

# *Seasonal intercomparison of observational rainfall datasets over India during the southwest monsoon season*

Article

Published Version

Prakash, S., Mitra, A. K., Momin, I. M. , Rajagopal, E. N., Basu, S., Collins, M., Turner, A. G., Achuto Rao, K. and Ashok, K. (2015) Seasonal intercomparison of observational rainfall datasets over India during the southwest monsoon season. *International Journal of Climatology*, 35 (9). pp. 2326-2338. ISSN 0899-8418 doi: <https://doi.org/10.1002/joc.4129>  
Available at <http://centaur.reading.ac.uk/37234/>

It is advisable to refer to the publisher's version if you intend to cite from the work.

To link to this article DOI: <http://dx.doi.org/10.1002/joc.4129>

Publisher: John Wiley & Sons

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Seasonal intercomparison of observational rainfall datasets over India during the southwest monsoon season

Satya Prakash,<sup>a</sup> Ashis K. Mitra,<sup>a\*</sup> Imran M. Momin,<sup>a</sup> E. N. Rajagopal,<sup>a</sup> S. Basu,<sup>a</sup> Mat Collins,<sup>b</sup> Andrew G. Turner,<sup>c</sup> K. Achuta Rao<sup>d</sup> and K. Ashok<sup>e</sup>

<sup>a</sup> National Centre for Medium Range Weather Forecasting, Earth System Science Organization, Ministry of Earth Sciences, Noida, India

<sup>b</sup> University of Exeter, UK

<sup>c</sup> NCAS-Climate & Department of Meteorology, University of Reading, UK

<sup>d</sup> Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, India

<sup>e</sup> Indian Institute of Tropical Meteorology, Ministry of Earth Sciences, Pune, India

**ABSTRACT:** The Indian monsoon is an important component of Earth's climate system, accurate forecasting of its mean rainfall being essential for regional food and water security. Accurate measurement of rainfall is essential for various water-related applications, the evaluation of numerical models and detection and attribution of trends, but a variety of different gridded rainfall datasets are available for these purposes. In this study, six gridded rainfall datasets are compared against the India Meteorological Department (IMD) gridded rainfall dataset, chosen as the most representative of the observed system due to its high gauge density. The datasets comprise those based solely on rain gauge observations and those merging rain gauge data with satellite-derived products. Various skill metrics and subjective comparisons are carried out for the Indian region during the southwest monsoon season (June–September). Relative biases and skill metrics are documented at all-India and subregional scales. In the gauge-based (land-only) category, Asian Precipitation – Highly-Resolved Observational Data Integration Towards Evaluation of water resources (APHRODITE) and Global Precipitation Climatology Center (GPCC) datasets perform better relative to the others in terms of a variety of skill metrics. In the merged category, the Global Precipitation Climatology Project (GPCP) dataset is shown to perform better than the Climate Prediction Center Merged Analysis of Precipitation (CMAP) for the Indian monsoon in terms of various metrics, when compared with the IMD gridded data. Most of the datasets have difficulties in representing rainfall over orographic regions including the Western Ghats mountains, in Northeast India and the Himalayan foothills. The wide range of skill metrics seen among the datasets and even the change of sign of bias found in some years are causes of concern. This uncertainty between datasets is largest in Northeast India. These results will help those studying the Indian monsoon region to select an appropriate dataset depending on their application and focus of research.

**KEY WORDS** South Asian monsoon; Indian summer monsoon rainfall; skill metrics; orography

Received 31 December 2013; Revised 15 July 2014; Accepted 15 July 2014

## 1. Introduction

The South Asian monsoon is one of the major components of Earth's climate system contributing towards variability of the global water cycle (Trenberth *et al.*, 2000). Accurate estimation and prediction of the South Asian summer monsoon rainfall are crucial for many societal applications. In the last decade, with the advancement of computational capabilities, data assimilation and better model parameterization schemes, the skill of numerical models of weather and climate has improved considerably (Slingo and Palmer, 2011; Turner *et al.*, 2011; Hoskins, 2013; Mitra *et al.*, 2013a). However, a commensurate skill improvement has not been fully realized for the tropical monsoons. Simulation and prediction of the South Asian monsoon by numerical models are challenging scientific

issues for the global modelling community (Sperber *et al.*, 2013). Owing to large biases in models for the monsoon region, realistic future climate change projection of the South Asian monsoon rainfall is still uncertain (Turner and Annamalai, 2012). Evaluation of model rainfall provides vital feedback for further model development. Evaluation of model outputs against observations is also crucial for a wide range of applications (Ebert *et al.*, 2007; Sorooshian *et al.*, 2011; Collins *et al.*, 2013).

Indian summer monsoon rainfall is at the heart of the South Asian monsoon system. Reliable *in situ* rainfall observations are available over India for a long period, in many areas dating back to the 19th century. Indian rainfall associated with the southwest monsoon is unique in terms of its spatial and temporal variability. This complex variability is associated with varied land types, mountains and monsoon flow patterns. Availability of rainfall data has enabled preparation of gridded rainfall datasets at various spatial resolutions and temporal scales by

\* Correspondence to: A. K. Mitra, NCMRWF, ESSO, MoES, A-50, Sector-62, Noida 201309, India. E-mail: ashis.mitra@nic.in

Table 1. Summary of the gridded rainfall datasets used.

Data	Full name	Spatial/temporal resolution	Input data sources	Key reference
IMD	India Meteorological Department gridded rainfall data	0.5°/daily	Rain gauges	Rajeevan and Bhate (2009)
GPCP	Global Precipitation Climatology Project Version 2.2	2.5°/monthly	IR and MW satellite observations, rain gauges	Adler <i>et al.</i> (2003)
CMAP	Climate Prediction Center (CPC) Merged Analysis of Precipitation Version 1201	2.5°/monthly	IR and MW satellite observations, rain gauges	Xie and Arkin (1996)
CPC	Climate Prediction Center Unified Rain gauge data Version 1.0	0.25°/daily	Rain gauges	Xie <i>et al.</i> (2007)
APHRO	Asian Precipitation – Highly-Resolved Observational Data Integration Towards Evaluation of water resources Version 1101R2 Monsoon Asia	0.5°/daily	Rain gauges	Yatagai <i>et al.</i> (2012)
CRU	Climate Research Unit Version 3.10.01	0.5°/monthly	Rain gauges	Harris <i>et al.</i> (2014)
GPCC	Global Precipitation Climatology Center Full Data Reanalysis Version 6	1°/monthly	Rain gauges, SYNOP and CLIMAT reports	Schneider <i>et al.</i> (2014)

