Daily rainfall statistics of TRMM and CMORPH: A case for trans-boundary Gandak River basin

BRIJESH KUMAR^{1,*} , KANHU CHARAN PATRA¹ and VENKAT LAKSHMI²

¹*Department of Civil Engineering, National Institute of Technology Rourkela, Odisha 769 008, India.*

²*Department of Earth and Ocean Sciences, University of South Carolina, Columbia, SC, USA.*

**Corresponding author. e-mail: brijeshag1684@gmail.com*

Satellite precipitation products offer an opportunity to evaluate extreme events (flood and drought) for areas where rainfall data are not available or rain gauge stations are sparse. In this study, daily precipitation amount and frequency of TRMM 3B42V.7 and CMORPH products have been validated against daily rain gauge precipitation for the monsoon months (June–September or JJAS) from 2005–2010 in the trans-boundary Gandak River basin. The analysis shows that the both TRMM and CMORPH can detect rain and no-rain events, but they fail to capture the intensity of rainfall.

The detection of precipitation amount is strongly dependent on the topography. In the plains areas, TRMM product is capable of capturing high-intensity rain events but in the hilly regions, it underestimates the amount of high-intensity rain events. On the other hand, CMORPH entirely fails to capture the high-intensity rain events but does well with low-intensity rain events in both hilly regions as well as the plain region. The continuous variable verification method shows better agreement of TRMM rainfall products with rain gauge data. TRMM fares better in the prediction of probability of occurrence of high-intensity rainfall events, but it underestimates intensity at high altitudes. This implies that TRMM precipitation estimates can be used for flood-related studies only after bias adjustment for the topography.

1. Introduction

Hydrological modelling of watersheds requires continuous and long-term meteorological data. The time series of input data should be at least at a daily time step for climate change and extreme events – flood and drought modelling (Jones *et al.* 1999; Folland *et al.* 2000). Precipitation is perhaps the most important input data for these studies. Obtaining reliable input data is a challenge, particularly in developing countries and trans-boundary basins. Currently, data sharing is quite limited across national boundaries due to economic and national security barriers (Vigliano *et al.* 2010; Plengsaeng *et al.* 2014). In this regard, satellite precipitation products can be very useful particularly

in trans-boundary basins, where the exchange of precipitation data from rain gauge data does not happen in near real time.

The necessity of reliable satellite precipitation records for extreme events (flood and drought) and climate change has been investigated in numerous error analysis studies worldwide (Rahman and Simon 2006; Vernimmen *et al.* 2012; Kneis *et al.* 2014). Rain gauge records are the most reliable source of precipitation estimate. However, the spatial distributions of rain gauges and the consistency in the record have certain limitations. High-resolution satellite precipitation products such as TRMM (Tropical Rainfall Measuring Mission) and CMORPH (Climate Prediction Center Morphing Technique) provide considerable improvements in

Keywords. TRMM 3B42V.7; CMORPH; Himalaya; extreme events; satellite rainfall estimation.

the spatial and temporal variability of the rainfall dataset. Satellite precipitation products are derived using different retrieval algorithms and some products are biased compared to the rain gauge datasets (Jiang *et al.* 2012). Therefore, satellite precipitation products must be validated with rain gauge data to estimate their accuracy. The previous assessment of the satellite products has shown considerable agreement against rain gauge precipitation in different parts of the world (Wang and Wolff 2010; Rahman *et al.* 2011; Wu and Zhai 2012; Chen *et al.* 2013; Prakash *et al.* 2015).

The Gandak River basin has very sparse rain gauge station network and rain gauges have missing data in the time series. It may cause difficulty in the simulation of floods in the Gandak River basin. The precipitation pattern is highly variable in space and time in this basin, and almost 90% rainfall occurs during the South Asian Monsoon (June–September). The average annual rainfall varies from 2030 mm in the northern mountainous area to 1100 mm in the southern plains region of the basin (GFCC 2004).

In the recent past, several studies have been carried out to validate precipitation data from TRMM and other satellite products in the Indian subcontinent. Detailed literature review for validation of satellite rainfall over Indian subcontinent has been presented in table 1. The table shows continuous effort made by researchers to validate the various satellite rainfall products over different zones of the Indian subcontinent including TRMM 3B42 V6 over the western states of India by Nair *et al.* (2009), TRMM 3B42 V7 over Northern Himalaya by Bharti and Singh (2015), TRMM 3B42 V6 and ERA40 in the east, over Bangladesh by Rahman *et al.* (2011), TRMM 3B42 and TRMM 3B42RT in central India, over the Mahanadi river basin by Kneis *et al.* (2014), TRMM 3B42 V6 over Nepal by Muller and Thompson (2013), and TRMM 3B42 V7 and CMORPH over the Tibetan Plateau and China by Wu and Zhai (2012), to name a few.

Table 1 explicitly highlights the gap in the validation of TRMM 3B42 V7 and CMORPH satellite product over the southern Himalayas.

This work is an attempt to validate a gauge adjusted satellite precipitation product, the TRMM 3B42 V7, and an exclusive satellite-based precipitation product, the CMORPH, to see their usefulness for extreme events analysis over southern Himalayan Gandak River basin. ‘We preferred the TRMM 3B42 V7 gauge adjusted satellite precipitation over others such as PERSIAN or GsMaP because the TRMM 3B42 V7 algorithm takes special care for mountainous regions using SRTM 30 m DEM as an auxiliary data (TRMM Precipitation Radar Team 2011). The PERSIAN and GsMaP retrieval algorithms use meteorological data from

ground observations while CMORPH is exclusively a satellite-based precipitation product (Sorooshian *et al.* 2000; Joyce *et al.* 2004; Shige *et al.* 2009).

The Gandak River basin is a trans-boundary basin – the basin area falls in India, China and Nepal (figure 1). The basin consists of some of the highest and most complex terrains, having elevations that range from 33 m in the southern plains to a maximum of 8143 m in the northern part, towards Himalayan Mountains.

