Gridded daily Indian monsoon rainfall for 14 seasons: Merged TRMM and IMD gauge analyzed values

ASHIS K MITRA^{1,*}, I M MOMIN¹, E N RAJAGOPAL¹, S BASU¹,
M N RAJEEVAN² and T N KRISHNAMURTI³

¹*National Centre for Medium Range Weather Forecasting (NCMRWF), Noida 201 309, India.*

²*MoES, Lodi Road, New Delhi 110 003, India.*

³*DEOAS, Florida State University, Tallahassee, FL 32306, USA.*

**Corresponding author. e-mail: ashis.mitra@nic.in ashis.kmitra@gmail.com*

Indian monsoon is an important component of earth's climate system. Daily rainfall data for longer period is vital to study components and processes related to Indian monsoon. Daily observed gridded rainfall data covering both land and adjoining oceanic regions are required for numerical model validation and model development for monsoon. In this study, a new gridded daily Indian rainfall dataset at $1^\circ \times 1^\circ$ latitude/longitude resolution covering 14 monsoon seasons (1998–2011) are described. This merged satellite gauge rainfall dataset (NMSG) combines TRMM TMPA rainfall estimates with gauge information from IMD gridded data. Compared to TRMM and GPCP daily rainfall data, the current NMSG daily data has more information due to inclusion of local gauge analysed values. In terms of bias and skill scores this dataset is superior to other daily rainfall datasets. In a mean climatological sense and also for anomalous monsoon seasons, this merged satellite gauge data brings out more detailed features of monsoon rainfall. The difference of NMSG and GPCP looks significant. This dataset will be useful to researchers for monsoon intraseasonal studies and monsoon model development research.

List of Acronyms

NMSG:	NCMRWF Merged Satellite Gauge product
TRMM:	Tropical Rainfall Measuring Mission
TMPA:	TRMM Multi-Satellite Precipitation Analysis
IMD:	India Meteorological Department
GPCP:	Global Precipitation Climatology Project
PR:	Precipitation radar
NCMRWF:	National Centre for Medium Range Weather Forecasting, India

1. Introduction

Observations of the earth system are important for understanding weather and climate. These observations are also pre-requisite to numerical models for preparation of model initial conditions through data assimilation. Observed rainfall data again is very crucial for model validation. Indian monsoon is an important component of earth's climate system. Daily rainfall data for longer period is vital to study components and processes related to Indian monsoon. Daily rainfall associated with transients like low-pressure systems and monsoon depressions contribute to the total seasonal

Keywords. Observed monsoon rain; satellite rain; daily merged rain data; TRMM rainfall; monsoon rainfall variability.

rainfall. For tropical and especially monsoon region, where deep convection plays crucial role, the validation of numerical models require accurate rainfall information over land and adjoining seas. In recent time, availability of computers, satellite and *in situ* sensors have revolutionized the modelling and observation of rainfall. New observing systems from satellites and air-borne platforms are providing enhanced measurements. High-resolution precipitation data products have become available from a variety of satellites and from their combination (Joseph *et al.* 2009; Kidd and Huffman 2011; Kidd *et al.* 2012). A variety of climatology of rainfall have been prepared based on these satellite rainfall data for different regions of the world. GridSat (Knapp *et al.* 2011) is an example of accessible satellite data from multiple sources. TRMM is one such state-of-the-art satellite system, which has opened up exciting opportunities to learn more about weather and climate systems of tropics and monsoons. Since 1997, TRMM has provided rainfall data for research and application activities related to hydro-meteorology. TRMM archived and real-time data provided by NASA GES DISC have been used widely around the world (Liu *et al.* 2012). Out of a variety of data available, TRMM gridded daily data (Huffman *et al.* 2007, 2010) provides comprehensive information on tropical daily rainfall. A regional climatology of monsoon precipitation in the southwest USA has been prepared from TRMM data (Wall *et al.* 2012). These datasets are very useful for planned studies like North American Monsoon Experiment (NAME). The TRMM PR data has revealed on how convective systems of different types and their associated horizontal scales contribute to the precipitation of south Asia monsoon (Romatschke and Houze 2011).