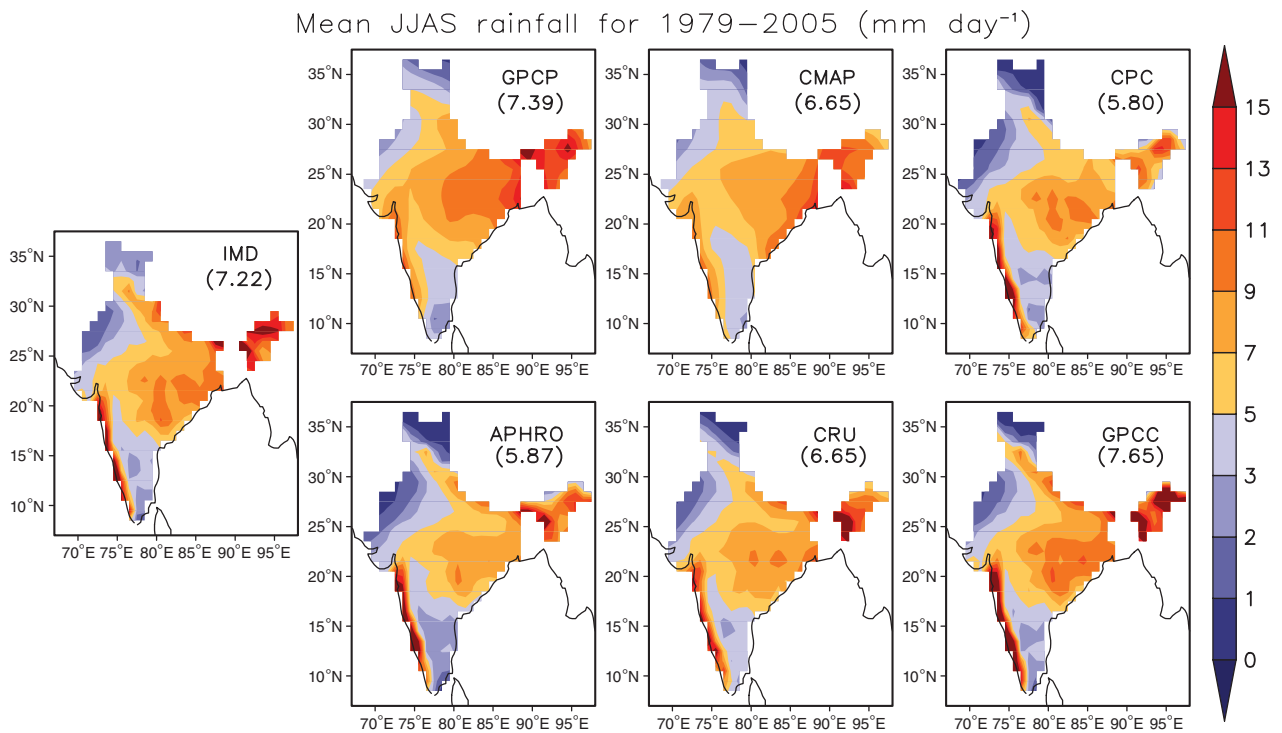


Figure 1. Spatial distributions of mean seasonal rainfall during the JJAS southwest monsoon averaged for the period 1979–2005. All-India mean values are given in the parentheses.

different research groups worldwide. As a result, a variety of global or regionally restricted gridded rainfall data are now easily available for research studies. Ground-based observations such as rain gauges and radars essentially provide accurate measurement of rainfall at the specific location. However, their uneven distributions over unpopulated regions and sampling issues, e.g. point observations from rain gauges, limit their wide applicability. On

the other hand, merged rainfall products provide spatially complete distributions of rainfall covering both land and oceans. Some of the gridded rainfall products are available over land only (from rain gauges) and some others feature a blend of gauges with satellite-based estimates providing gridded rainfall data over both land and oceans, popularly known as merged products (Kidd and Levizzani, 2011; Sorooshian *et al.*, 2011; Kucera *et al.*, 2013). However,