Rainfall data were collected from 30 rain gauge stations maintained by the Department of Hydrology and Meteorology, Nepal, and the India Meteorological Department (IMD). All 30 rain gauge stations data were analysed for their consistency using double mass curve method and only six rain gauge stations were found reliable following this evaluation. The example consistency check (double mass curve graph for Marsharakh) is presented in figure 2. Widely used satellite precipitation products TRMM 3B42 V7 and CMORPH are quantitatively evaluated with rain gauge precipitation at these six rain gauge stations. The results provide insights into the capability of the satellite precipitation products to detect the frequencies and amount of precipitation in the complex topography of Gandak River basin. This study may provide the foundation for bias correction in the satellite products over the southern Himalayan basins.

2. Data and methods

2.1 Satellite data

We used the TRMM 3B42 V7 and CMORPH precipitation products in this analysis. The TRMM 3B42 data are available from 1998 to present while CMORPH from 2002 to present. TRMM 3B42 data can be found at http://gdata1.sci.gsfc.nasa.gov/daacbin/G3/gui.cgi?instance_id=TRMM_3-Hourly and CMORPH from ftp://ftp.cpc.ncep.noaa.gov/precip/global_CMORPH/3-hourly 025 deg.

The TRMM 3B42 V6 processing algorithm is described in Huffman *et al.* (2007) while changes made in the processing algorithm of TRMM 3B42 V7 are described in Huffman *et al.* (2010) and are summarised here. TRMM 3B42 V7 has a new processing algorithm – the Goddard profiling algorithm (GPROF) 2010 for the passive microwave (PMW) based estimation. It uses existing TRMM storm profile records, PMW based brightness temperature, and precipitation rates and it replaces a reference database created using a cloud model in version 6. Additionally, version 7 also incorporates satellite data from Microwave Humidity Sounder (HMS), Special Sensor Microwave Imager

Table 1. A detailed literature review for validation of satellite precipitation over the Indian subcontinent.

Authors	Data	Region/country	Key findings
Bharti and Singh (2015)	TRMM 3B42 V7	Northern Himalaya, India	<ul style="list-style-type: none"> Product does well up to 2000 m altitude. Critical altitude was found at 3100 m. The product has a high magnitude bias at 99.99 percentile. That means the product is not good for high rain events
Prakash <i>et al.</i> (2015)	TRMM 3B42 V7 and CPC_RFE 2.0	Whole India	<ul style="list-style-type: none"> TRMM product overestimates over the eastern India while underestimates along the west coast. RFE 2.0 Underestimates over the entire country except eastern India.
Guo <i>et al.</i> (2014)	CMORPH	Tibet, China	<ul style="list-style-type: none"> CMORPH is failing to capture morning peak. Diurnal cycle in rainfall frequency is basically due to the topography.
Kneis <i>et al.</i> (2014)	TRMM 3B42 and TRMM 3B42RT	Mahanadi basin, India	<ul style="list-style-type: none"> Discrepancies between rain gauge precipitation and TRMM precipitation for higher intensities. Significant negative bias in TRMM 3B42RT data.
Xue <i>et al.</i> (2013)	TRMM 3B42 V7 and TRMM 3B42 V6	Wangchu Basin, Bhutan	<ul style="list-style-type: none"> 3B42 V7 product has improved upon 3B42 V6's underestimation for the entire basin. A significant improvement in the discharge simulation using 3B43 V7 compared to TRMM 3B42 V6.
Muller and Thompson (2013)	TRMM 3B42 V6	Nepal	<ul style="list-style-type: none"> Reported error in the TRMM 3B42 V6 product and proposed a frequency-domain based bias correction.
Krakauer <i>et al.</i> (2013)	TRMM 3B43, GSMaP, CMORPH, and PERSIANN	Nepal	<ul style="list-style-type: none"> Monthly evaluation of these satellite products indicates that TRMM B43 precipitation is promising on others for water resources planning.
Wu and Zhai (2012)	TRMM 3B42 V7 and CMORPH	Tibetan plateau, China	<ul style="list-style-type: none"> The ability to accurately detect rainfall for both the products depends on topography. TRMM and CMORPH underestimates and overestimates the frequency of lighter rain events, respectively.
Duo <i>et al.</i> (2011)	CPC_RFE 2.0	Tibet, China	<ul style="list-style-type: none"> Product performs well with mean CC 0.74, POD 73% and FAR between 1 to 12% except few aspersions towards Himalayan range. They concluded that the product can be used for flood related studies.
Shrestha <i>et al.</i> (2011)	CPC_RFE 2.0	Narayani, Himalaya, Nepal	<ul style="list-style-type: none"> The product underestimates rain events. The simulated discharge with CPC_RFE2.0 rainfall underestimated runoff peak by 50%, while, after application of three ratios based bias adjustment, it shows considerable improvement.
Rahman <i>et al.</i> (2011)	ERA40, TRMM 3B42 V7	Bangladesh	<ul style="list-style-type: none"> Three verification methods proposed by Murphy (1993) indicate that both the products perform well and can be used for design flood estimation studies.

Table 1. (Continued.)

Authors	Data	Region/country	Key findings
Islam <i>et al.</i> (2010)	TRMM 3B42 V6	Nepal	<ul style="list-style-type: none"> • TRMM trends are similar to rain gauge data. • TRMM underestimates the amount with few exception of overestimation.
Nair <i>et al.</i> (2009)	TRMM 3B42 V6	Maharashtra, India	<ul style="list-style-type: none"> • TRMM 3B42 V.6 product found able detect dry and wet monsoon. • Moreover, the timing of ground-based rain events coincides with the satellite rain events.

Figure 1. Location map of the study area.

(SSMI) and it also has the capability to ingest new satellite data in the future. A new infrared brightness temperature data is introduced for the period of Climate Prediction Centre (CPC), 4 km global IR data (i.e., January 1998–February 2000). Furthermore, a single reprocessed Advanced Microwave Sounding Unit-B (AMSU-B) satellite dataset replaces the previous version for which two dissimilar calibration periods were used, and this removes some of the internal discrepancy existing in TRMM Multi-satellite Precipitation Analysis (TMPA) Version 6. The algorithm also incorporates a monthly scale gauge bias correction

based on Global Precipitation Climatology Centre (GPCC) reanalyzed data (version 6.0). GPCC version 6.0 data includes anomalies instead of amounts from denser rain gauge network.