These satellite rainfall data are useful, still it requires calibration at regional levels with ground-based gauge and radar observations. In a recent study for a sample tropical region of USA, it was noted that at shorter time-scales and for higher rain rates, the TRMM rainfall have biases compared to ground measurements (Wang and Wolff 2012). This bias fluctuates from year to year. Satellite rainfall estimates show regional and seasonal variations in biases also (Lu *et al.* 2010). Region specific biases have to be corrected by inclusion of said conventional observations (Mitra *et al.* 2003, 2009). Merging gauge-based information to satellite values is a viable option for getting more accurate rainfall information. Daily TRMM rainfall captures the sub-seasonal monsoon rainfall variability (Rahman *et al.* 2009), however, there is a need to remove the biases on daily basis by merging information from other ground-based measurements.

Study of features related to intraseasonal oscillations of tropics and monsoons are very important to understand, simulate and predict the dynamical and physical processes related to tropical south Asian monsoons (Kripalani *et al.* 2004; Goswami 2005; Xavier 2012). State-of-the-art models have limitations and drawbacks in representing monsoon/tropical rainfall and their variabilities. Model biases and deficiencies related to various processes have to be studied for continuous model development. Representation of monsoon intraseasonal variability in latest coupled models *via* large-scale organization of convection has to be documented. These studies need daily gridded data of rainfall covering land and oceanic region. To verify the skill of tropical rainfall and their variability from numerical models, we need daily observed gridded rainfall data (Chakraborty 2010). In turn, reliable model or reanalysis data could be used to study the characteristics of water balance, hydrological cycle and its space-time variability (Prasanna and Yasunari 2009). In this study, we describe 14 monsoon seasons daily gridded rainfall dataset for Indian region, prepared by merging daily TRMM TMPA satellite estimates with IMD daily gauge analysed information. In an earlier study, the algorithm used here was tested rigorously in detail for one season (Mitra *et al.* 2009). Compared to other similar daily data available, this data has more information from land-based gauge analysis, and will be useful for Indian monsoon studies.

2. Data and methodology

A successive correction based algorithm was used earlier to merge satellite rainfall with gauge data to obtain a daily gridded large-scale rainfall for Indian monsoon region (Mitra *et al.* 2003, 2009). In a recent study (Rozante *et al.* 2010), merged TRMM and gauge data have been produced for South American regions using successive correction algorithm. This successive correction algorithm is superior to just replacing box averaged values and has been found satisfactory to produce large-scale daily monsoon rainfall. The TRMM TMPA data is good in capturing the pattern and phase of intraseasonal variability of Indian monsoon (Rahman *et al.* 2009). The biases of the TRMM data can be corrected by merging gauge-based information. In our present dataset, the satellite product used as a first guess is the TRMM TMPA for daily estimates (Huffman *et al.* 2007, 2010). This original TRMM TMPA data was 3-hourly at 0.25° lat./long. grids, which have been accumulated over time and averaged over space to represent daily (24 hours accumulated) 1° lat./long. grids to be

used in our dataset. The other ground-based rain data used here is from IMD. IMD prepared daily gridded rainfall datasets at 0.5° lat./long. resolution (Rajeevan and Bhat 2009). The higher resolution IMD analysis at 0.5° lat./long. grids used around 3000 quality checked stations. This gridded data was used as proxy rain-gauge data in our successive correction method based on interpolation algorithm. Each grid point of IMD analysis was considered as a proxy gauge station, which actually represents value obtained from several rain-gauges surrounding the grid of IMD analysis. In IMD analysis, Indian land area on a 0.5° lat./long. grid consists of around 1250 grid points. Daily IMD analysed rainfall values at these 1250 grids are considered as proxy gauge data for our 1° lat./long. resolution merged satellite gauge rainfall analysis for the Indian monsoon region. Since the basic