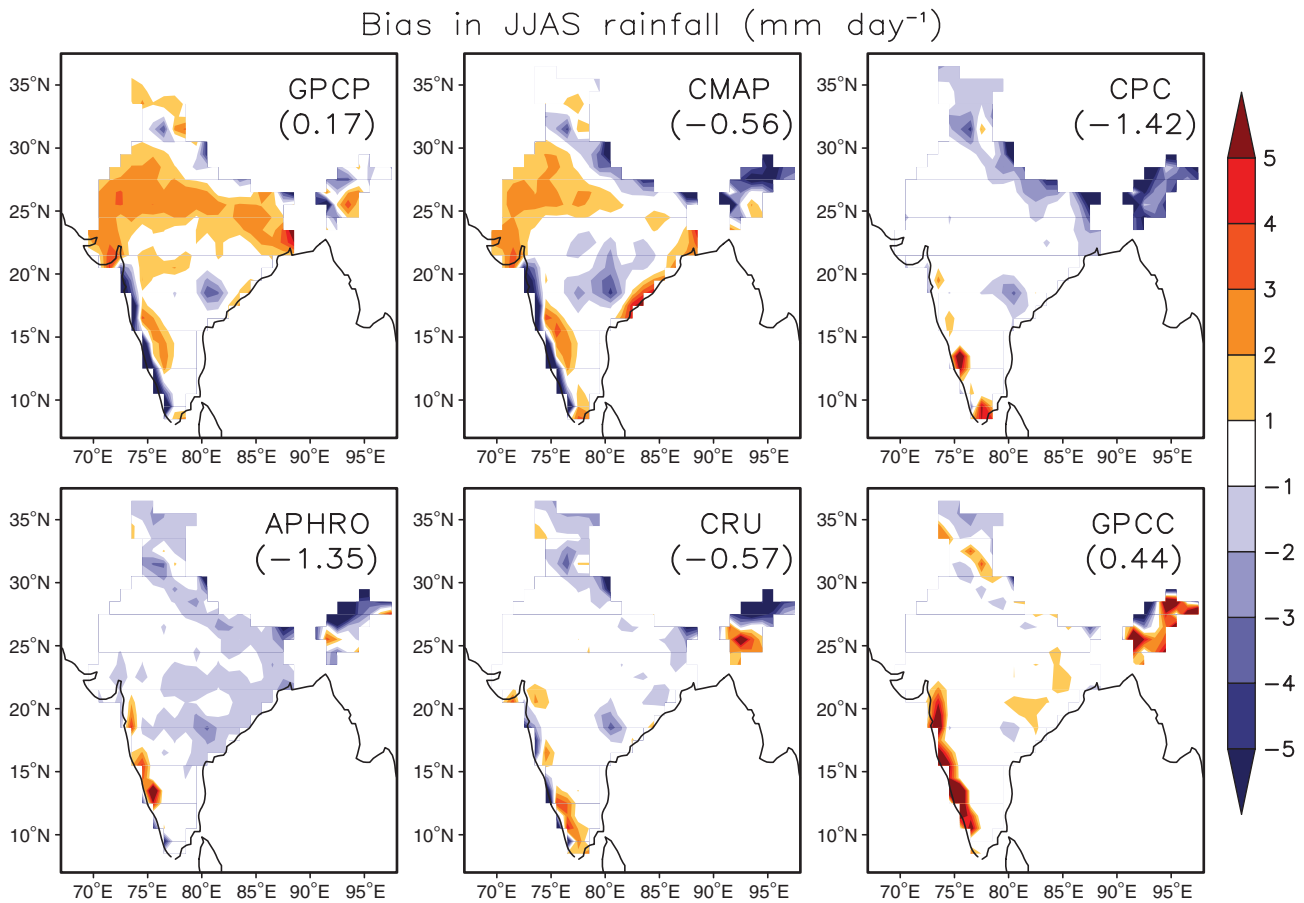


Figure 2. Spatial distributions of mean bias with respect to IMD rain gauge data during the JJAS southwest monsoon averaged for the period 1979–2005. All-India mean values are given in the parentheses.

the choice of the best rainfall data among these available datasets is again very important for specific applications, model validation and model development.

In this study, we perform a large-scale comparison of two merged and four rain gauge-based (land-only) rainfall products over India with the India Meteorological Department (IMD) rain gauge-derived gridded rainfall data for the southwest monsoon season (June–September). The comparison is done for seasonal mean measures over a 27-year period starting from the onset of the satellite era in 1979–2005. This also coincides with the Atmospheric Model Intercomparison Project (AMIP) period commonly used in general circulation model (GCM) studies. Except for the IMD data, all six datasets are available over the entire South Asian monsoon region. The intercomparison is aimed at finding out the relative skill of each dataset over the Indian monsoon region when compared with IMD gridded observations. The results will be useful for the monsoon rainfall data users for various applications and model development research.

## 2. Data and methods

We used seven rainfall datasets in this study for a 27-year period spanning 1979–2005. The summary of different rainfall products, including their names and versions used,

native spatial resolutions, the major sources of input data type that comprise each product with respective references are given in Table 1. The updated IMD gridded rainfall at  $0.5^\circ$  latitude/longitude is used as reference data to validate all other rainfall datasets described below (Rajeevan and Bhate, 2009). The IMD gridded rainfall data use a dense network of rain gauges over India (more than 3000) and well represents known large-scale monsoon features. Compared to other datasets used here, IMD has the maximum number of gauges used in its gridded product since the other datasets obtain gauge information from subsets of the IMD network.

### 2.1. Merged products

The latest versions of two merged gridded rainfall products namely, the Global Precipitation Climatology Project (GPCP; Adler *et al.*, 2003) version 2.2 and Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP; Xie and Arkin, 1996) version 1201 available since 1979 are used in this study. These two global rainfall products combine multisatellite and rain gauge observations and have similar temporal and spatial coverage. Both products are produced by different merging techniques and use partly different types of input data. The GPCP dataset includes inputs from the Special Sensor Microwave/Imager (SSM/I) emission and scattering

## Correlation coefficient

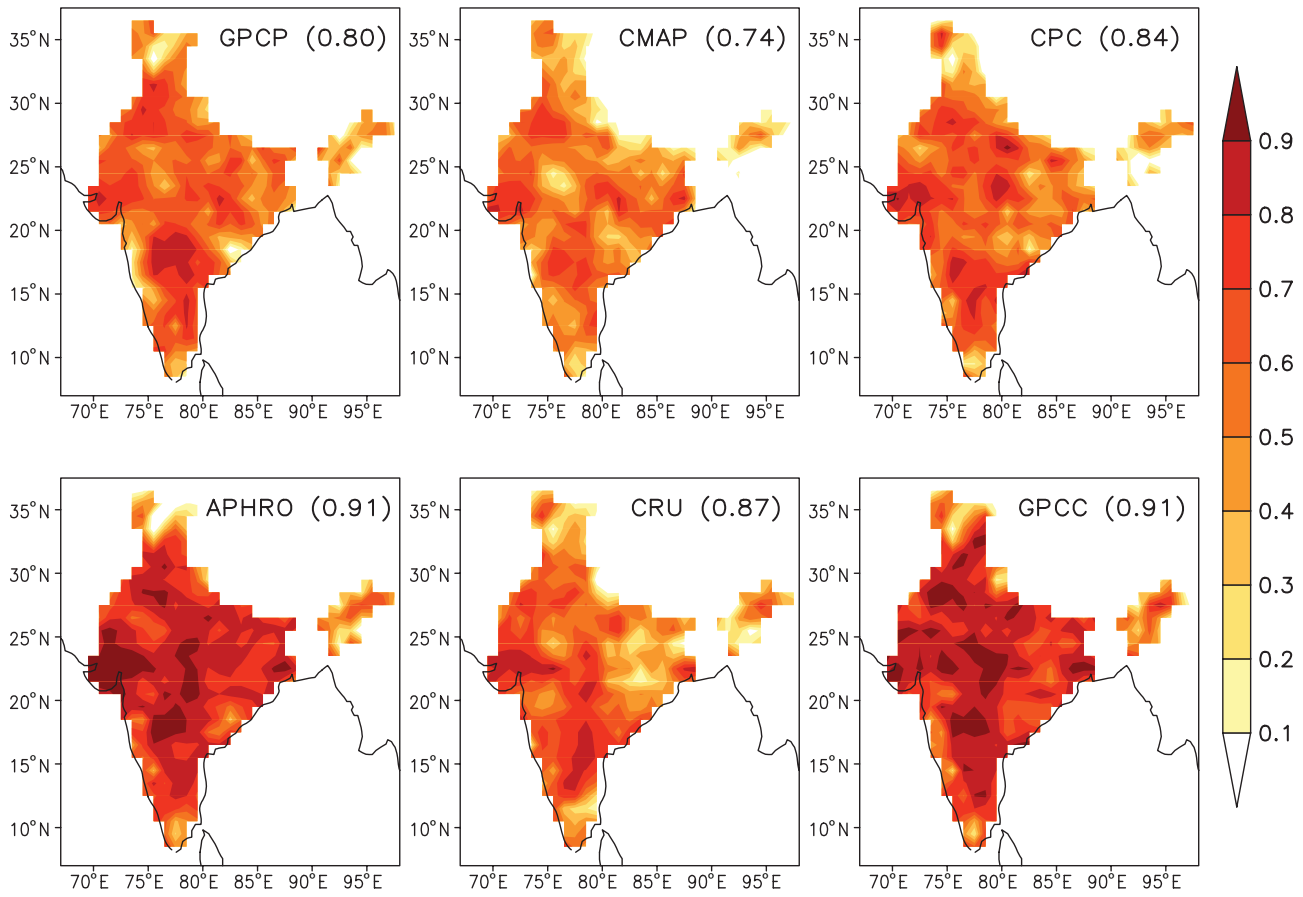


Figure 3. Spatial distributions of temporal correlation coefficient of each data set with the IMD data during the JJAS southwest monsoon averaged for the period 1979–2005. Pattern correlation values are given in the parentheses.