The TRMM precipitation radar (TPR) estimate was corrected both globally and regionally, over the mountainous regions, and are described in a technical document (TRMM Precipitation Radar Team 2011) and the key findings are summarised here. In version 6, the algorithm was found to be inaccurate over the mountainous regions and affected by a high level of surface clutter for rain echo. It has also been seen to miss-locate surface echoes because of

Figure 2. Double mass curve of Marsharakh. The curve reveals a consistent rain gauge station data.

erroneous elevation data and cover-up by strong signals from intense rainfall. The TRMM 3B42 V7 algorithm accounts for elevation data of mountains like the Himalayas and Andes from the Shuttle Radar Topography Mission (SRTM 30 m). Furthermore, the TRMM 3B42 V7 algorithm uses a repeat search algorithm for the surface echo that advances its detection and thus, it determines cluster free rain regions in the entire storm profile.

The CMORPH algorithm produces a very high temporal and spatial resolution precipitation data with near-total global coverage. This technique uses rainfall estimates exclusively derived from low orbit satellite microwave observations (Joyce *et al.* 2004) and whose features are transported using spatial propagation information that is obtained entirely from geostationary satellite infrared (IR) sensors. At present, CMORPH incorporates precipitation estimates derived from the passive microwave sensors -14 and 15 (SSM/I), 16, 17 and 18

Table 2. Detailed description of the precipitation datasets used in this study.

Dataset		Spatial resolution	Temporal resolution	Description of the data	Comment
Rain gauge data	IMD	Daily	Irregular	Ground-based 15 station rain gauge data provided by India Meteorological Department (IMD), Gov. of India. Strict data acquisition quality. Data is available from 2000–2012.	Data gaps and inconsistent
	DHM	Daily	Irregular	Ground-based 15 station rain gauge data provided by Department of Hydrology and Meteorology (DHM), Gov. of Nepal. Strict data acquisition quality. Data is available from 2000–2012.	Data gaps and inconsistent
TRMM 3B42		3 h	Gridded (0.25×0.25)	TRMM 3B42 V7 were derived by algorithm 3B42 (Huffman <i>et al.</i> 2007) and provided by the NASA Goddard Space Flight Center. The data coverage is 50°S–50°N and 180°W–180°E.	Continuous and consistent
CMORPH		3 h	Gridded (0.25×0.25)	The data has produced using NOAA's CMORPH, which uses rainfall estimates exclusively derived from low orbit satellite microwave observations (Joyce <i>et al.</i> 2004). The data coverage is 60°S–60°N and 180°W–180°E.	Continuous and consistent

2.2 Gauge data

Ground-based rain gauge data are available on a daily basis (accumulated over 24-hr period) and, collected/reported at 8.30 AM local time. In the Gandak River basin, there are 15 rain gauge station data maintained by the Department of Hydrology and Meteorology (DHM), Government of Nepal and 15 rain gauge station data maintained by the India Meteorological Department (IMD), Government of India in the Indian portion of the Gandak River basin.

Rain gauge data collected from IMD and DHM are having numerous data gaps. There are many methods to estimate missing precipitation data, viz., simple average, normal ratio, linear or multi-regression, inverse distance weighting methods, the coefficient of correlation weighting method (CCWM) to name a few. Ramesh *et al.* (2005) concluded that CCWM is superior to other methods, and it can provide a better deterministic estimation of missing data in any climatic region. CCWM (equations 6 and 7) was used to estimate the missing data in the time series.

$$P_m = \frac{\sum_{i=1}^n P_i (r_{mi})^K}{\sum_i^n (r_{mi})^K} \quad (6)$$

$$r_{mi} = \frac{\text{Cov}(P_i P_m)}{\sigma_{P_i} \sigma_{P_m}} \quad (7)$$

where P_m is the missing precipitation at station m , P_i is the precipitation at the i th station, r_{mi} is the correlation coefficient between station m and i and K is the exponent that can be optimised for the particular case.

A thorough study was conducted to check the consistency and continuity of the dataset from each station. The consistency check was done using the double mass curve method while the length of missing data at each site was considered for the continuity test. An example to verify the consistency of time series for a station, namely Marsharakh, is given in figure 2. After a detailed study, we found only six consistent raingauge stations in the study area. These rain gauges are shown in figure 1 and are: (i) Marsharakh (India), (ii) Chatiya (India), (iii) Bagaha (India), (iv) Kuncha (Nepal), (v) Thamachit (Nepal), and (vi) Baghara (Nepal).

2.3 Validation methods

In this study, TRMM 3B42 and CMORPH precipitation data for monsoon season, June, July, August and September (JJAS) from 2005–2010 are evaluated with daily rain gauge data.

Murphy (1993) described three types of ‘goodness’ to verify satellite precipitation estimates,

viz., consistency, quality, and value. Wilks (2006) defined a partial list of scalar attributes of forecast qualities: accuracy, bias, reliability, resolution, discrimination and sharpness. Stanski *et al.* (1989) divided verification of forecast into visual, dichotomous (categorical variable verification), and continuous variable verification categories. We used forecast/satellite verification categories proposed by Stanski *et al.* (1989) to verify TRMM and CMORPH satellite precipitation estimate in this study.

2.3.1 Visual verification

In visual verification, we examine the difference between the satellite estimates and observed rain gauge values. Visual verification is an instantaneous judgement to distinguish the error between the satellite estimates and rain gauge precipitation. We often use exploratory graph techniques such as time series plot, histogram, and Cumulative Distribution Function (CDF) referring to the use of visual verification.

2.3.2 Yes/no dichotomous

The yes/no dichotomous verification (categorical variable verification) is based on the statistical quantification of the scalar attribute of the forecast (Wilks 2006). In this method, ‘yes’ describes that an event will happen and ‘no’, the event will not happen. A threshold may be set to distinguish ‘yes’ and ‘no’ event in the rainfall (1, 2, 5, 10, 20, 50 mm d⁻¹) as per the time scale of the dataset WMO/TD-No. 1485 (2008). Wu and Zhai (2012) described a threshold 0.1 mm day⁻¹, i.e., equivalent to the minimum rain gauge observation. This is a low threshold and will be able to distinguish the ‘yes’ and ‘no’ event of light rain. In this study, we have also applied the threshold of 0.1 mm day⁻¹. Brown *et al.* (1997) provided a contingency table for the query of ‘yes’ and ‘no’ estimation and it is presented in table 2.