gauge analysed data is at 0.5° lat./long. resolution, we decided to make our gridded data at 1° lat./long. resolution, so that the argument of using IMD input data as proxy gauges is justified. For our coarser resolution 1×1 analysis grid, the data from higher resolution 0.5×0.5 grids enables us with sufficient observations (1250 proxy data points). IMD gauge its analysis checks for the quality of *in situ* measured gauge data values, which we readily use (the analysed grid point values) as one input to our analysis. Therefore, it may be noted that, use of these 1250 point data based on ground measurements is good enough to represent the large-scale monsoon rainfall at $1^\circ \times 1^\circ$ lat./long. grids. This new data product described here is named as NCMRWF merged satellite gauge (NMSG) product. As the TRMM TMPA data covers the period starting from 1998 to current, in this study we

Figure 1. (a) Mean monsoon rainfall (JJAS) in mm/day for NMSG, GPCP and their difference. (b) Same as (a), except for JJA of 2009.

have prepared daily data of 14 monsoon seasons covering monsoon months of June to September (1998–2011).

Another popular dataset used in this study for inter-comparison purpose is the version 1.1 of the GPCP one degree daily 1DD products from the GPCP project (Huffman *et al.* 2001). This daily GPCP data is used by many research groups working on monsoon intraseasonal oscillations from observations and model. The GPCP 1DD product has been scaled to the new GPCP V2.1 monthly product, released during July 2009. The major

enhancement included in the GPCP V2.1 data is the inclusion of the new GPCC full and monitoring product gauge analyses, which provide better estimates of land surface precipitation. This newer version was reported to be superior to its earlier GPCP 2.0 version. In our current study, we use the GPCP 1DD V1.1 daily product, which is compatible to monthly GPCP Satellite-Gauge V2.1. Hence, this 1DD V1.1 is the latest available daily data from GPCP for 1997–2009 period. In the following discussion section, our NMSG daily data is compared with the GPCP 1DD V1.1 daily data.

Figure 2. (a) Mean time-latitude section of mean monsoon rainfall from NMSG, GPCP and their difference (NMSG–GPCP) averaged over longitudes 70°–85°E. (b) Same as (a), except for JJA 2009.

Figure 2. (Continued.)

3. Discussions

As mentioned earlier, GPCP daily gridded data covering both land and ocean regions were available from GPCP project for the period October 1996 to August 2009, and it is being used extensively by most monsoon researchers engaged in studies related to subseasonal variability. The NMSG data is available from 1998 to 2011. For intercomparison of daily NMSG and GPCP, we use the monsoon daily data during 1998 to 2009. June to September are raining monsoon period in India. Monsoon season of 2009 was an anomalous drought year. Here, mainly the mean monsoon

rain and the anomalous drought year of 2009 from NMSG and GPCP are compared for the Indian region. Figure 1(a) shows mean monsoon rain from GPCP, NMSG and the difference of NMSG minus the GPCP. Clearly, the impact of regional gauge analysis information is seen over land. In a mean sense, the mountain regions of the west coast of India (Western Ghats) and the Himalayan foot hills show higher rainfall in the NMSG data. For the anomalous 2009 dry year also, we see similar differences between GPCP and NMSG (figure 1b). Rain gauge can capture the warm rainfall over Western Ghats, whereas remote sensing is not sensitive to warm rainfall. Additionally during 2009, over

central India associated monsoon trough region, the NMSG data shows lower values compared to GPCP.

Figures 2 and 3, show latitude-time cross-section of rainfall averaged over longitudes of India region from NMSG, GPCP and their differences. In figure 2(a), latitude-time cross-section of rainfall for longitudes averaged over 70° to 85° E is shown from the climatology of both the datasets. In the mean sense, we find different northward propagating rainfall patterns during 122 days of monsoon. The difference plot of NMSG and GPCP on the bottom panel shows the impact of inclusion of the gauge information. The heavy rainfall over west

coast of India contributes to the difference. Similar differences with much higher magnitudes of both positive and negative values are seen for the anomalous year 2009 (figure 2b). These differences are quite significant and have to be considered for correct interpretation of monsoon rainfall intra-seasonal variability from model and observations. Figure 3(a) and (b) are similar to the latitude-time cross section of rainfall within the season, but averaged over longitudes 75° to 90° E. This region represents the monsoon trough and the associated Bay of Bengal domain, a representative region of monsoon activity. For anomalous 2009 monsoon season we find significant differences between NMSG

Figure 3. (a) Mean time-latitude section of mean monsoon rainfall from NMSG, GPCP and their difference (NMSG–GPCP) averaged over longitudes 75° – 90° E. (b) Same as (a), except for JJA 2009.

