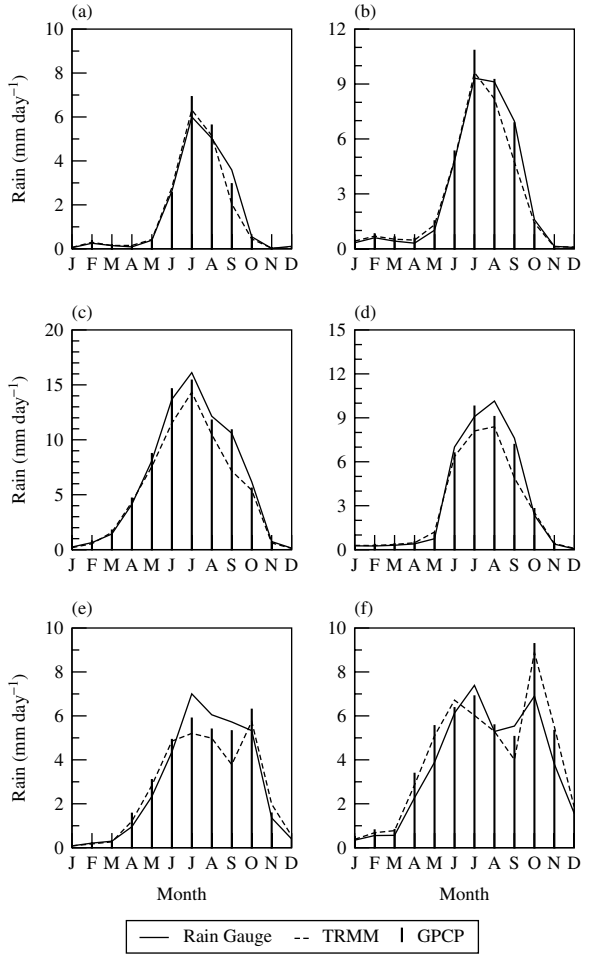

Figure 4. Monthly annual cycle of rainfall from rain gauge (IMD+ BWDB), TRMM 3B42V6 and GPCP for selected land domains over Indian region. (a) Domain 1, (b) Domain 2, (c) Domain 3, (d) Domain 4, (e) Domain 5, (f) Domain 6.

is a near real-time product, including the monthly precipitation data from global telecommunication system (GTS) stations, i.e. the synoptic weather and climate stations. The basic difference between the GPCC and the gauge data is that in the GPCC, the IR and microwave data are used for rainfall estimation along with the rain gauge calibration. Moreover, the GPCP and the TRMM merged products use very similar procedures with different initial inputs, thus simplifying the intercomparison.

2.1.5. GPCC precipitation product

The GPCC dataset is a global land-surface precipitation analysis on monthly timescale based on *in situ* observation data (Beck *et al*., 2005). It comprises near real-time weather and climate data (Synop and CLIMAT) and non-real-time precipitation observations provided by World Meteorological Organisation (WMO) member countries.

2.1.6. Aphrodite precipitation dataset

The Asian Precipitation Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources (APHRODITE; Kamiguchi *et al*., 2010) is a rain gauge-based daily precipitation dataset, which is also used for comparison in this study. APHRODITE uses a new weighted mean method based on Spheremap (Willmott *et al*., 1985) for gridding the station observations (Yatagai *et al*., 2009).

2.2. Methodology for creating a merged precipitation product

There is a number of efforts to compare and validate different satellite rainfall products with other rainfall measurements on global and regional scales. Many of these studies focused on validation and intercomparison of TRMM standard products as well as TRMM Multisatellite Precipitation Analysis (TMPA) products for various regions, e.g., Africa (Adeyewa and Nakamura, 2003; Nicholson *et al*., 2003) and south and southeast Asia (Chokngamwong and Chiu, 2007). Reviews of the other validation studies are comparisons by Chiu *et al*. (2006) of TRMM retrievals (the TRMM Precipitation Radar (PR) 2A25, TRMM Microwave Imager (TMI) 2A12, TRMM Combined Instrument (TCI) 2B31, TRMM 3B42 V5 & V6 and TRMM 3B43 V5 & V6) with Water District (WD) Rain-Gauge rates over New Mexico. The WD data showed high biases for satellite-only algorithms, whereas the merged satellite-gauge products (3B43 or 3B42) and the WD gauge data are better correlated. Bowman (2004) used TRMM TMI and PR data retrievals for comparison with 25 rain gauges on an ocean buoy in the tropical Pacific. Agreement of gauge data with TRMM TMI is found to be better than with TRMM PR over the tropical Pacific Ocean. Adeyewa and Nakamura (2003) validated the TRMM PR and 3B43 rainfall products with the Global Precipitation Climatology Centre (GPCC) global precipitation analysis rain gauge data over the major climatic regions of Africa. Significant seasonal and regional differences are observed. One major conclusion drawn from their assessment is that TRMM PR data do not appear to be adequate to serve as a standalone satellite product. The best agreements are achieved with the TRMM 3B43 data. Nicholson *et al*. (2003) who used gauge data from a network of 920 stations over West Africa to evaluate the TRMM (PR, TMI, 3B43) rainfall products for 1998 obtained a similar result. While the TRMM PR and TMI products showed a net tendency to overestimate gauge measurements, the 3B43 merged product showed excellent agreement with gauge measurements on monthly to seasonal timescales. Chokngamwong and Chiu (2007) compared gauge data from more than 100 rain gauges over Thailand with the TRMM level 3 rainfall products (TRMM 3B42, 3B43 for V5 and V6), results revealed that the 3B42-V6 data correlated best with the gauge data. The above studies for different regions of the globe reiterates that the TRMM Multi-Satellite Precipitation Analysis (TMPA) can be used for flood forecasting studies.

As a first step, station observed precipitation data over Bangladesh from FFWC and BWDB are gridded to the resolution of the IMD rainfall analysis grids for the entire period January 1988 to December 2007. From 1998 to 2007, the TRMM 3B42V6 satellite rainfall measurement is merged with other two datasets to provide a merged product of $0.5^\circ \times 0.5^\circ$ resolution. Daily station datasets obtained from the BWDB are shown in Figure 1(a). The domains selected for comparison between rain gauge and TRMM 3B42V6 are shown in Figure 1(b). Similarly, river systems in the Ganga, Brahmaputra and Meghna (GBM) domain, the GBM river basin and the complex topography over the GBM domain are shown in Figure 1(c) and (d), respectively. Statistics of station observations for the merging period starting from January 1998 to December 2007 are given in Table 1. Stations with at least 70% of observations for the entire duration and exceeding 95% during the analysis period (1998–2007) are used in the analysis, in total 46 Bangladesh rainfall stations are