From the contingency table, four combinations of chance are possible between rain gauge and satellite observations. These are listed below.

- (i) Hit: both gauge and satellite event occur
- (ii) Miss: Gauge event occurs, but satellite event does not occur
- (iii) False alarm: Gauge event does not occur, but satellite event occurs
- (iv) Null or correct negative: both events do not occur

Based on these dichotomous, a variety of statistics were developed by Wilks (2006) and WMO/TD-No. 1485 for validation of satellite observations.

Table 4. Contingency table for yes/no dichotomous between daily gauge and satellite precipitation.

		Gauge estimate (G)		
		Yes	No	Total
Satellite estimate(s)	Yes	$\begin{cases} a \\ \text{Hit} \end{cases}$	$\begin{cases} b \\ \text{False alarm} \end{cases}$	$\begin{cases} a+b \\ \text{Satellite observed Yes} \end{cases}$
	No	$\begin{cases} c \\ \text{Miss} \end{cases}$	$\begin{cases} d \\ \text{Correct negative} \end{cases}$	$\begin{cases} c+d \\ \text{Satellite observed No} \end{cases}$
Total		$\begin{cases} a+c \\ \text{Gauge observed Yes} \end{cases}$	$\begin{cases} b+d \\ \text{Gauge observed No} \end{cases}$	$\begin{cases} N \\ \text{Total} \end{cases}$

The statistical parameters used in this study are described below (meaning of notations a, b, c and d are given in table 4).

Bias score:

$$B = \frac{a+b}{a+c} = \frac{\text{Satellite observed events}}{\text{Gauge observed events}} \quad (8)$$

The Bias score only depends on the marginal of the events; it does not measure correspondence. $B=1$, indicates an unbiased estimation whereas $B<1$ and $B>1$ indicates underestimation and overestimation, respectively.

Probability of detection (Hit rate):

$$\text{POD}_{\text{rain}} = \frac{a}{a+c} = \frac{\text{Hits}}{\text{Gauge observed events}} \quad (9)$$

$$\begin{aligned} \text{POD}_{\text{no rain}} &= \frac{d}{b+d} \\ &= \frac{\text{Correct negative}}{\text{Non Gauge observed events}} \end{aligned} \quad (10)$$

The probability of detection measures the fraction of all gauge events correctly observed by the satellite events. The best POD score is 1, but the occurrence of the best POD score does not mean that satellite precipitation is equivalent to gauge precipitation.

False alarm ratio:

$$\begin{aligned} \text{FAR}_{\text{rain}} &= \frac{b}{a+b} \\ &= \frac{\text{False alarms}}{\text{Satellite observed events}} \end{aligned} \quad (11)$$

$$\begin{aligned} \text{FAR}_{\text{no rain}} &= \frac{c}{c+d} \\ &= \frac{\text{Miss}}{\text{Non-satellite observed events}} \end{aligned} \quad (12)$$

False alarm ratio is the fraction between false alarms to the total number of satellite observed events. The value varies from $0 \leq \text{FAR} \leq 1$. The best score is zero, but the best score of FAR does not mean perfect satellite observation.

Probability of false alarm detection (POFD false alarm rate):

$$\text{POFD} = \frac{b}{b+d} = \frac{\text{False alarms}}{\text{Non Gauge observed events}} \quad (13)$$

This is the ratio between the false alarms to the non-gauge observed events. The value varies from $0 \leq \text{POFD} \leq 1$ and the best score for POFD is zero.

Accuracy (fraction correct):

$$\text{ACC} = \frac{a+d}{N} = \frac{\text{Correct satellite events}}{\text{All satellite events}} \quad (14)$$

This is the fraction of all satellite events that were correct. The ACC value ranges from $0 \leq \text{ACC} \leq 1$ and the best score is 1. The best score of ACC means a perfect match of gauge precipitation by satellite precipitation.

Threat score (Critical Success Index):

$$\text{TS} = \text{CSI} = \frac{\text{Hits}}{\text{All satellite or Gauge events}} \quad (15)$$

TS or CSI is the fraction between hits to all satellite or Gauge events. The value varies from $0 \leq \text{TS} \leq 1$ and the best score is 1. The best score indicates perfect detection.

Skill score:

$$\text{SS} = \frac{A - A_{\text{ref}}}{A_{\text{perf}} - A_{\text{ref}}} \quad (16)$$

where A is accuracy score, e.g., ACC or TS, A_{perf} is the accuracy of a perfect forecast, i.e., 1. A_{ref} is the accuracy of the reference forecast and can be calculated as:

$$\begin{aligned} A_{\text{ref}} &= \left(\frac{(a+b)}{N} \right) * \left(\frac{(a+c)}{N} \right) \\ &\quad + \left(\frac{(d+c)}{N} \right) * \left(\frac{(d+b)}{N} \right) \end{aligned} \quad (17)$$

SS=1 indicated perfect detection of gauge precipitation by satellite precipitation. SS>0 indicates a skilful detection, better than the reference, whereas SS<0 indicates a less than skilful detection, lower than reference detection.

2.3.3 Continuous variable verification

The verification of continuous variables measures how the value of satellite estimates differs with the value of gauge observations. This can be done by various summary scores and are described below

(i) Bias (mean error, systematic error):

There can be two types of bias, viz.

- Mean error (additive bias), which describes the error between the mean of satellite ($\bar{S}$) and mean of rain gauge estimate ($\bar{G}$). The mean error may be (-ve) or (+ve) based on the underestimation and overestimation of records.
- Systematic error (multiplicative bias) is the ratio between $\bar{S}$ and $\bar{G}$. The best score of systematic error is 1. The equation is given below:

$$B_{\text{add}} = (\bar{S} - \bar{G}), \quad B_{\text{mult}} = \frac{\bar{S}}{\bar{G}}. \quad (18)$$

(ii) Mean absolute error:

Mean absolute error measures the absolute deviation of satellite precipitation from gauge precipitation. The formula is given below:

$$\text{MAE} = \frac{1}{N} \sum_i |S_i - G_i|. \quad (19)$$

(iii) Mean squared error (MSE) and root MSE (RMSE):

$$\text{MSE} = \frac{1}{N} \sum_i (S_i - G_i)^2, \quad \text{RMSE} = \sqrt{\text{MSE}} \quad (20)$$

This statistical parameter is sensitive to outliers and favours satellite estimates to avoid large deviation from the mean.