Figure 3. (Continued.)

and GPCP (figure 3b). These differences are of the order of 10 to 20 mm/day, with both positive and negative signs.

Apart from TRMM 3B42V6 (TMPA) there are two more TRMM products, 3A25PR and 3A12TMI available for different applications. In our NMSG daily merged data, we have used 3B42V6 as the first guess. To examine the suitability of first guess, we examined the biases of different TRMM products compared to an independent IMD daily gridded data at $1^\circ \times 1^\circ$ lat./long. resolution (Rajeevan *et al.* 2006). For seeing the bias and categorical skill scores, monsoon seasons of 2002 and 2003 were selected. In terms of rainfall, monsoon 2002 was an anomalous drought year, and 2003 was a normal year. Figure 4 shows the biases of the

different TRMM rainfall products with respect to the said verifying IMD data. Four diagrams in the upper panel are for the drought year 2002, and the lower panel is for the normal 2003 monsoon season. Diagrams from left to right are the biases (product minus IMD) of TRMM products like 3A25PR, 3A12TMI, 3B42V6 and our merged data respectively. Shades of red colour indicate dry bias, and that of blue indicated wet bias. It is clearly seen that 3A25PR and 3A12TMI products have higher biases over different parts of India. Rainfall associated with west coast orography, central India monsoon trough region, Himalayan mountain region and the north-east India region have higher biases. However, 3B42V6 is seen to have lesser biases compared to the other two TRMM

Figure 4. Bias of TRMM rain against observation for 2002 and 2003 monsoon period.

Figure 5. Skill scores of GPCP and NMSG rainfall against observation over central India region.

products. It is appropriate that the 3B42V6 product is used as the first guess in our NMSG merged data product. It is quite interesting to see that the bias of the merged product is least for both the 2002 and 2003 monsoon seasons for Indian region.

In figure 5, the categorical skill scores like equitable threat score (ETS) and the hit rate (HR) are shown for 2002 and 2003 monsoon seasons for GPCP and NMSG daily rainfall data. Objective scores are computed to compare the GPCP and NMSG with the IMD products to provide quantitative information about the goodness of the merged product from a statistical point of view. Here, we assume that the IMD rainfall data is closest to the true observed state of the atmosphere. Standard statistical parameters like ETS and HR are computed for the comparisons. A brief description of these categorical statistics is given by Ebert *et al.* (2007). HR (success rate) is the ratio of the number of correctly forecast points above a threshold to that of the number of forecast points above the corresponding threshold. ETS is commonly used as an overall efficiency measure for intercomparison of precipitation products. ETS gives the fraction of observed and/or detected rain that was correctly detected and is adjusted for the number of hits that could be expected purely due to random chance. Different thresholds from 1 to 6 cm/day are selected to examine the skills. A geographical domain covering the monsoon trough region of the central India is selected (box covering region 72.5°–83.5°E; 22.5°–26.5°N) for examining the skill of the GPCP and merged NMSG datasets. This region contains data from all the three datasets involved in computing the skill scores. Clearly for 2002 drought year, no samples are seen in higher rainfall category of 4 cm and above. The ETS and HR scores are higher for NMSG compared to GPCP. Similarly for 2003 normal monsoon year, for all rainfall threshold values of 1 to 6 cm, both the scores are much superior (higher scores) for the NMSG product compared to GPCP data. In an earlier study, we had shown the positive impact of including more local gauges to TRMM TMPA satellite estimates (Mitra *et al.* 2009). Objective scores then showed the NMSG data for 2006 monsoon to be more realistic and correct the biases of TRMM TMPA rainfall. Therefore, as more local gauge information has gone to the NMSG dataset, it is obvious that the daily NMSG is more realistic compared to the daily GPCP data. We have seen separately, that the NMSG agrees quite well to IMD land only analysis over land, confirming the current analysis algorithm retaining valuable information from gauges. This new 14 monsoon seasons dataset is an extension of the earlier work of merging TRMM TMPA with rain-gauges to produce a complete daily land-ocean dataset. Recently, this