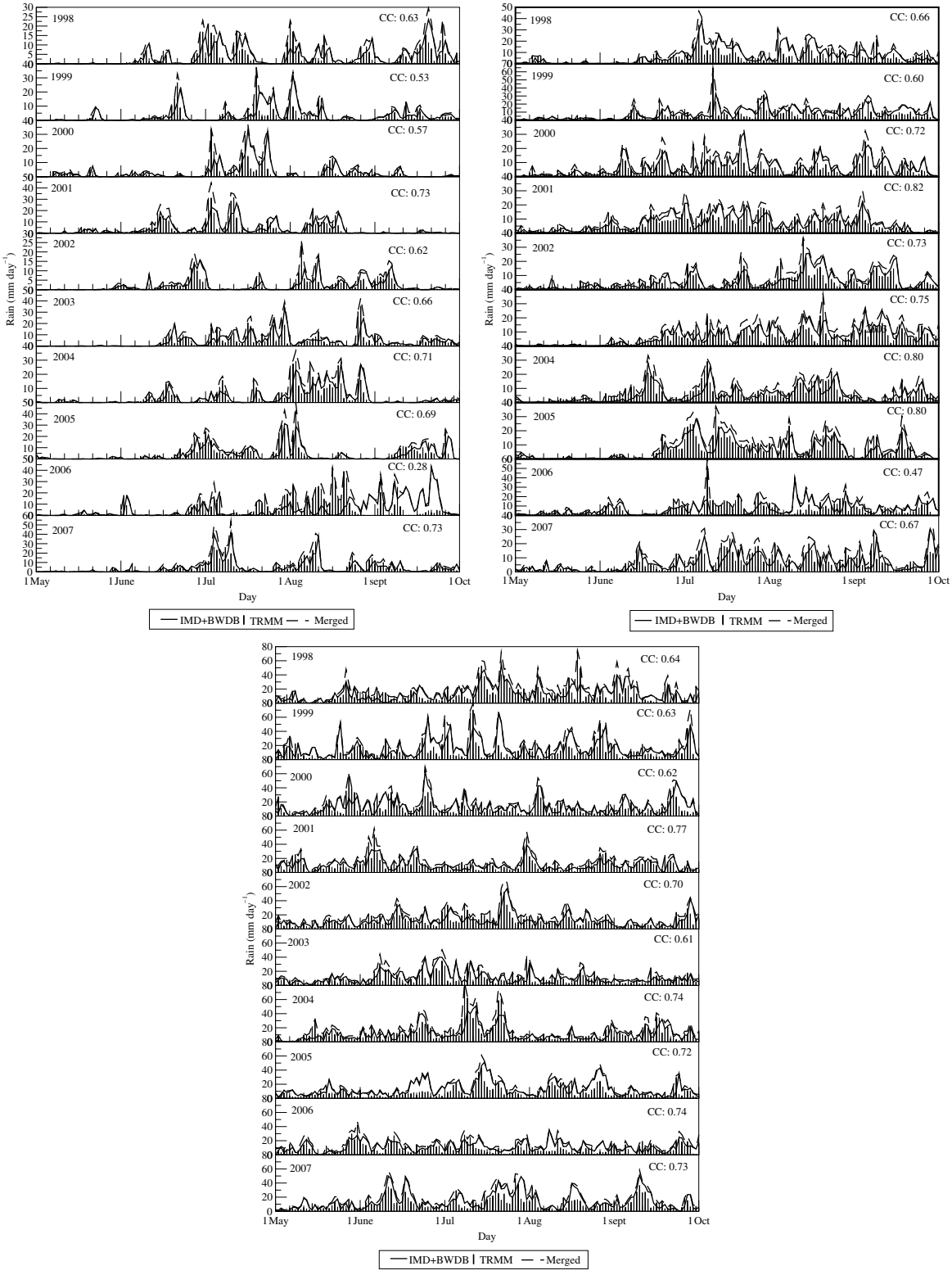

Figure 5. (a) Daily evolution of rainfall from gauge IMD + BWDB (thick line) and TRMM 3B42V6 (vertical bar) for corresponding years from 1998 to 2007 (10 years) over domain-1, the dashed line represents the merged product. Correlation co-efficient values between rain gauge (IMD + BWDB) and TRMM are shown for each year in the top right corner of each panel. b) Same as Figure 5(a), but for domain-2. (c) Same as Figure 5(a), but for Domain 3.

used. The Bangladesh station datasets are made into gridded datasets of uniform grid boxes equivalent to the native grids of the Indian daily gridded dataset. The TRMM 3B42V6 satellite observation datasets with a high temporal resolution of 3 h and a spatial resolution of 0.25° from the space platform are downloaded and processed into daily native grids of the IMD observed rain gauge datasets with a resolution of 0.5° × 0.5° latitude and longitude boxes. The seasonal mean rainfall pattern

and the standard deviations for June to September from the daily averaged TRMM 3B42V6 datasets are compared with the gridded gauge datasets (IMD + BWDB) for the period 1998–2007. Good agreements are achieved between the two datasets, as shown in Figure 2.

2.2.1. *Gridding the rainfall datasets*

To create a rainfall data over the GBM basin, the nearest neighbour interpolation scheme is used. In this method, interpolated values are computed from a weighted sum of the observations. Given a grid point, the search distance is defined as the distance from that point to a given station. The interpolation is restricted to the radius of influence. For search distances equal to or greater than the radius of influence, the grid point value is designated a missing code when there is no station located within this distance.

A simple inverse distance weighted average method (Shepard, 1968) is based on weights that are inversely proportional to the power of the distance from the centre (Xe) of the searching radius (R_s).

The general formula for simple inverse distance weighted average is given below. For a point Xe, the estimated value Z_e is defined as:

$$Z_e = \sum_{i=1}^{n} \{(Z(i) \times W(R(i))\} \tag{1}$$

where $Z\ (i)$ is the sample value at point $X\ (i)$; R_s is the search radius about Xe; $R\ (i)$ the distance of $Xe-X\ (i)$ within R_s, and n the number of data samples within R_s.

The weight function $W\ (i)$ at $X\ (i)$ is calculated by:

$$W(i) = \frac{\left\{\left(\frac{1}{R(i)^P}\right)\right\}}{\left\{\sum_{i=1}^{n} \frac{1}{R(i)^P}\right\}} \tag{2}$$

where $R\ (i)$, represents the distance between the estimate and the i^{th} sample location, and p is the inverse-distance exponent. The simple inverse distance weighted average method is an effective method for creating geospatial averages of irregularly distributed station observations (Weber and Englund, 1994). The search radius is confined to twice the grid size throughout the gridding process. The inverse-distance exponent (P) value is kept at 2 in this analysis.

There are a few advantages as well as disadvantages of gridding station datasets. One of the advantages is the areal coverage of instantaneous rain at a given location can be estimated. The observed station provides information for the nearby areas (i.e., areas without an observational station). On the other hand, a few disadvantages are, in case of an instantaneous flash flood at a given point, the intensity information would be reduced due to gridding and also that the sphere of influence (search radius R_s) does not take the topography of the region into account.

2.2.2. *Merging different precipitation products*

A simple merging technique is adopted in this study. It is noted that discrepancies exist between two observations of the same variable from different platforms (Figures 4 and 5(a)–(c)). The satellite observations have a unique advantage over the point observations, as they have a much larger aerial coverage compared to the point observations. However, the intensity of rainfall may not be captured accurately by space borne instruments, which is one of the major limitations of space borne observations. A significant portion of the India-Bangladesh border showed a negative bias in the rain gauge observations, due to the scarce observational network as well as high mountain terrain (hills around the Shillong Plateau), as shown in Figure 3(b) and (d).

Table 2. Statistics for area average on the entire 10 years data (daily data from 1998 to 2007).