data, Geostationary Operational Environmental Satellites (GOES) Precipitation Index (GPI), Outgoing Longwave Radiation (OLR) Precipitation Index (OPI) data, Television and Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS), Atmospheric Infrared Sounder (AIRS) data and rain gauge observations. CMAP includes only the SSM/I data, infrared-based GPI and OPI data and Microwave Sounding Units (MSU) data along with gauge information.

## 2.2. Gauge-only products

Four gridded rain gauge-based land-only rainfall products are used for this evaluation. Firstly, CPC Unified daily rain gauge data (Xie *et al.*, 2007) version 1.0 available since 1979 is produced to create a suite of unified precipitation products by combining all information sources available at CPC, NOAA, USA. This rainfall product uses around 30 000 global rain gauge observations. Secondly, the Asian Precipitation – Highly-Resolved Observational Data Integration Towards Evaluation of water resources (APHRODITE) version 1101R2 for Monsoon Asia (Yatagai *et al.*, 2012) is prepared at the daily scale through an international cooperative program by collecting and analysing thousands of Asian rain gauge observations in addition to those reporting to the World Meteorological

Organization (WMO) Global Telecommunication System (GTS). The technique used for the preparation of this dataset is similar to the technique used by Xie *et al.* (2007) for the production of CPC data. However, APHRODITE has better collection of gauge measurements from the South Asian region including India. Thirdly, the updated Climate Research Unit (CRU) version 3.10.01 rainfall data are prepared from monthly observations at meteorological stations across the global land areas. The CRU dataset provides anomalies from a 1961 to 1990 mean and allow absolute monthly values to be obtained when combined with the climatology (Harris *et al.*, 2014). Finally, the most recent version of the Global Precipitation Climatology Center (GPCC) Full Data Reanalysis version 6 (Schneider *et al.*, 2014) comprises precipitation data on a monthly basis from a variety of sources such as WMO GTS, synoptic weather reports (SYNOP) and monthly climate reports (CLIMAT). It may be noted that some datasets have many common input satellite and gauge information (Table 1).

For comparison purposes, in this study we resample each rainfall dataset to 1° latitude/longitude resolution using bi-linear interpolation technique and take the seasonal (JJAS) mean. We perform visual geographical verification and time-series analysis over India.

Anomaly correlation coefficient

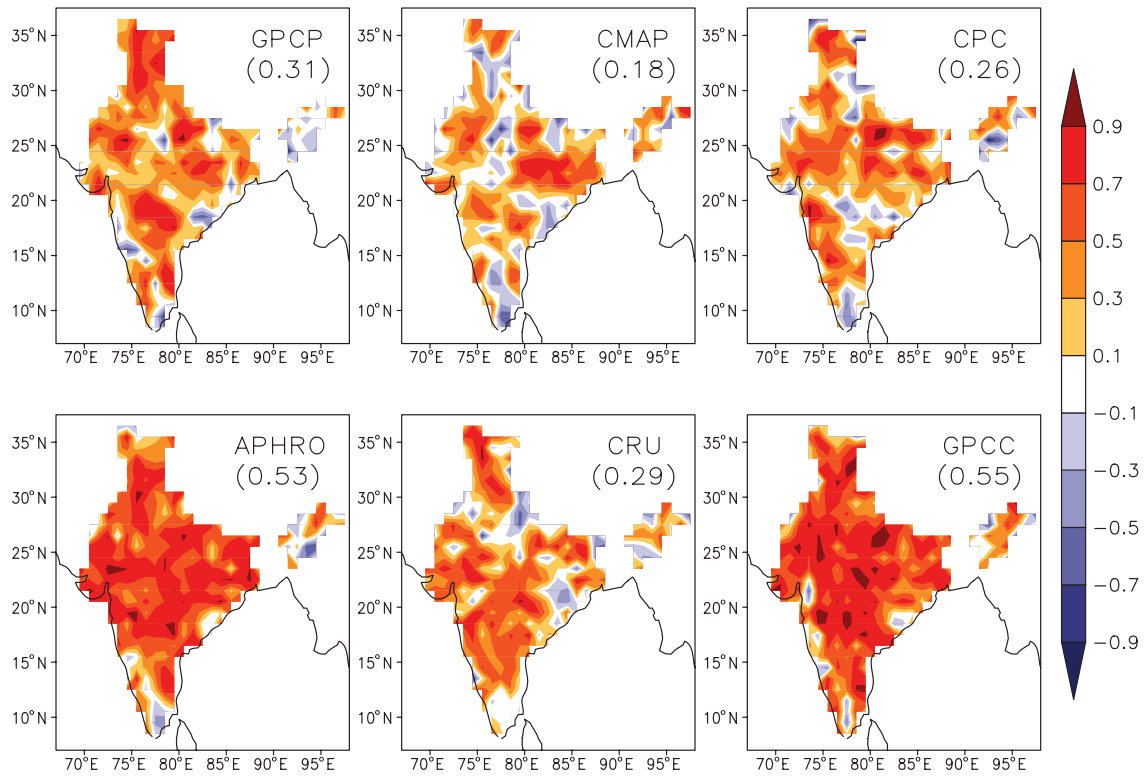


Figure 4. Spatial distributions of anomaly correlation coefficient of each data set with the IMD data during the JJAS southwest monsoon averaged for the period 1979–2005. All-India mean values are given in the parentheses.