(iv) Correlation coefficient:

This is a statistical measure of the random error (scatter around the best fit) and can be formulated as below

$$CC = \frac{\text{Cov}(S_i, G_i)}{\sigma_S \sigma_G} \quad (21)$$

where $\text{Cov}(S_i, G_i)$ is the co-variation between the gauge and satellite data and $\sigma_S \sigma_G$ are the variance

of the satellite and gauge data respectively. The range varies from $-1 \leq CC \leq 1$ and the best score is 1.

3. Result and discussions

Three satellite data verification methods described by Stanski *et al.* (1989), viz., visual verification, dichotomous verification, and continuous variable verification have been discussed in section 2.3. These methods were applied to analyse the satellite products at six rain gauge locations from 2005–2010 for monsoon months (JJAS). The results are discussed in this section.

3.1 Visual verification

Figure 4 shows the scatter plot comparison between 6-year daily average satellite and rain gauge rainfall at the six stations. The stations are given here in ascending order of elevation from mean sea level (i) Marsharakh 62 m, (ii) Chatiya 70 m, (iii) Bagaha 93 m, (iv) Kuncha 725 m, (v) Thamachit 1942 m, and (vi) Baghara 2659 m. TRMM daily average precipitation overestimates the amount (Duncan and Biggs 2012; Bharti and Singh 2015) for low rain events (intensities $\leq 7 \text{ mm d}^{-1}$) at all stations. Furthermore, for medium rain events ($7 \text{ mm d}^{-1} \geq \text{event intensity} \leq 15 \text{ mm d}^{-1}$), it is either close to the rain gauge data or overestimates the amount at first four stations (elevation $\leq 2000 \text{ m}$) but for rest two stations (elevation $\geq 2000 \text{ m}$), it underestimates amount apparently. TRMM underestimates the amount for high rainfall events (intensities $\geq 15 \text{ mm d}^{-1}$) at all stations. On the other hand, CMORPH is reliable only for very low rain events (intensities $\leq 5 \text{ mm d}^{-1}$) and in the plains (elevation $\leq 1000 \text{ m}$). CMORPH underestimates amount whose intensities are $\geq 5 \text{ mm d}^{-1}$ at all rain gauge stations.

Figure 5 is a bar diagram representation of monthly satellite and rain gauge precipitation. It shows that the monthly precipitation value of CMORPH is much lower than the gauge rainfall for all stations. On the other hand, the TRMM rainfall is comparable to the gauge rainfall for first four rain gauge stations (elevation $\leq 2000 \text{ m}$) whereas for two others, it underestimates.

The precipitation intensities can be categorised by either fixed quantity of precipitation which occurs in a 24-hr period or by the percentile of daily precipitation amounts. Here, in figure 6, we grouped satellite precipitation along with the gauge precipitation in various intensity classes to construct a histogram of relative frequency. This figure shows that almost 75% rain events are in class 0–10 mm day^{-1} meaning that only 25% rain

Figure 4. Scatter plot of the TRMM and CMORPH with daily Gauge data at (a) Marsharakh, (b) Chatiya, (c) Bagaha, (d) Kuncha, (e) Thamachit and (f) Baghara.

events have rain intensity $\geq 10 \text{ mm d}^{-1}$. In the frequency class $0\text{--}10 \text{ mm day}^{-1}$, TRMM better reproduces the rain events with little underestimation whereas, CMORPH overestimates the relative frequency in this class that means CMORPH may have high false alarms for low-intensity rain events. Furthermore, in the class intervals $\geq 10 \text{ mm d}^{-1}$ TRMM overestimates the relative frequency for first three plain stations (elevation $\leq 1000 \text{ m}$), and for the other three (elevation $\geq 1000 \text{ m}$) it underestimates the rainfall relative frequency. This suggests that TRMM may produce a false alarm of the high-intensity rain events in the plains whereas it may miss at high altitudes. On the other hand, CMORPH completely fails to capture the relative frequency in the class intervals with intensity $\geq 10 \text{ mm d}^{-1}$ for all stations.

Figure 7 shows the Cumulative Distribution Function (CDF) plot between the satellite and rain gauge data. CDF shape of TRMM estimates nearly matches to rain gauge CDF for first three stations (elevation $\leq 1000 \text{ m}$) but it is lower for the other three stations (elevation $\geq 1000 \text{ m}$). This implies that TRMM underestimates the probability of occurrence of rain events in high altitudes, i.e., TRMM may miss high-intensity rainfall events. On the other hand, the CDF plot of CMORPH is above to CDF plot of rain gauge data. It means

CMORPH has a greater probability of producing a false alarm for all rainfall events at all stations.

3.2 Yes/no dichotomous

Yes/no-dichotomous statistics, described in section 2.3.2, determines how well the satellite estimates can detect rain events and their frequencies. The results presented in table 5, exhibit the reliability of satellite-based precipitation is strongly dependent on topography.

Bias score is ≥ 1 at all stations for both the satellite estimates, except at Baghara rain gauge where it is 0.93 and 0.99 for CMORPH and TRMM respectively. This implies that both satellite measures overestimate the number of rain events. POD_{rain} is more than 0.80 always; it indicates that both the satellite products can probably detect rainy events in more than 80% cases at all stations. The $\text{POD}_{\text{no-rain}}$ is considerable (almost 60%) for both satellites in the plain stations (elevation $\leq 1000 \text{ m}$) but it falls to 16% for stations falling in the $\geq 1000 \text{ m}$ elevation from MSL. It means both satellite products are poor at detecting no-rain events at higher altitudes and this may result in higher false alarms in these areas. FAR_{rain} , for both satellites, ranges between 0.11 and 0.50.

Figure 5. Monthly time series plot of TRMM, CMORPH and Gauge data for monsoon months (JJAS, 2005–2010).

TRMM and CMORPH both show a higher false alarm ratio towards the plains. On the other hand, the $FAR_{no-rain}$ results are opposite to FAR_{rain} . $FAR_{no-rain}$ is a maximum of 0.66 for CMORPH at most elevated location – Baghara, whereas it drops to a minimum of 0.10 at the low elevated place – Marsharakh. TRMM has relatively better $FAR_{no-rain}$ (0.55) meaning that, TRMM is better at revealing the no-rain events towards high elevated areas.