Gauge *versus* TRMM	Domain 1	Domain 2	Domain 3
Correlation	0.59	0.69	0.68
RMSE (mm day^{-1})	3.76	3.89	7.17
BIAS (mm day^{-1})	0.05	0.11	0.69
Mean in gauge (IMD + BWDB) (mm day^{-1})	1.5	2.9	6.2
Mean in TRMM (mm day^{-1})	1.5	2.8	5.5
Standard deviation in gauge (IMD + BWDB) (mm day^{-1})	4.1	5.0	9.3
Standard deviation in TRMM (mm day^{-1})	4.2	4.9	8.6

To exploit the advantages of both the systems, a simple merging technique is used to capture the heavy rainfall events. The correlation between the two products (rain gauge and TRMM 3B42V6) for each year is significantly higher at the 99% level over the near core regions of the GBM basin (Domains 1, 2 and 3) in Figure 5(a)–(c). Therefore, merging these two products at each grid point provides additional information to the grids where observations are sparse. The merging is done in such a way that the highest rainfall at each grid point from both products is included. The merged product would be of use to climate risk management groups and flood forecasters.

3. Results

Satellite-based precipitation products provide high temporal (3 h) and spatial resolutions (order of a few kilometres), but satellites do have biases and random errors that are caused by various factors such as sampling frequency, non-uniform field-of-view of the sensors, and uncertainties in the rainfall retrieval algorithms. It is therefore essential to validate the satellite-derived products with more conventional rain estimates to quantify the direct usability of the products.

3.1. Validation

The spatial mean rainfall and standard deviations for the June to September monsoon season from the rain gauge observations (IMD + BWDB) and the TRMM3B42V6 have some differences (Figure 2(a)–(d)). The spatial patterns of RMSE and BIAS between rain gauge and TRMM 3B42V6 are shown for the entire period (Figure 3(a) and (b)) as well as for the monsoon season only (Figure 3(c) and (d)). The high values of RMSE are noted along the west coast and northeast India,

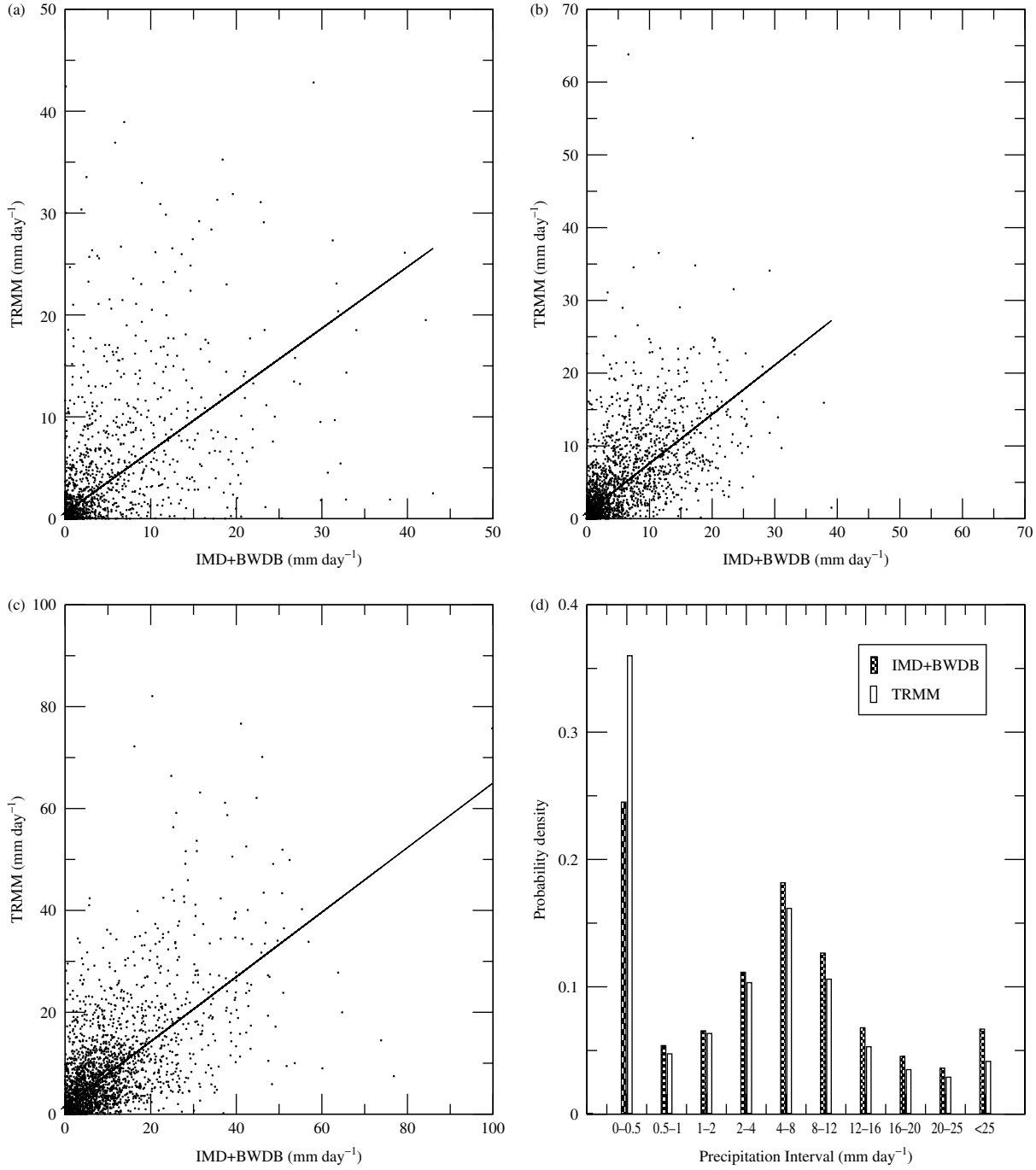

Figure 6. (a) Scatter plot for daily area averaged rainfall over Domain 1 between rain gauge (IMD + BWDB) and TRMM observations for the entire period (1998–2007). The no. of points used is 3652 and the correlation value is 0.6. The equation of fit is given by $y = 0.603x \times +0.5796$. (b) Same as Figure 6(a), but for Domain 2. The no. of points used is 3652 and correlation value is 0.68. The equation of fit is given by $y = 0.675x + 0.8369$. (c) Same as Figure 6(a), but for Domain 3. The no. of points used is 3652 and correlation value is 0.64. The equation of fit is given by $y = 0.634x + 1.5759$. (d) Frequency distribution of daily area averaged rainfall for Domain 3 between rain gauge (IMD + BWDB) and TRMM observations for the entire period (1998–2007).

around 10–30 mm day^{-1} (heavy rainfall regions) (Figure 3(a) and (c)). The bias between the two datasets is not large except for a few patchy regions along the west coast and northeast hilly regions and regions near Bangladesh (Figure 3(b) and (d)). Negative bias exists near the India-Bangladesh border in the rain gauge estimate around the hilly regions of the Shillong Plateau, as shown in Figure 3(b) and (d).

As a simple validation check, arbitrary domains are selected (Figure 1(b)) over the Indian land area covering Bangladesh and the representation of the monthly annual cycle (Figure 4) and daily rainfall evolution over Bangladesh (Domain 3) from gauge observation and satellite measurement (Figure 5(c)) are verified. It is noted from Figure 4 that the evolution of monthly mean annual cycle over the Domains (1, 2, 3 and 4) near the GBM basin is similar in representation from both rain gauge observation and the TRMM 3B42V6 product. The mean monthly cycle is similar over Domains 1, 2, 3 and 4, unlike the domains south of it (Domains 5 and 6). The TRMM 3B42V6

data are also compared with one of the most widely used datasets over land and ocean, the GPCP dataset. The results show that the differences are not large over India between the TRMM 3B42V6 and the GPCP estimate (Figure 4) since both use the Global Precipitation Climatology Centre (GPCC) gauge analysis. Nevertheless, some differences are noted among the three domains for the entire data period between the rain gauge observations and the TRMM 3B42V6 (Table 2). An encouraging and salient point here is that the RMSE values (random errors) are well below the natural variability (standard deviations of the entire data) of IMD + BWDB rain gauge and the TRMM 3B42V6 observations.