daily NMSG dataset has been released to monsoon researchers for the April to October period. It will be useful in studies relating to comparisons of monsoon intraseasonal variability from observations and models.

4. Summary and future works

In this study, a new daily Indian rainfall dataset covering 14 monsoon seasons are described. This merged satellite gauge dataset combines TRMM TMPA rainfall with IMD gauge information. Compared to GPCP daily rainfall data, the current NMSG daily data has more information due to inclusion of local gauge values. In terms of bias and skill scores this dataset is superior to other daily datasets. This dataset will be useful for monsoon intraseasonal studies and monsoon model development research. For depicting large-scale daily rainfall pattern, this merged dataset may be reasonable. Over India, rain-gauge data is the only information which is used both in preparing this data, and also for verification. It may be appropriate in a future study to verify this dataset with other independent *in situ* data like rainfall measurements from Radar or AWS/ARG network. For representing much finer scale diverse rainfall patterns and their variability, current satellite-derived rainfall data may not be adequate (Nair *et al.* 2009), and we still need much more ground-based observations. To capture the intensity of rainfall, gauge, AWS and radar data will be needed. Evaluation, validation, assessment and intercomparison of satellite rainfall data and their calibration at various space-time scales will be a continuous research area of work for coming few years (Turk *et al.* 2008). In future, this dataset will be extended to more number of years by using realistic satellite rainfall estimates and ground-based information. For monsoon intraseasonal studies, this daily merged data is available from April to October period, but we have discussed only the monsoon (June to September) rainfall in this paper. We plan to include more gauge information from neighbouring countries to make it a south-Asian daily monsoon rainfall dataset for more number of years. Experience gained to include TRMM data product and their utilization will be helpful in upcoming GPM mission scheduled for launch in 2014.

Acknowledgements

Authors are thankful to Dr Shailesh Nayak, Secretary MoES for his encouragement and guidance. Gauge analyzed rainfall data was available from IMD. TRMM TMPA rainfall data was made

available *via* ftp from NASA TRMM project. Daily GPCP 1DD data was available from GPCP project *via* ftp. April to October period daily rainfall data described here are downloaded from <ftp.ncmrwf.gov.in> located at </home/ftp/pub/outgoing/akmitra> by anonymous ftp. This work is carried out under SAPRISE CWC project.