RMSE ( $\text{mm day}^{-1}$ )

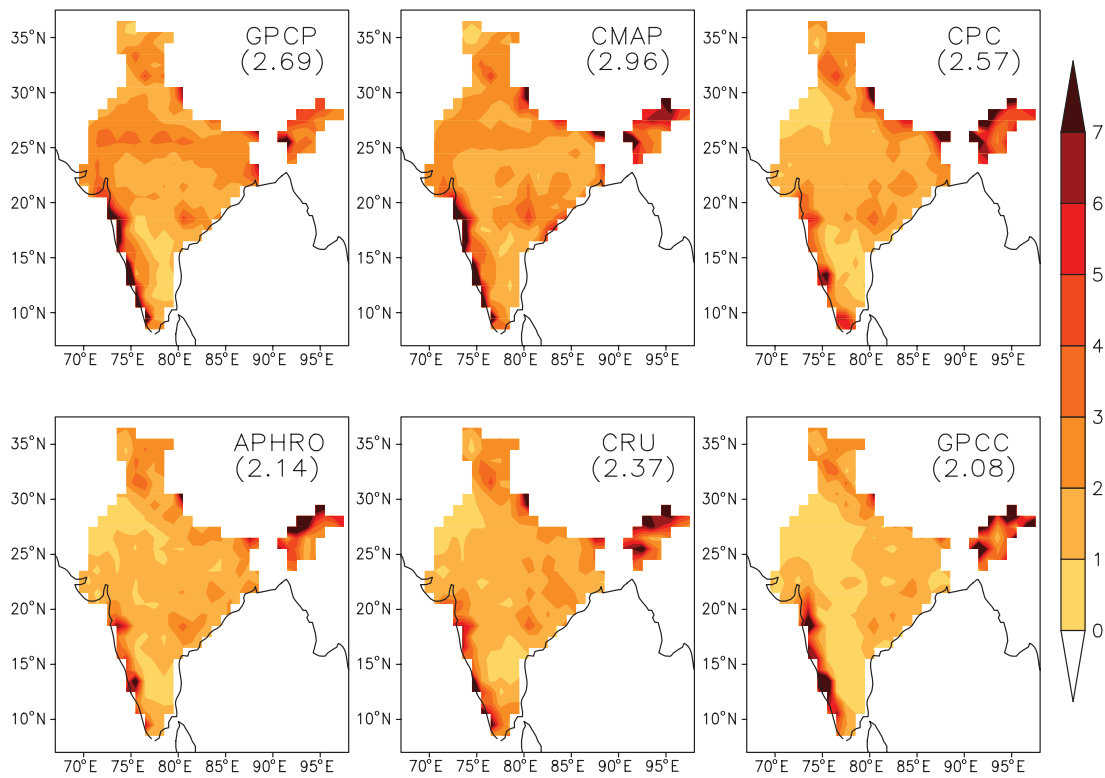


Figure 5. Spatial distributions of RMSE with respect to IMD rain gauge data during the JJAS southwest monsoon averaged for the period 1979–2005. All-India mean values are given in the parentheses.

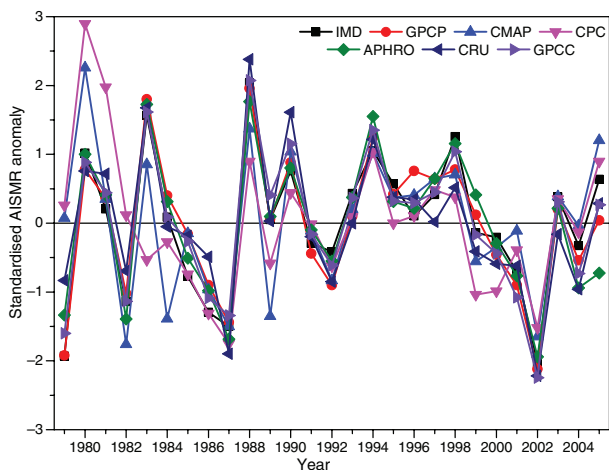


Figure 6. Interannual variations of standardized JJAS AISMR anomaly.

### 3. Results and discussions

#### 3.1. Spatial distributions of error over India

In this section, we present the spatial distributions of mean monsoon rainfall and various skill metrics against IMD gridded data for the study period over India. Figure 1 shows the mean seasonal rainfall over India from all seven datasets. The large-scale patterns of the monsoon rainfall such as heavier orographic rainfall along the Western

Ghats mountains and in Northeast India, and lower rainfall over the northwest desert region and rain-shadow region of Southeast Peninsular India is qualitatively captured by all the rainfall products. However, quantitative comparisons reveal some differences among them. The all-India mean rainfall values shown in Figure 1 reveal that, except for GPCP and GPCC, the four remaining rainfall datasets show lower values compared to observed IMD rainfall. Both the merged rainfall products GPCP and CMAP show less rainfall along the Western Ghats mountain region and higher amounts over Northwest India. In CMAP, the northeast part of India shows relatively lower values. In the merged product category, overall GPCP performs better than CMAP in comparison with IMD data. Better performance of GPCP over CMAP in the tropical regions is also reported by Yin *et al.* (2004). Over the Western Ghats, all gauge-only datasets show higher and more realistic rainfall amounts compared to the merged data products. All the gauge-only datasets exhibit lower values over extreme Northern India covering Jammu & Kashmir (J&K) region when compared to the IMD reference observations. These facts are again confirmed in Figure 2 showing the spatial distributions of bias with respect to IMD. The overestimation of rainfall along the monsoon trough region (central India) and underestimation along the Western Ghats mountain region and the foothills of the Himalayas by both the merged products are evident.