The daily evolution of mean precipitation from 1998 to 2007 over the domains near and around the GBM basin area is shown in Figure 5(a)–(c). The correlation co-efficient (CC) value between the rain gauge observation (IMD + BWDB) and the TRMM 3B42V6 for the respective years is shown on the right hand corner of each panel. It is noted that the CC values are significantly higher over Domain 3 in comparison to Domains 1 and 2 (Figure 5(a)–(c)). It is noteworthy that CC values over Domain 3 are consistently higher than for the other two domains for all years, wherein all the CC values are above the 99% significance level. The results provide confidence for merging both the products. Rahman *et al*. (2009) discussed in detail the comparison of the TRMM products (3B42V5, 3B42V6) with the IMD rain gauge datasets over India. There are also a few studies which deal with validation and use of both rain gauge and satellite products over India (Krishnamurti *et al*., 2009; Mitra *et al*., 2009).

The scatter plots of daily area averaged rain estimate over the nearest three domains around the GBM basin are shown in Figure 6(a)–(c). From the scatter plots it is evident that the TRMM underestimates over most of the domains, but captures the heavy rainfall to some extent. The equations of fit for each plot are also shown in the figure caption. The comparison of frequency distribution of daily area averaged precipitation interval between the gauge observation and the TRMM 3B42V6 over Domain 3 is shown in Figure 6(d). It can be easily identified from the distribution that the TRMM slightly underestimates heavy precipitation. The correlation co-efficient between the gauges and the TRMM is significantly higher at the 99% level of significance over the domains, 0.59 (Domain 1), 0.69 (Domain 2) and 0.68 (Domain 3) in Figure 6(a)–(c). The result is satisfactory, but not enough to form a basis for merging both rain gauge and the TRMM 3B42V6 rainfall estimate to catch the heavy rainfall over each grid in the GBM domain.

A similar scatter plot and distribution of precipitation interval between the TRMM 3B42V6 and the rain gauge observations is performed from individual grid points over the domain 3 and is shown in Figure 7(a) and (b). The correlation co-efficient between the rain gauge observation (IMD + BWDB) and the TRMM 3B42V6 is found out to be 0.42 from 4.77×10^5 grid points (above the 99% significance level). The equation of fit is likewise given in the figure caption. The frequency distribution of gridded rainfall over domain 3 between the rain gauge and the TRMM observations for the entire period is shown in Figure 7(b). The results reveal that the TRMM 3B42V6 overestimates light rainfall, in the range of 0–2 mm day^{-1} and underestimates moderate and moderately high rainfalls, in the ranges 2–25 and 25–75 mm day^{-1}, respectively. It is encouraging to see that the TRMM 3B42V6 is close to the grid observation in estimating the heavy rainfalls, >75 mm day^{-1}.

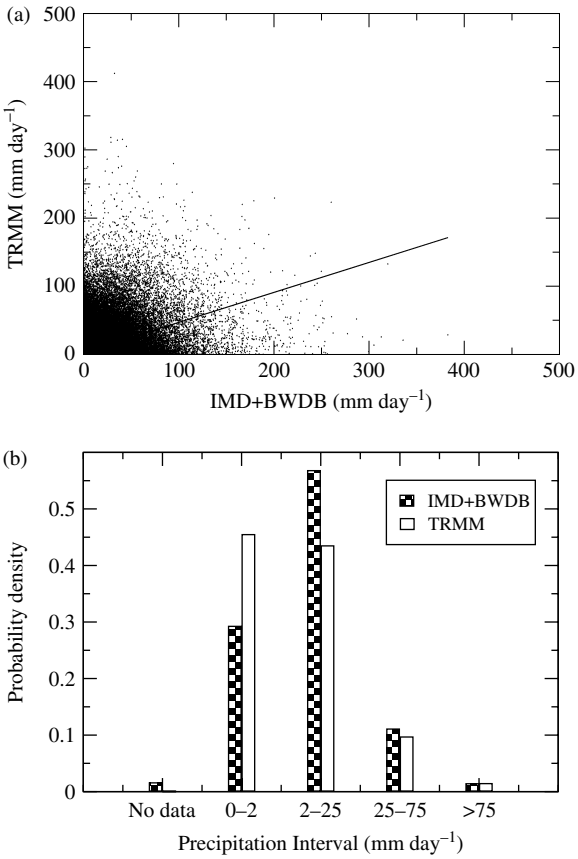

Figure 7. Scatter plot from each grid point over domain-3 between rain gauge (IMD + BWDB) and TRMM observations for the entire period (1998–2007). The no. of points used is 477 787 points and correlation value is 0.43. Equation of fit is given by $y = 0.439x + 3.0247$. (b) Frequency distribution of grid rainfall for domain-3 between rain gauge (IMD + BWDB) and TRMM observations for the entire period (1998–2007).

3.2. Percentile calculation

In order to see the discrepancies between the rain gauges (IMD + BWDB) and the TRMM 3B42V6 in heavier rainfall estimates clearly, the 95-percentile rank is calculated. The estimated above 95-percentile values from the rain gauges (IMD + BWDB) and the TRMM 3B42V6 are plotted in a scatter diagram (Figure 8).

The percentile rank is calculated using the following formula:

$$n = \frac{P}{100} \times N + \frac{1}{2} \tag{3}$$

where n is rank; P is percentile probability and N is the number of samples. The 95-percentile rank for the rain gauges (IMD + BWDB) yielded a value of 30.265 mm day^{-1} and the 95-percentile rank for the TRMM 3B42V6 resulted in a value of 30.463 mm day^{-1}. The scatter diagrams are made based on the 95-percentile rank for rain gauges (IMD + BWDB) (Figure 8(a)), TRMM 3B42V6 (Figure 8(b)) and for both observations (Figure 8(c)). The overestimation of precipitation in the rain gauge compared to the TRMM is evident from the scatter plots (Figure 8(a)–(c)). The total rainfall grids for the entire period (1998–2007) over Domain 3 correspond to 4.77×10^5 points. The 95-percentile rainy grids correspond to 24 091 grid values in IMD + BWDB and 24 073 grid values in the TRMM, whereas only 1.5% of the total grid points