The POFD analysis shows that TRMM estimates are better than CMORPH in plains but poorly

performs in elevated stations. The minimum PODF is 0.37 for TRMM at Chatiya while maximum (0.93) for CMORPH at Thamachit. The accuracy of TRMM ranges from 0.81 to 0.60 whereas for CMORPH, it is 0.76 to 0.59 (a perfect accuracy score is 1). Thus, both satellite estimates have a minimum 0.19 in accuracy. The lower value of TS score towards plains shows more asymmetry between rain events and non-rain events. The SS score is more than 0 for both satellites at all stations which indicates skilful detection by both the satellite products. The SS score range for TRMM estimates

Figure 6. Histogram of relative frequency of daily TRMM, CMORPH and Gauge data (for JJAS, 2005–2010) at rain gauge stations (a) Marsharakh, (b) Chatiya, (c) Bagaha, (d) Kuncha, (e) Thamachit and (f) Baghara.

is 0.47–0.2 while for CMORPH 0.45–0.02 which means that CMORPH does not perform well.

3.3 Continuous variable verification

The continuous variable verification is basically a statistical quantification of the differences in the amount of satellite and rain gauge precipitation. This includes the mean absolute error (MAE); root mean square error (RMSE), correlation coefficient (CC), bias additive (mean error) and bias multiplicative (systematic error). Results of the continuous variable verification are presented in table 6.

MAE value ranges from 7.63 to 16.21 for TRMM whereas it ranges from 6.66 to 18.43 for CMORPH. TRMM and CMORPH both have a minimum MAE of 6.63 and 6.66 respectively, at Chatiya, in the plain. On the other hand, TRMM and CMORPH both show a maximum MAE of 16.21 and 18.43 respectively at Baghara in the elevated area. RMSE values indicate quite a similar perception as MAE. It ranges from 13.98 to 27.11 for TRMM and 16.06 to 27.53 for CMORPH. The CC values are in the range 0.25–0.75 and 0.10–0.58 for TRMM and CMORPH respectively. TRMM indicates +ve mean error, 0.82, 1.75, 2.63, at first three plain stations (elevation ≤ 1000 m), whereas

–ve, –3.14, –6.87, –9.43 for rest three elevated stations (elevation ≥ 1000 m). On the other hand, CMORPH indicates a –ve mean error at all the rain gauge stations. A +ve mean error indicates an overestimation of rainfall amount by satellite estimates whereas a –ve mean error indicates an underestimation. The systematic error indicates >1 and <1 values for TRMM, at first three plain stations and another three elevated stations respectively whereas <1 for CMORPH for all stations. The systematic error value >1 and <1 indicates overestimation and underestimation of rain amount respectively. Thus, the results of mean error and systematic error indicate that TRMM overestimates the rain amount in the plains whereas underestimates towards mountains. On the other hand, both systematic error and mean error indicates that CMORPH clearly underestimates the rain amount regardless of it being the plains or mountains. The statistics of continuous variable verification, RMSE, MAE, CC and two types of error mean and systematic indicate that CMORPH is associated with more quantitative error compared to TRMM.

Figure 8 shows that the CC between satellite and gauge data is sensitive to the elevation. Correlation coefficient follows the decreasing trend with an increase in elevation. The gradient of a decreasing trend for both satellites, TRMM and CMORPH,

Figure 7. Cumulative distribution functions (CDFs) between daily TRMM, CMORPH and Gauge data at (a) Marsharakh, (b) Chatiya, (c) Bagaha, (d) Kuncha, (e) Thamachit and (f) Baghara.

Table 5. Yes/no dichotomous validation statistics of daily satellite precipitation products relative to gauge data.

Station	Data	Yes/no dichotomous								
		Bias	POD _{rain}	POD _{no-rain}	FAR _{rain}	FAR _{no-rain}	POFD	Acc.	TS	SS
Marsharakh	TRMM	1.66	0.80	0.62	0.5	0.11	0.38	0.69	0.45	0.38
	CMORPH	1.72	0.85	0.61	0.5	0.1	0.39	0.68	0.46	0.38
Chatiya	TRMM	1.34	0.84	0.63	0.37	0.16	0.37	0.73	0.56	0.46
	CMORPH	1.34	0.83	0.61	0.38	0.18	0.39	0.7	0.55	0.42
Bagaha	TRMM	1.59	0.9	0.59	0.43	0.09	0.41	0.71	0.53	0.44
	CMORPH	1.62	0.9	0.57	0.44	0.1	0.43	0.69	0.52	0.45
Kuncha	TRMM	1.8	0.95	0.23	0.47	0.17	0.77	0.60	0.51	0.20
	CMORPH	1.75	0.93	0.26	0.47	0.2	0.74	0.58	0.51	0.18
Thamachit	TRMM	1.5	0.95	0.16	0.37	0.32	0.84	0.64	0.61	0.2
	CMORPH	1.55	0.94	0.07	0.39	0.54	0.93	0.6	0.59	0.02
Baghara	TRMM	0.99	0.88	0.47	0.11	0.55	0.53	0.81	0.79	0.35
	CMORPH	0.93	0.82	0.45	0.12	0.66	0.55	0.76	0.74	0.24

is the same (0.0001) but it differs in intercept, 0.67 and 0.45, for TRMM and CMORPH respectively. To examine the significance of the decreasing trend of correlation coefficient with elevation, a hypothesis testing method t -test is performed. The t -test statistics of TRMM shows that $t_{\text{calculated}} (-ve, 5.439) > t_{\text{critical}} (1.792)$ at 95% level of significance. On the other hand, CMORPH has $t_{\text{calculated}}$

$(-ve, 4.1545) > t_{\text{critical}} (1.792)$ at 95% level of significance. This implies that the correlation coefficient is negatively significant with elevation at 95% level of significance for both CMORPH and TRMM. R^2 value for TRMM is 0.61 and 0.49 for CMORPH. The higher value of R^2 for TRMM indicates a better linear fit and thus it can be more easily bias adjusted than CMORPH. Figure 9 compares the spatial

Figure 8. Elevation *vs.* correlation coefficient (CC) diagram between (a) daily Gauge *vs.* TRMM and (b) daily Gauge *vs.* CMORPH for all the rain gauge stations.