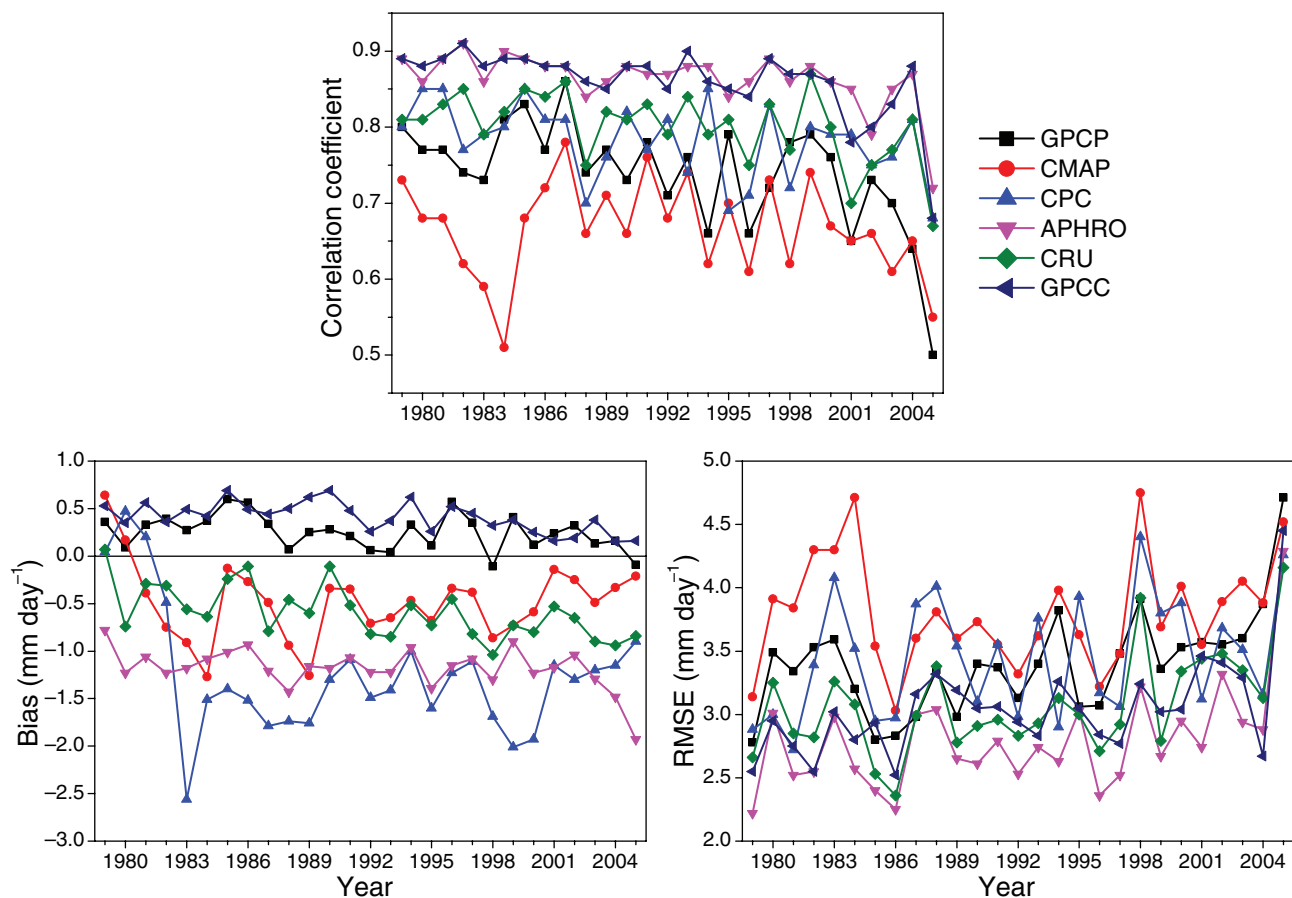


Figure 7. Interannual variations of correlation coefficient, bias and RMSE in JJAS AISMR with respect to IMD rain gauge data.



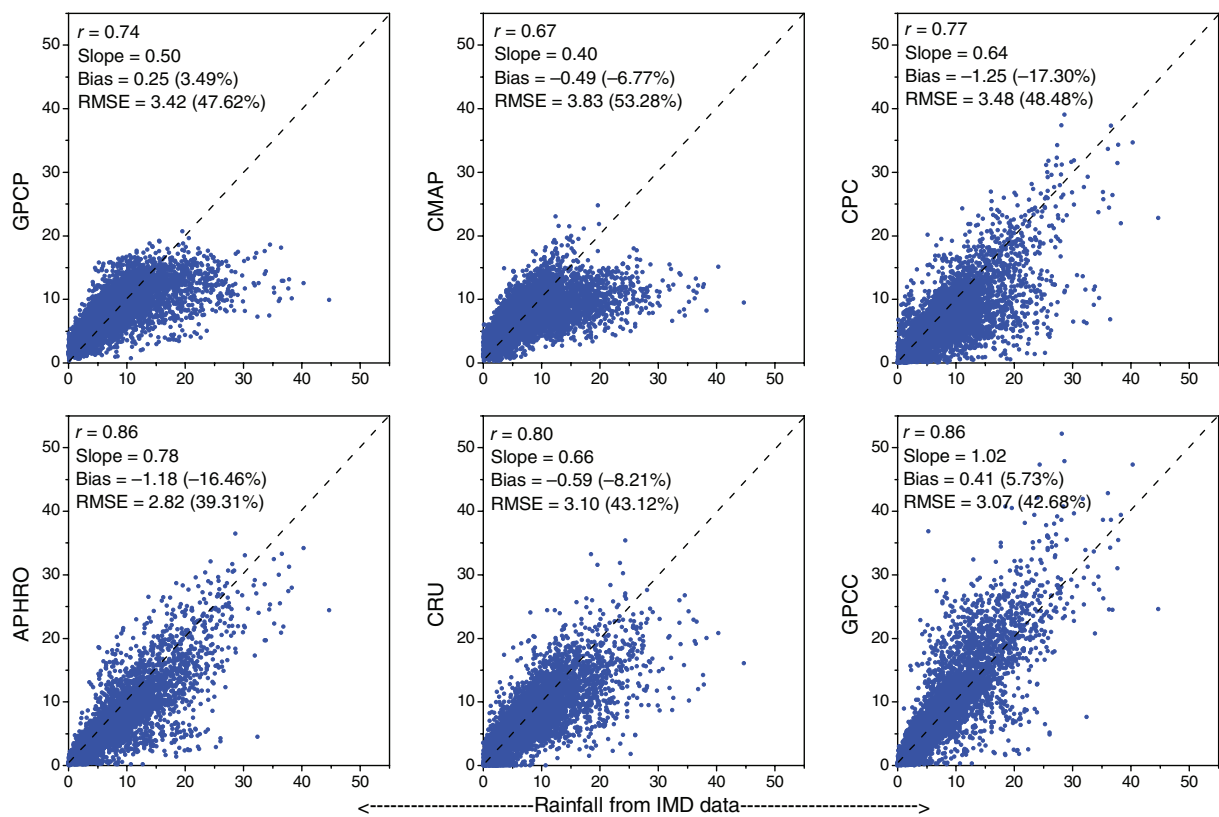
Scatter plots of JJAS rainfall ( $\text{mm day}^{-1}$ ) [ $N = 9099$ ]

Figure 8. Scatter plots of JJAS AISMR between IMD rain gauge data and other rainfall products for the period 1979–2005. Dashed black line represents 1 : 1 best-fit line and correlation coefficient ( $r$ ), slope, bias and RMSE are given in each plot.

The all-India average bias (values shown in parentheses in Figure 2) of GPCP is smaller compared to CMAP. However, both the merged products will be superior in regions where gauge data are very rare. The gauge-only (land) datasets also show smaller biases compared to merged data. GPCC shows the lowest bias in the average all-India value. However, GPCC notably overestimates rainfall along the Western Ghats and in Northeast India by about  $5 \text{ mm day}^{-1}$  when compared to IMD rainfall data. These are relatively high mountain regions, and it is possible that the interpolation and calibration algorithms could contribute to these higher values. Even the mountain regions of Northeast India show high biases in gauge-only datasets. Excluding mountain regions (considering plains only), all four gauge-only datasets compare well over India.