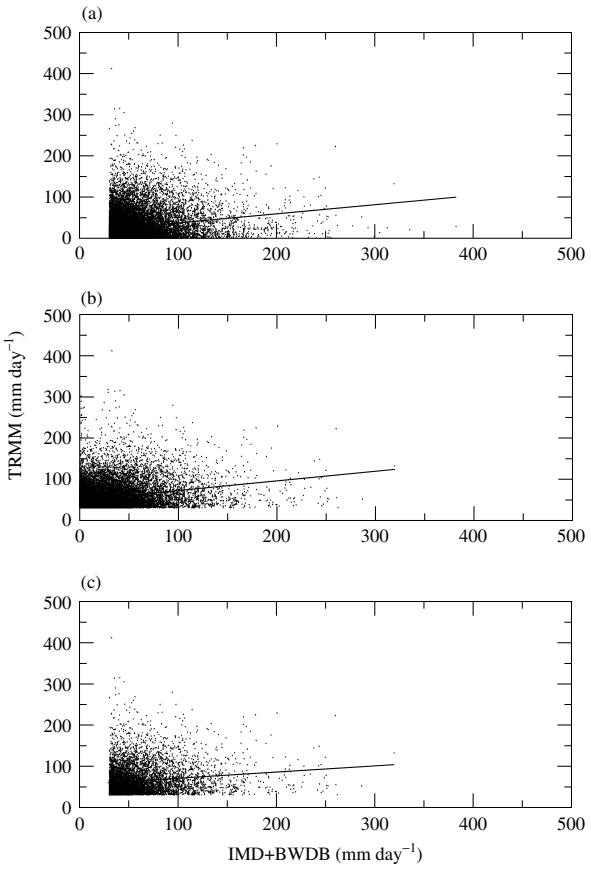

Figure 8. (a) Scatter plot from each grid point over Domain 3 between rain gauge (IMD + BWDB) and TRMM observations based on TRMM 95 percentile threshold. The equation of fit is given by $y = 0.221x + 14.992$. (b) Scatter plot from each grid point over Domain 3 between rain gauge (IMD + BWDB) and TRMM observations based on rain gauge (IMD + BWDB) 95 percentile threshold. The equation of fit is given by $y = 0.232x + 49.534$. (c) Scatter plot from each grid point over Domain 3 between rain gauge (IMD + BWDB) and TRMM observations based on TRMM 95 percentile threshold. The equation of fit is given by $y = 0.1495x + 56.04$.

(7356 points) capture heavy rainfall above the 95-percentile in both platforms on a corresponding day basis. Out of 24 073 grid values in the TRMM estimating heavy rainfall (>95-percentile), less than one-third (7356) coincide with the IMD + BWDB grids estimating heavy precipitation (>95-percentile). Moreover, the rain gauges and the TRMM show large bias in estimating heavier precipitation, nevertheless the correlation between the rain gauges (IMD + BWDB) and the TRMM does not fall below the 95% significance level, which is encouraging and forms the basis for merging both the datasets. The equations of fit for the scatter diagrams are given in the figure caption.

3.3. Merged product

The merged product is a combination of both gauge (IMD + BWDB) observations and the TRMM 3B42V6 product with a heavier estimation of rainfall at each grid point. It is noted in Figure 9 that the rainfall intensity over the major grid points is higher in the merged product, especially over northeast India, near the border of Bhutan with India, India and Bangladesh and the GBM domain in particular.

In order to represent correctly where the difference in the precipitation comes from in the estimates, the GPCC datasets, the gauges or the satellites, local climatology is constructed for the 1998 monsoon from the IMD + BWDB gauges (Figure 10(a)) and the GPCC data (Figure 10(b)), and the difference between those two is shown in Figure 10(d). Since the GPCC gridded climatology values are added to the satellite sensor values (TMI and PR estimates in TRMM) to create the final precipitation analysis (TRMM 3B42V6), the difference between GPCC and the gauges highlights the problems with the gauge observations. From Figure 10, the discrepancies are mostly between the gauge estimates (IMD + BWDB) and the GPCC estimates. A similar discrepancy (Figure 10(e)) is also found when the recently launched Asian gridded precipitation product APHRODITE (Figure 10(c)) is considered. The notion of taking the greatest of several estimates intentionally biases the result: this might be acceptable for applications such as flood forecasting since the resulting increase in false alarms increases the probability of detection as well.

3.4. Case studies

The basis for merging is to obtain a higher estimate of rainfall during floods and this is validated for a few cases, one such case is in the 1998 monsoon year (Figure 9). During the summer of 1998, over 60% of Bangladesh is inundated for nearly 3 months. The seasonal mean rainfall (July to September) in the rain gauge observations (Figure 9(a)) near the India-Bangladesh border areas is not heavier. This might be due to sparse observations over the border regions. By merging the TRMM 3B42V6 information, the merged product gives a higher estimate over these grids (Figure 9(d)), which is additional information for flood forecasters in Bangladesh.

3.4.1. Case study 1: July to August 2004

To check for heavy flood events during recent years, a known heavy flood during the week of 27 July to 3 August 2004 (widespread flooding over Bangladesh) is considered for examination. The rain estimates shown in Figure 11(a) are the mean accumulated rainfall between 27 July and 3 August 2004. It can be noted from Figure 11(a) that the heavy floods over Bangladesh are not captured well in the rain gauge observation, whereas the TRMM (Figure 11(b)) captures some heavy rain near the India-Bangladesh border. In the bottom panel it is seen that the rainfall underestimation is between 10 and 20 mm day^{-1} in the rain gauge near the India Bangladesh border (Gauge – TRMM) (Figure 11(c)). It is obvious that the gauge and the TRMM merged product gives a higher estimate as shown in Figure 11(d).

3.4.2. Case study 2: August 2007

Similarly, there were heavy floods during the week of 3–13 August 2007, wherein by 13 August 2007 the confirmed death toll in Bangladesh was 405. This event is considered as case 2. The rain estimates shown in Figure 12(a) is the mean accumulated rain between 3 and 13 August 2007. The heavy rainfall west and northwest of the India-Bangladesh border is not captured well in the rain gauge observations, whereas the TRMM 3B42V6 captures this feature (Figure 12(b)). In the bottom panel underestimation in the gauge observation is between 5 and 10 mm day^{-1} over these areas (Gauge – TRMM) (Figure 12(c)). The merged product produces a higher estimate around the Bangladesh region (Figure 12(d)).