Figure 9. Graduated bubble map of correlation coefficient (CC) between daily Gauge data and (a) TRMM and (b) CMORPH over the DEM of study area for all six rain gauge stations.

Table 6. *Continuous variable verification method statistics of daily satellite precipitation products relative to gauge data.*

Station	Data	Continuous variable verification method				
		MAE	RMSE	CC	Mean error	Systematic error
Marsharakh	TRMM	8.16	17.79	0.72	0.82	1.1
	CMORPH	7.45	19.87	0.41	-5.99	0.25
Chatiya	TRMM	7.63	15.97	0.71	1.75	1.23
	CMORPH	6.66	16.06	0.58	-5.49	0.27
Bagaha	TRMM	10.84	22.59	0.75	2.63	1.28
	CMORPH	8.56	22.21	0.56	-6.48	0.3
Kuncha	TRMM	15.07	25.35	0.55	-3.14	0.79
	CMORPH	14.74	27.53	0.29	-12.2	0.19
Thamachit	TRMM	8.67	13.98	0.25	-6.87	3.38
	CMORPH	3.18	16.17	0.1	-0.84	0.71
Baghara	TRMM	16.21	24.35	0.5	-9.43	0.53
	CMORPH	18.43	27.11	0.33	-18.08	0.1

variation of the correlation coefficient of TRMM and CMORPH with elevation. It clearly shows a larger error in the mountains as indicated by lower CC. When we compare two CC maps one by one, it shows that TRMM and CMORPH have maximum CC, 0.75 and 0.58 and minimum 0.25 and 0.10 respectively. Thus, spatial correlation coefficient map shows a better prediction of rain amount by TRMM than CMORPH over the entire basin.

4. Conclusions

Management of extreme events (flood and drought) are critical to reducing vulnerability. Accurate rainfall is the key to predict the streamflow for extreme events using hydrological models. Availability of station based gauge rainfall data is limited due to the trans-boundary problem, complexity in climate and topography within the study area. Two satellite precipitation products, viz., TRMM 3B42 V7 and CMORPH, are validated with the available six rain gauge station data. The validation period is from 2005–2010 for monsoon period (JJAS). Based on the statistical discussion, we summarise here the following implications in the satellite precipitation data.

- The reliability of daily satellite precipitation products is dependent on the topography of the basin. Both the products were able to differentiate days with precipitation and days without precipitation but the gauge adjusted TRMM 3B42 V7 product is found more accurate than the exclusively satellite-based precipitation product CMORPH.
- TRMM daily average precipitation overestimates the rain amount for low rain intensities over the entire area, whereas it underestimates the rain amount for high rain intensities over the high altitude mountainous areas. CMORPH does well for low-intensity rain events but it underestimates the rain amount for high rain intensities over the entire area.
- TRMM captures the frequency of rain events for low-intensity while CMORPH overestimates the frequency of low intense rain events. On the other hand, TRMM overestimates the frequency of high-intensity rain events over the plains and underestimates over mountains.
- Satellite estimates generally overestimate the rain events. TRMM and CMORPH are found to be reliable to detect ‘rainy events’ with more than 80% accuracy for plains while it does detect ‘no-rain days’ in the mountainous areas.
- The correlation coefficient decreases with increase in altitude and it is found that it is negatively significant for *t*-test hypothesis testing method at 95% level of significance.
- The error strongly depends on the change in elevation, so it is recommended to incorporate it as a parameter in bias adjustment.

Based on above conclusions, we recommend the use of satellite products after the bias adjustment in high altitudes. The gauge adjusted TRMM 3B42 V7 product can be used for extreme event analysis such as flood frequency analysis and drought monitoring in the mountainous Himalayan basins.

Acknowledgements

Authors are very thankful to editor and anonymous reviewers for their insightful comments and valuable suggestions. Without their support this paper would not have been possible in the present shape.

References

- Arnold D N, Boffi D and Falk R S 2002 Approximation by quadrilateral finite element; *Math. Comput.* **71** 909–922.
- Bharti V and Singh C 2015 Evaluation of error in TRMM 3B42V7 precipitation estimates over the Himalayan region; *J. Geophys. Res.: Atmospheres*, doi: 10.1002/2015JD023779.
- Brito A E, Chan S H and Cabrera S D 2003 SAR image super resolution via 2-D adaptive extrapolation; *Multidim. Syst. Sign. Proces.* **14** 83–104.
- Brown B G, Thompson G, Brientjes R T, Bullock R and Kane T 1997 Intercomparison of in-flight icing algorithms: Part II: Statistical verification results; *Wea. Forecasting* **12** 890–914.
- Chen Y, Ebert E E, Kevin J E Walsh, Davidson and Noel E 2013 Evaluation of TMPA 3B42 daily precipitation estimates of tropical cyclone rainfall over Australia; *J. Geophys. Res.: Atmospheres* **118** 966–978.
- Din S U, Al-Dousari, Ramdan A and Al-Ghadban A 2008 Site-specific precipitation estimate from TRMM data using bilinear weighted interpolation technique: An example from Kuwait; *J. Arid Environ.* **72** 1320–1328.
- Duncan J and Biggs E M 2012 Assessing the accuracy and applied use of satellite-derived precipitation estimates over Nepal; *Appl. Geogr.* **34** 626–38.
- Duo C, Tundrop P, Ghancan N, Bajracharya S, Shrestha M and Jianping G 2011 Validation of the satellite-derived rainfall estimates over the Tibet; *Acta Meteor. Sinica* **25**(6) 734–741.
- Folland C, Frich P and Basnett T 2000 Uncertainties in climate datasets – A challenge for WMO; *WMO Bull.* **49**(1) 59–68.
- GFCC 2004 Updated comprehensive plan of flood management of Gandak river system; Ganga Flood Control Commission, Ministry of Water Resources, Gov. India.
- Guo J, Zhai P, Wu L, Cribb M, Li Z, Ma Z, Wang F, Chu D, Wang P and Zhang J 2014 Diurnal variation and the influential factors of precipitation from surface and satellite measurements in Tibet; *Int. J. Climatol.* **34** 2940–2956.
- Gribbon K T and Bailey D G 2004 A novel approach to real-time bi-linear interpolation; In: *Proc. Second IEEE Int. Workshop on Electronic Design, Test and Applications*, 126p.