Now, we discuss various metrics of statistical skill such as correlation coefficient (CC), anomaly correlation coefficient (ACC) and root mean square error (RMSE). The definitions of these common metrics are described in detail by Stanski *et al.* (1989). The spatial distributions of CC, ACC and RMSE with respect to observed rainfall data for the same period over India are presented in Figures 3–5, respectively. The corresponding all-India values are given in parenthesis in each plot. In Figure 3, the single value in parenthesis represents the pattern correlation with respect to IMD dataset. Pattern correlation measures

the association between two distinct rainfall datasets in the representation of large-scale patterns (Sperber *et al.*, 2013). The ACC score (Figure 4) is used to assess the skill of a particular dataset in terms of its capability in capturing the anomalous monsoon seasons seen in interannual variations over the study period. Between the two merged products, GPCP noticeably shows a higher correlation and pattern correlation than CMAP. In addition, GPCP has higher ACC and lower RMSE compared to CMAP. Therefore, between the merged datasets, GPCP seems to be superior. However, along the Western Ghats, both merged datasets show lower correlations and higher RMSE, which are consistent with higher biases from both datasets as shown in Figure 2. In particular, CMAP has very high RMSE values over Northeast India. Other mountain regions of J&K and the Himalayas also show low correlation values in both the merged products. The pockets of negative ACC in CMAP are of concern, indicating in some cases, the anomalies of interannual variation of monsoon rainfall. Among the four gauge-only rainfall products, higher CC and ACC are seen over most parts of the country in GPCC and APHRODITE. Both APHRODITE and GPCC show the largest pattern correlations of 0.91 among the six rainfall products. Lower correlations are seen over Northeastern India and J&K in all four gauge-only products. CPC and CRU show rather lower correlations with IMD gridded observations over Eastern India, a region

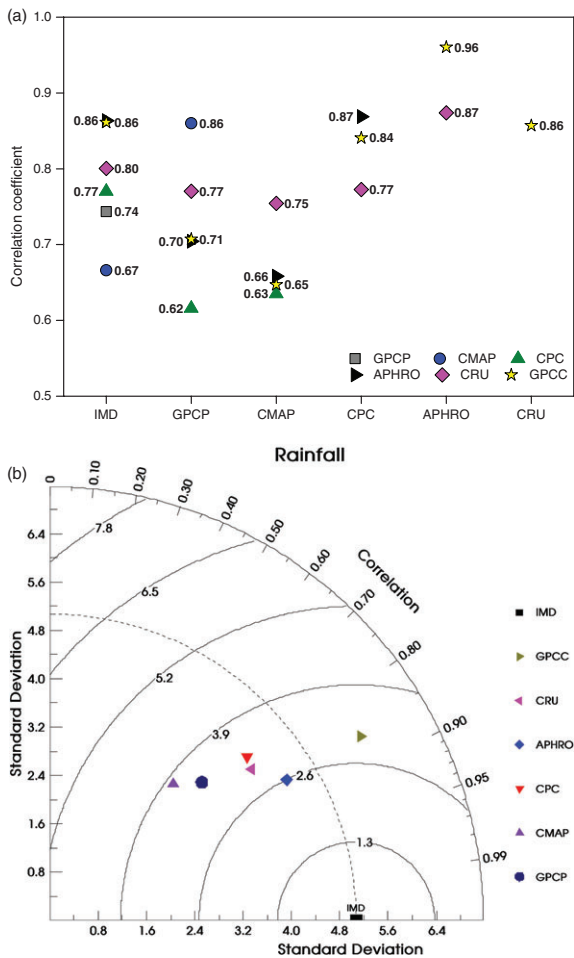


Figure 9. (a) Cross-correlation coefficients of AISMR among seven rainfall products and (b) Taylor diagram showing the comparison of JJAS AISMR from six rainfall products against IMD gauge-based data.

where a lot of monsoon low-pressure systems contribute to the total rainfall of the season (Krishnamurthy and Ajayamohan, 2010). Both CPC and CRU also show low correlations near the foothills of the Himalayas. Along the Western Ghats, all four gauge-only datasets have good CC and ACC, which was not seen in the merged datasets. In terms of RMSE, all three gauge-only datasets except GPCP show lower values over the Western Ghats mountain region. This suggests that either the gauge density used in the merged products is not high enough in the Western Ghats, or that the satellite algorithms behave poorly over narrow orography. The lower RMSE of CPC, CRU and APHRODITE over such orographic regions is not seen in the merged datasets, GPCP and CMAP. The lowest RMSE and the higher CC and ACC of APHRODITE and GPCP (except for the Western Ghats) products over India suggest that APHRODITE and GPCP are more reliable over India among the gauge-only datasets in comparison to the reference IMD gridded rainfall.

### 3.2. Comparison at the all-India scale

In this section, the comparison of six gridded rainfall products with the reference IMD observed data is performed at the country-wide scale for the period 1979–2005. The

mean rainfall from June to September over India is considered as all-India summer monsoon rainfall (AISMR). The standardized anomaly of AISMR during each year from these rainfall products is shown as a time-series in Figure 6. This anomaly is computed against the respective means of each dataset. In most of the years, the interannual variability of monsoon rainfall is reasonably well captured by all rainfall products. However, during certain years (1983, 1984, 1989 and 1999) CPC and CMAP products show a noticeable difference from the other products including errors of sign. APHRODITE shows a negative anomaly in 2005, whereas the other rainfall datasets show positive anomaly. The interannual variations of the skill metrics (CC, bias and RMSE) of AISMR from the six rainfall products with respect to the IMD gridded reference data are shown in Figure 7. The bias plot shows a systematic overestimation of rainfall by GPCP and GPCP data. Moreover, CPC, APHRODITE, CRU and CMAP systematically underestimate AISMR compared to IMD. The lowest bias in AISMR is evident in the GPCP data. CMAP has the lowest CC and the largest RMSE. Among the gauge-only data, GPCP and APHRODITE have higher CC and lower RMSE than the others. The wide range of CC, bias (and even difference of sign) and RMSE are points of concern in selecting a particular dataset.