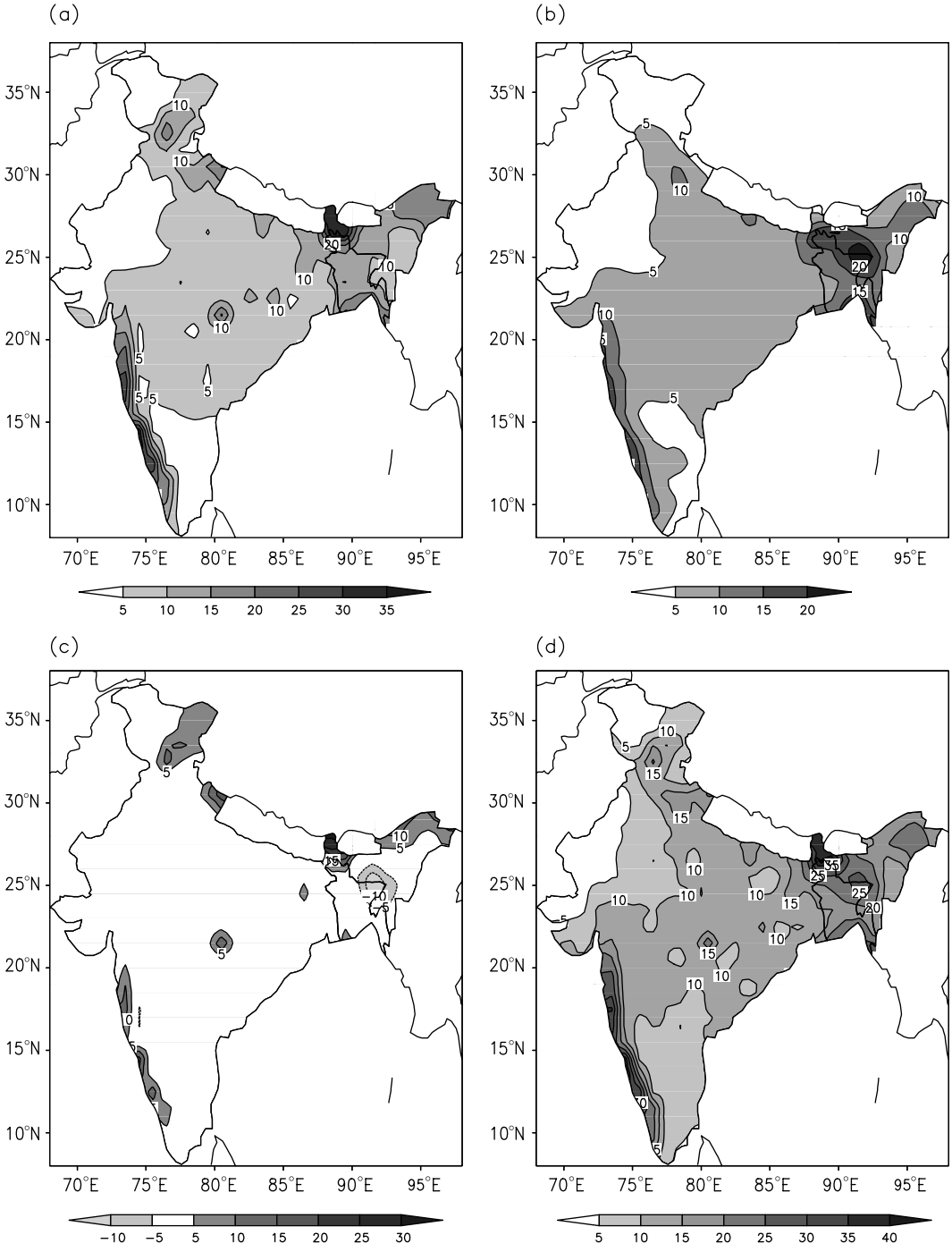

Figure 9. (a) Monsoon seasonal mean rainfall (June, July, August, September) for the year 1998 from daily observation of IMD and BWDB gridded data (IMD + BWDB) and (b) TRMM. (c) Difference in seasonal mean rainfall for June, July, August, September and (d) merged product of the above two products.

In the above two cases, the flooding is severe in the Brahmaputra river. In both cases, underestimation is seen in the rain gauge data compared to the TRMM-3B42V6 data along the border areas, therefore merging the satellite information provides additional rain estimates in these regions which will help the flood forecasters and also help in correcting the biases in the model climate forecasts aimed at flood forecasting over Bangladesh.

Bookhagen and Burbank (2010) presented a detailed hydrological budget over the Himalayas. A similar hydrological budget would be needed for flood forecasting over the GBM domain. The GBM domain has a complex terrain due to the presence of the high Himalayas on its northern periphery as well as the presently available datasets pose major limitations on modelling the hydrological budget over the GBM domain. Despite using the high temporal resolution remote sensing datasets and a large network of gauge observations, this study has several shortcomings. The study has only focussed on the precipitation from the TRMM, rain gauges and gridded precipitation products from Bangladesh and India, respectively, but does not include either snowmelt or runoff from the high Himalayan Mountains or river discharge measurements from

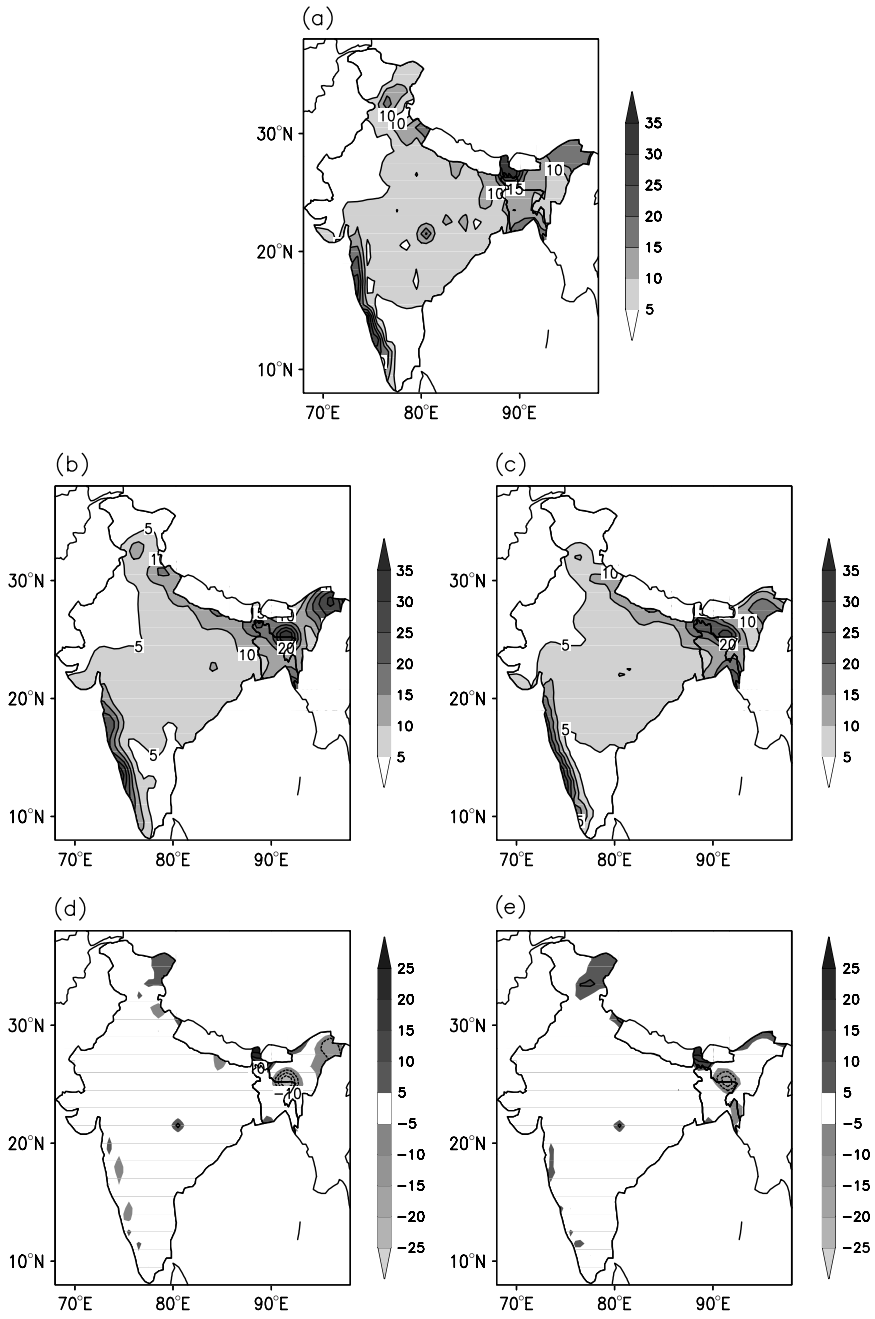

Figure 10. (a) Monsoon seasonal mean rainfall (June, July, August, September) for the year 1998 from daily observation of IMD and BWDB gridded data (IMD + BWDB). (b) Monsoon seasonal mean rainfall (June, July, August, September) for the year 1998 from daily observation of GPCC and (c) monsoon seasonal mean rainfall (June, July, August, September) for the year 1998 from daily observation of Aphrodite data. (d) Difference in seasonal mean rainfall for June, July, August, September between GPCC and rain gauge and (e) difference in seasonal mean rainfall for June, July, August, September between Aphrodite and rain gauge.

the major rivers to quantify the inundation or flooding over Bangladesh. Nevertheless, the data serves as additional information to the flood forecasters and climate modelers aimed at flood forecasting.