References

- Chakraborty A 2010 The skill of ECMWF MRF during the YOTC 2008; *Mon. Weather Rev.* **138**(10) 3787–3805.
- Ebert E E, Janowiak J and Kidd C 2007 Comparison of near real time precipitation estimates from satellite observations and numerical models; *Bull. Am. Meteor. Soc.* **88** 47–64.
- Goswami B N 2005 South Asian Monsoon; In: *Intra-seasonal Variability in the Atmosphere-Ocean Climate System* (eds) Lau W K M and Waliser D E, Chap. 2, pp. 19–61.
- Huffman G J, Adler R F, Morrissey M M, Curtis S, Joyce R, McGavock B and Susskind J 2001 Global precipitation at one-degree daily resolution from multi-satellite observations; *J. Hydrometeorol.* **2** 36–50.
- Huffman G J et al. 2007 The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales; *J. Hydrometeorol.* **8**(1) 38–55.
- Huffman G J, Adler R F, Bolvin D T and Nelkin E J 2010 The TRMM multi-satellite precipitation analysis (TMPA); In: *Satellite Rainfall Application for Surface Hydrology* (eds) Gebremichael M and Hossain F, Springer-Verlag, ISBN 978-90-481-2914-0, pp. 3–22.
- Joseph R, Smith T M, Sapiano M R P and Ferraro R R 2009 A new high-resolution satellite-derived precipitation dataset for climate studies; *J. Hydrometeorol.* **10** 935–952.
- Kidd C and Huffman G 2011 Global precipitation measurement (review); *Meteorol. Appl.* **18**(3) 334–353.
- Kidd C P, Bauer, Turk J, Huffman G J, Royce R, Hsu K L and Braithwaite D 2012 Intercomparison of high-resolution precipitation products over north-west Europe; *J. Hydrometeorol.* **13**(1) 67–83.
- Knapp K R et al. 2011 Globally gridded satellite observations for climate studies; *Bull. Am. Meteor. Soc.* **92**(7) 893–907.
- Kripalani R H, Kulkarni A, Sabade S S, Revadekar J V, Patwardhan S K and Kulkarni J R 2004 Intra-seasonal oscillations during monsoon 2002 and 2003; *Curr. Sci.* **87**(3) 325–331.
- Liu Z, Ostrenga D, Teng W and Kempler S 2012 TRMM precipitation data and services for research and applications; *Bull. Am. Meteor. Soc.* **93** 1317–1325, doi: [10.1175/BAMS-D-11-00152.1](https://doi.org/10.1175/BAMS-D-11-00152.1).
- Lu C, Yuan H, Tollerud E I and Wang N 2010 Scale-dependent uncertainties in global QPFs and QPEs from NWP model and satellite fields; *J. Hydrometeorol.* **11** 139–155.
- Mitra A K, Dasgupta M, Singh S V and Krishnamurti T N 2003 Daily rainfall for Indian monsoon region from merged satellite and raingauge values: Large-scale analysis from real-time data; *J. Hydrometeorol.* **4**(5) 769–781.
- Mitra A K, Bohra A K, Rajeevan M N and Krishnamurti T N 2009 Daily Indian precipitation analyses formed from a merge of rain-gauge with TRMM TMPA satellite derived rainfall estimates; *J. Meteor. Soc. Japan* **87A** 265–279.
- Nair S, Srinivasan G and Nemani R 2009 Evaluation of multi-satellite TRMM rainfall estimates over a western state of India; *J. Meteor. Soc. Japan* **87**(6) 927–939.
- Prasanna V and Yasunari T 2009 Time-space characteristics of seasonal and interannual variations of atmospheric water balance over south Asia; *J. Meteor. Soc. Japan* **87**(2) 263–287.
- Rahman S H, Sengupta D and Ravichandran M 2009 Variability of Indian summer monsoon rainfall in daily data from gauge and satellite; *J. Geophys. Res.* **114** D17113.
- Rajeevan M N and Bhate J 2009 A high-resolution daily gridded rainfall dataset (1971–2005) for meso-scale meteorological studies; *Curr. Sci.* **96**(4) 558–562.
- Rajeevan M, Bhate J, Kale J D and Lal B 2006 High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells; *Curr. Sci.* **91**(3) 296–306.
- Romatschke U and Houze R A Jr 2011 Characteristics of precipitating convective systems in south Asia monsoon; *J. Hydrometeorol.* **12**(2) 157–180.
- Rozante J R, Moreira D S, Goncalves L G and Vila D A 2010 Combining TRMM and surface observations of precipitation: Technique and validation over South America; *Wea. Forecasting* **25**(6) 885–894.
- Turk F J, Arkin P, Ebert E E and Sapiano M R P 2008 Evaluating high-resolution precipitation products: Meeting summary; *Bull. Am. Meteor. Soc.* **89**(12) 1911–1916.
- Wall C L, Zipser E J and Liu C 2012 A regional climatology of monsoon precipitation in the south-western US using TRMM; *J. Hydrometeorol.* **13**(1) 310–323.
- Wang J and Wolff D B 2012 Evaluation of TRMM rain estimates using ground measurements over central Florida; *J. Appl. Met. Clim.* **51**(5) 926–940.
- Xavier P 2012 Intraseasonal convective moistening in CMIP3 models; *J. Climate* **25**(8) 2569–2577.