- Huffman G J, Adler R, Bolvin D, Gu G, Nelkin E, Bowman K, Hong Y, Stocker E and Wolff D 2007 The TRMM Multisatellite Precipitation Analysis (TCMA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales; *J. Hydrometeorol.* **8** 38–55.
- Huffman G J, Adler R, Bolvin D and Nelkin E 2010 The TRMM Multi Satellite Precipitation Analysis (TMPA); *Satellite Rainfall Applications for Surface Hydrology*, Springer Netherlands, pp. 3–22.
- Islam M N, Das S and Uyeda H 2010 Calibration of TRMM derived rainfall over Nepal during 1998–2007; *Open Atmos. Sci. J.* **4** 12–23.
- Jiang S, Ren L, Hong Y, Yong B, Yang X, Yuan F and Ma M 2012 Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the bayesian model averaging method; *J. Hydrol.* **452–453** 213–225.
- Joyce R J, Janowiak J E, Arkin P A and Xie P 2004 CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution; *J. Hydrometeorol.* **5** 487–503.
- Jones P D, Horton E B and Folland C K 1999 The use of identify climate change extremes; In: *Weather and Climate Extremes*, Springer Netherlands, pp. 131–149.
- Kneis D, Chatterjee C and Singh R 2014 Evaluation of TRMM rainfall estimates over a large Indian river basin (Mahanadi); *Hydrol. Earth Syst. Sci.* **18** 2493–2502.
- Krakauer Y N, Pradhanang S M, Lakhankar T and Jha A K 2013 Evaluating satellite products for precipitation estimation in mountain regions: A case study for Nepal; *Remote Sens.* **5** 4107–4123.
- Kuhel H and Sacchi M D 2003 Least squares wave equation migration for AVP/AVA inversion; *Geophysics* **68** 262–273.
- Muller M F and Thompson S E 2013 Bias adjustment of satellite rainfall data through stochastic modeling: Methods development and application to Nepal; *Adv. Water Resour.* **60** 121–134.
- Murphy A H 1993 What is a good forecast? An essay on the nature of goodness in weather forecasting; *Wea. Forecasting* **8** 281–293.
- Nair S, Srinivasan G and Nemani R 2009 Evaluation of multi-satellite TRMM derived rainfall estimates over a western state of India; *J. Meteorol. Soc. Japan* **87** 927–939.
- Plengsaeng B, When U and Zaag P V D 2014 Data-sharing bottlenecks in trans-boundary integrated water resources management: A case study of the Mekong River Commission's procedures for data sharing in the Thai context; *Water Int.* **39** 933–951.
- Prakash S, Mitra A K and Pai D S 2015 Comparing two high-resolution gauge-adjusted multi-satellite rainfall products over India for the southwest monsoon period; *Meteorol. Appl.* **22** 679–688.
- Rahman M M, Arya D S, Goel N K and Mitra A K 2011 Rainfall statistics evaluation of ECMWF model and TRMM data over Bangladesh for flood related studies; *Meteorol. Appl.* **19** 501–512.
- Rahman S H and Simon B 2006 Summer monsoon intraperiodal oscillation over eastern Arabian Sea – as revealed by TRMM microwave imager products; *J. Earth Syst. Sci.* **115** 575–586.
- Ramesh S V, Teegavarapua and Chandramoulia V 2005 Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records; *J. Hydrol.* **312** 191–206.
- Shige S, Yamamoto T, Tsukiyama T, Kida S, Ashiwake H, Kubota T, Seto S, Aonashi K and Okamoto K 2009 The gsmap precipitation retrieval algorithm for microwave sounders – Part I: Over-ocean algorithm; *IEEE T; Geosci. Remote* **47** 3084–3097.
- Shrestha M S, Artan G A, Bajracharya S R, Gautam D K and Tokar S A 2011 Bias-adjusted satellite-based rainfall estimates for predicting floods: Narayani Basin; *J. Flood Risk Manag.* **4** 360–373.
- Stanski H R, Wilson L J and Burrows W R 1989 Survey of common verification methods in meteorology; World Weather Watch Technical Report No. 8, WMO/TD No. 358, WMO: Geneva; 114p.
- Sorooshian S, Hsu K, Gao X, Gupta H V, Imam B and Braithwaite D 2000 Evolution of the PERSIANN system satellite-based estimates of tropical rainfall; *Bull. Am. Meteorol. Soc.* **81** 2035–2046.
- TRMM Precipitation Radar Team 2011 Tropical rainfall measuring mission (TRMM) precipitation radar algorithm: Instrument manual for version 7. Tech. Rep. JAXA/NASA, 170p. http://www.eorc.jaxa.jp/TRMM/documents/PR_algorithm_product_information/_pr_manual/PR_Instruction_Manual_V7_L1.pdf.
- Vernimmen R R E, Hooijer A, Mamenun Aldrian E and Dijk A I J M van 2012 Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia; *Hydrol. Earth. Syst. Sci.* **16** 133–146.
- Viglione A, Borga M, Balabanis P and Blöschl G 2010 Barriers to the exchange of hydrometeorological data in Europe: Results from a survey and implications for data policy; *J. Hydrol.* **394** 63–77.
- Wang J and Wolff David B 2010 Evaluation of TRMM ground-validation radar-rain errors using rain gauge measurements; *J. Appl. Meteorol. Clim.* **49** 310–324.
- Wilks D S 2006 Statistical Methods in the Atmospheric Science; Academic Press, 627p.
- Wu L and Zhai P 2012 Validation of daily precipitation from two high-resolution satellite precipitation datasets over the Tibetan plateau and the regions to its east; *Acta Meteor. Sinica* **26** 735–745.
- Xue X, Hong Y, Limaye A S, Gourley J J, Huffman G J, Khan S I, Dorji C and Chen S 2013 Statistical and hydrological evaluation of TRMM-based Multi-satellite precipitation analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins?; *J. Hydrol.* **499** 91–99.

MS received 6 October 2015; revised 4 March 2016; accepted 10 March 2016