4. Conclusion

This paper has focused extensively on the Ganga, Brahmaputra and Meghna (GBM) river basin domain precipitation and the usability of multi-source precipitation datasets. The efforts in comparing multi-source precipitation dataset over the GBM domain are more complicated due to its geographical location and steep topography, but its potential application is enormous. Although there are discrepancies in the observations among different platforms, the aim to achieve a higher probability rainfall estimate over the GBM could be acceptable for applications such as flood forecasting, since the resulting increase in false alarms increases the probability of detection as well. The merged rainfall dataset ($0.5^\circ \times 0.5^\circ$ resolution) developed using the Indian gridded dataset, the Bangladesh rain gauge station observations and the TRMM 3B42V6 products would serve as a useful precipitation product for climate

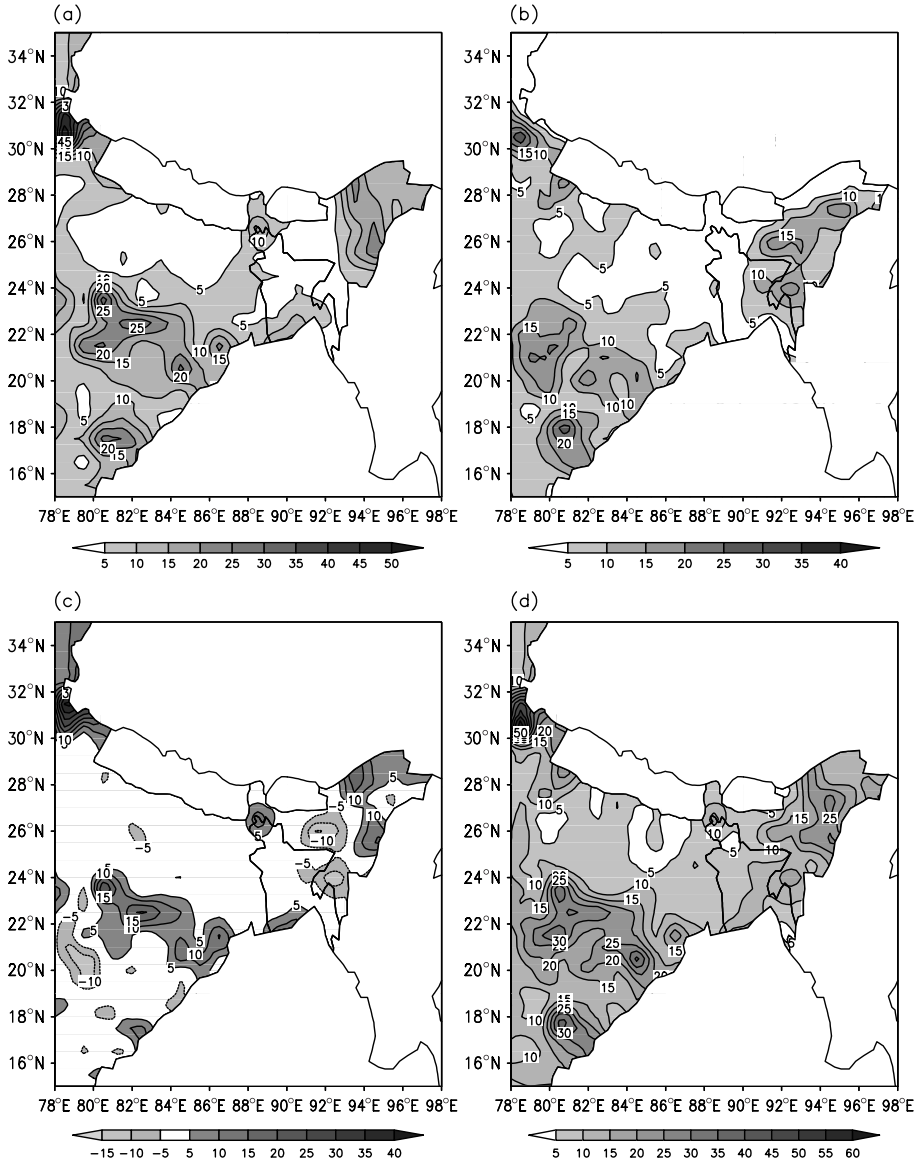

Figure 11. (a) 27 July to 3 August 2004 (flood) from daily observation of IMD and BWDB Gridded data (IMD + BWDB) and (b) TRMM. (c) Difference in mean rainfall between Gauge and TRMM and (d) merged product of the above two products.

risk management groups involved in flood forecasting activities, water modelling and climate change adaptation of heavy floods. The data would also serve as a validation tool for the high-resolution climate model simulation over the GBM basin. The data development is aimed at providing additional information on rain estimates over the upstream areas and low-lying areas near the Ganges and Brahmaputra catchments of the India-Bangladesh border. Based on the results from this study the TRMM-3B42V6 dataset is reasonable in capturing the spatial and temporal rainfall distribution over the GBM region. Also, it was found that by merging the satellite and the rain gauge measurements, additional information over data sparse regions in the GBM domain is obtained. The GBM domain has one of the most complex topographies in the globe and observations over the domain are very limited. The hydrological budget over this domain is very dynamic and complex to model (Bookhagen and Burbank, 2010). Given the above-mentioned limitations, this paper only provides additional information by merging the rain gauge observations with remote sensing observations and does not provide a complete hydrological budget, which would be desirable for flood forecasting.

Acknowledgements

We thank A. R. Subbiah for providing facilities to carry out this work. This work is supported by ADPC, Bangkok, Thailand. The work is carried out as part of the DANIDA supported project being implemented by IWM, Bangladesh, which is gratefully acknowledged. We wish to acknowledge the BWDB, FFWC of Bangladesh for providing the station rainfall datasets and the IMD of India for providing gridded rainfall data and NASA/GSFC for archiving and providing the 3B42-V6 data from their ftp site. The authors specially like to thank Dr. George Huffman for his comments in improving the manuscript and also like to thank the editor Prof. P. J. Burt and two other anonymous reviewers for their constructive comments. Free Software packages like Fedora, Intel Fortran compiler; GrADS and Xmgrace have been used for data analysis and graphics.

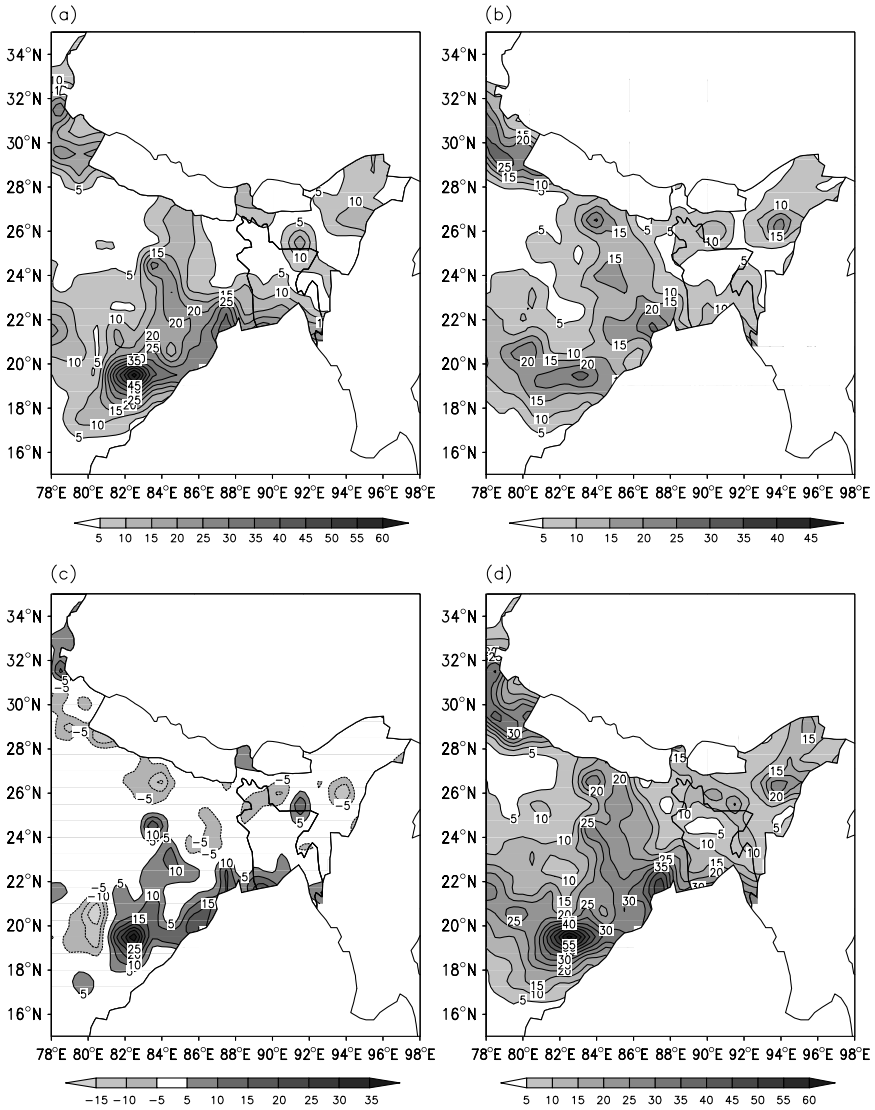

Figure 12. Same as Figure 11 but for 6 August to 13 August 2007.

References

Adeyewa ZB, Nakamura K. 2003. Validation of TRMM radar rainfall data over major climatic regions in Africa. *J. Clim.* **42**: 331–347.

Adler RF, Bolun DT, Curtis S, Nelkin EJ. 2000. Tropical rainfall distributions determined using TRMM combined with other satellite and raingauge information. *J. Appl. Meteorol.* **39**: 2007–2023.

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie PP, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P, Nelkin E. 2003. The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis 1979-present. *J. Hydrometeorol.* **4**: 1147–1167.

Ali AMS. 2007. September 2004 flood event in South-western Bangladesh: a study of its nature, causes, human perception and adjustments to a new hazard. *Nat. Hazards* **40**: 89–111.

Beck C, Grieser J, Rudolf B. 2005. A new monthly precipitation climatology for the global land areas for the period 1951 to 2000. DWD, Klimastatusbericht KSB 2004, ISSN 1437–7691, ISSN 1616–5063 (Internet), ISBN 3-88148-402-7, 181–190.

Bookhagen B, Burbank DW. 2010. Toward a complete Himalayan hydrological budget: spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *J. Geophys. Res.* **115**: F03019, DOI: 10.1029/2009JF001426.

Bowman KP. 2004. Comparison of TRMM precipitation retrievals with rain gauge data from ocean buoys. *J. Clim.* **18**: 178–190.

Brammer H. 1990a. Floods in Bangladesh: I. Geographical background to the 1987 and 1988 floods. *Geogr. J*. **156**: 12–22.

Brammer H. 1990b. Floods in Bangladesh: II. Flood mitigation and environmental aspects. *Geogr. J*. **156**: 158–165.

Chiu LS, Liu Z, Vongsaard J, Morain S, Budge S, Neville P, Bales C. 2006. Comparison of TRMM and water district rain rates in New Mexico. *Adv. Atmos. Sci.* **23**: 1–13.

Chokngamwong R, Chiu L. 2007. Thailand daily rainfall and comparison with TRMM products. *J. Hydrometeorol.* **9**: 256–266.

Chowdhury AMR. 1988. The 1987 flood in Bangladesh: an estimate of damage in twelve villages. *Disasters* **12**: 294–300.

Chowdhury AMR. 1989. Flood-1988 as seen by satellites. *Bangladesh Quest* **1**: 52–54.

Huffman GJ, Adler RF, Arkin P, Chang A, Ferraro R, Gruber A, Janowiak J, McNab A, Rudolf B, Schneider U. 1997. The GPCP combined precipitation datasets. *Bull. Am. Meteorol. Soc.* **78**: 5–20.

Huffman GJ, Adler RF, Bolvin DT, Gu G, Nelkin EJ, Bowman KP, Hong Y, Stocker EF, Wolff DB. 2007. The TRMM multi-satellite precipitation analysis: quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. *J. Hydrometeorol.* **8**: 38–55.

Huffman GJ, Adler RF, Morrissey MM, Bolvin DT, Curtis S, Joyce R, McGavock B, Susskind J. 2001. Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeorol.* **2**: 36–50.

Kamiguchi K, Arakawa O, Kitoh A, Yatagai A, Hamada A, Yasutomi N. 2010. Development of APHRO_JP, the first Japanese high-resolution daily precipitation product for more than 100 years. *Hydrol. Res. Lett.* **4**: 60–64.

Krishnamurti TN, Mishra AK, Simon A, Yatagai A. 2009. Use of a dense rain-gauge network over India for improving blended TRMM products and downscaled weather models. *J. Meteorol. Soc. Jpn.* **87**: 393–412.

Mitra AK, Bohra AK, Rajeevan MN, Krishnamurti TN. 2009. Daily Indian precipitation analysis formed from a merge of rain-gauge data with the TRMM TMPA satellite-derived rainfall estimates. *J. Meteorol. Soc. Jpn.* **87**: 265–279.

Nicholson SE, Some B, McCollum J, Nelkin E, Klotter D, Berte Y, Diallo BM, Gaye I, Kpabeba G, Ndiaye O, Noukpozounkou JN, Tanu MM, Thiam A, Toure AA, Traore AK. 2003. Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa. Part II: validation of TRMM rainfall products. *J. Appl. Meteorol.* **42**: 1355–1368.

Pant GB, Kumar KR. 1997. *Climates of South Asia*. John Wiley: Chichester; 320 pp.

Rahman SH, Sengupta D, Ravichandran M. 2009. Variability of Indian summer monsoon rainfall in daily data from gauge and satellite. *J. Geophys. Res.* **114**: D17113, DOI: 10.1029/2008JD011694.

Rajeevan M, Bhate J, Kale JD, Lal B. 2006. High resolution daily gridded rainfall data for the Indian region: analysis of break and active monsoon spell. *Curr. Sci.* **91**: 296–306.

Saiful AI, Anisul H, Sujit BK. 2010. Hydrologic characteristics of floods in Ganges–Brahmaputra–Meghna (GBM) delta. *Nat. Hazards* **54**: 797–811.

Shepard D. 1968. A two-dimensional interpolation function for irregularly spaced data. *Proceedings of the 1968 ACM National Conference*, New York, NY; 517–524.

Weber DD, Englund EJ. 1994. Evaluation and comparison of spatial interpolators II. *Math. Geol.* **26**(5): 589–603.

Willmott CJ, Rowe CM, Philpot WD. 1985. Small-scale climate maps: a sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *Am. Cartogr.* **12**: 5–16.

Yatagai A, Arakawa O, Kamiguchi K, Kawamoto H, Nodzu MI, Hamada A. 2009. A 44 year daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *SOLA* **5**: 137–140.