Figure 8 shows a scatter plot of monsoon rainfall from the IMD gridded reference data and six other rainfall datasets compared at each grid point over India. The single value of correlation coefficient ( $r$ ), slope, bias (percentage with respect to mean IMD gridded data) and RMSE (percentage with respect to mean IMD gridded data) shown in each panel of Figure 8 indicates the overall (all-India) skill of each dataset. Comparing the two merged datasets, GPCP shows better scores than CMAP. Among the gauge-only datasets, APHRODITE and GPCP are the two best performers in terms of these skill metrics against the IMD gridded data. Both GPCP and APHRODITE show the largest correlation of 0.86, but RMSE is lower in APHRODITE than GPCP. For CPC, both the bias and RMSE are the highest among the land-only products. However, GPCP overestimates monsoon rainfall over India whereas APHRODITE systematically underestimates rainfall in comparison with the IMD gridded data. In order to compare all seven rainfall products used in this study, the cross-correlation values for the AISMR for the 27-year period are shown in Figure 9(a). It is shown that the correlation among the seven different datasets lies in the range of 0.62–0.96. GPCP is highly correlated with APHRODITE on interannual time scales. GPCP and APHRODITE also have higher correlations with the IMD gridded rainfall data. In addition, the merged datasets, GPCP and CMAP, have very good correlations with each other, reflecting some of the common satellite and gauge sources used by these two datasets. It is also useful to examine the correlation of the datasets along with their respective variability in terms of standard deviations and RMSEs. Figure 9(b) shows the popular Taylor diagram (Taylor, 2001) presenting the correlations simultaneously with the RMSE and standard deviations of AISMR. Even

COMPARISON OF MONSOON RAINFALL

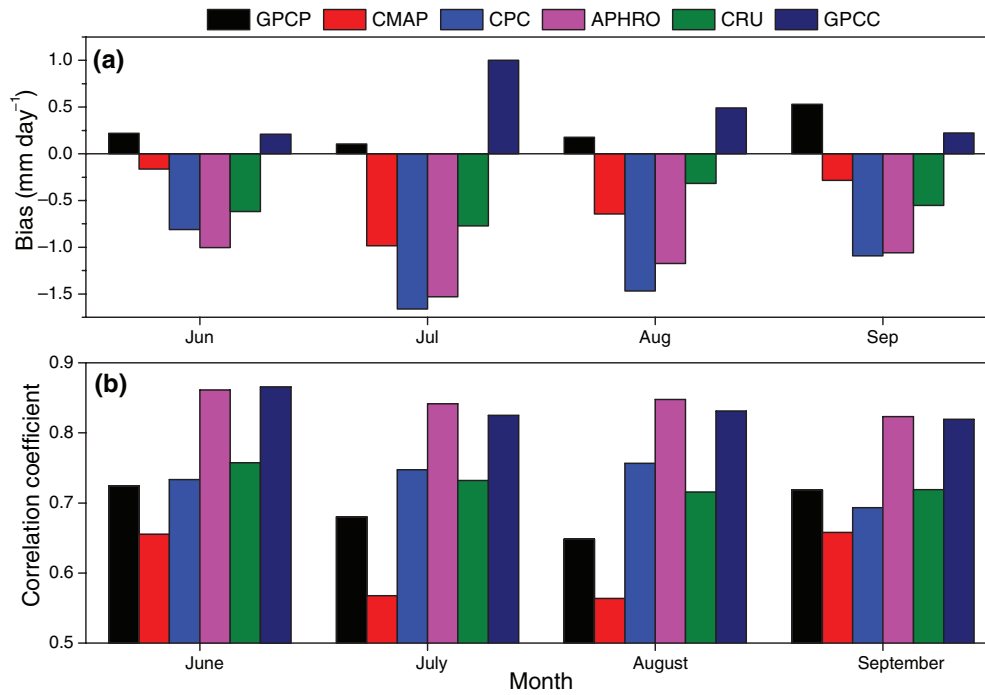


Figure 10. (a) Bias and (b) correlation coefficient of AISMR in different rainfall datasets at monthly scale over 1979–2005 for the southwest monsoon season.

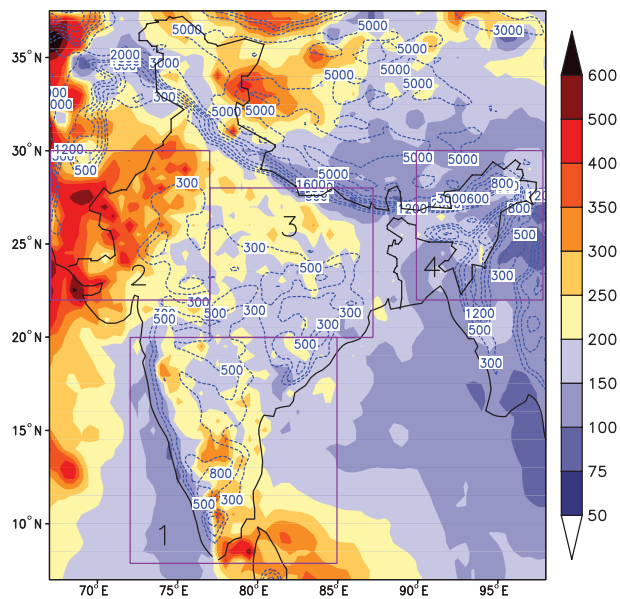


Figure 11. Coefficient of variation (%; shaded) of rainfall and orography (metres; dashed contours) for the Indian monsoon region. The four subregions of India shown by box 1, 2, 3 and 4 used for detailed analysis and have reference to Figures 12, 13 and 15.

if the correlations are good, the wide range of standard deviations is a cause of concern. The Taylor diagram clearly shows APHRODITE and GPCC to be the best performers in terms of all-India interannual variability, with each underestimating and overestimating interannual variability, respectively, relative to the IMD reference observations.

After examining the skill at seasonal scale, we enquire which month(s) within the monsoon season contribute maximum to the overall seasonal performance. Figure 10 shows the biases and correlations of different datasets against IMD gridded data for different months within the monsoon season. The skill at monthly scale from the different datasets is similar to the seasonal skills. However, as the rainfall amounts are higher during the two peak monsoon months of July and August, the biases are also seen to be proportionately higher. Overall, this suggests that there is consistent performance of the various datasets throughout the season.

### 3.3. Comparison at the subregional scale

In this section, the Indian monsoon rainfall is compared at a subregional scale. We selected four subregions of India (Figure 11) for the analysis based on spatial rainfall homogeneity and topography characteristics. Figure 11 shows the height of mountains (dashed contours). The colour shades show the subseasonal coefficient of variation of rainfall in percentage with respect to mean monsoon values. The choice of subseasonal variability from daily merged gridded rainfall data (Mitra *et al.*, 2009) helps in deciding distinct rainfall regions. The four selected regions are unique in terms of variations in rainfall and orography, and hence are good test-beds for the evaluation of any rainfall product. Region 1 comprises highly complex orography with large east–west gradients of the monsoon rainfall on both sides of the Western Ghats mountain region. The windward side receives very heavy rainfall, whereas the leeward side gets low rainfall during the monsoon season (Figure 1). This region also includes the dry rain-shadow area over Southeast Peninsular India. As we compare

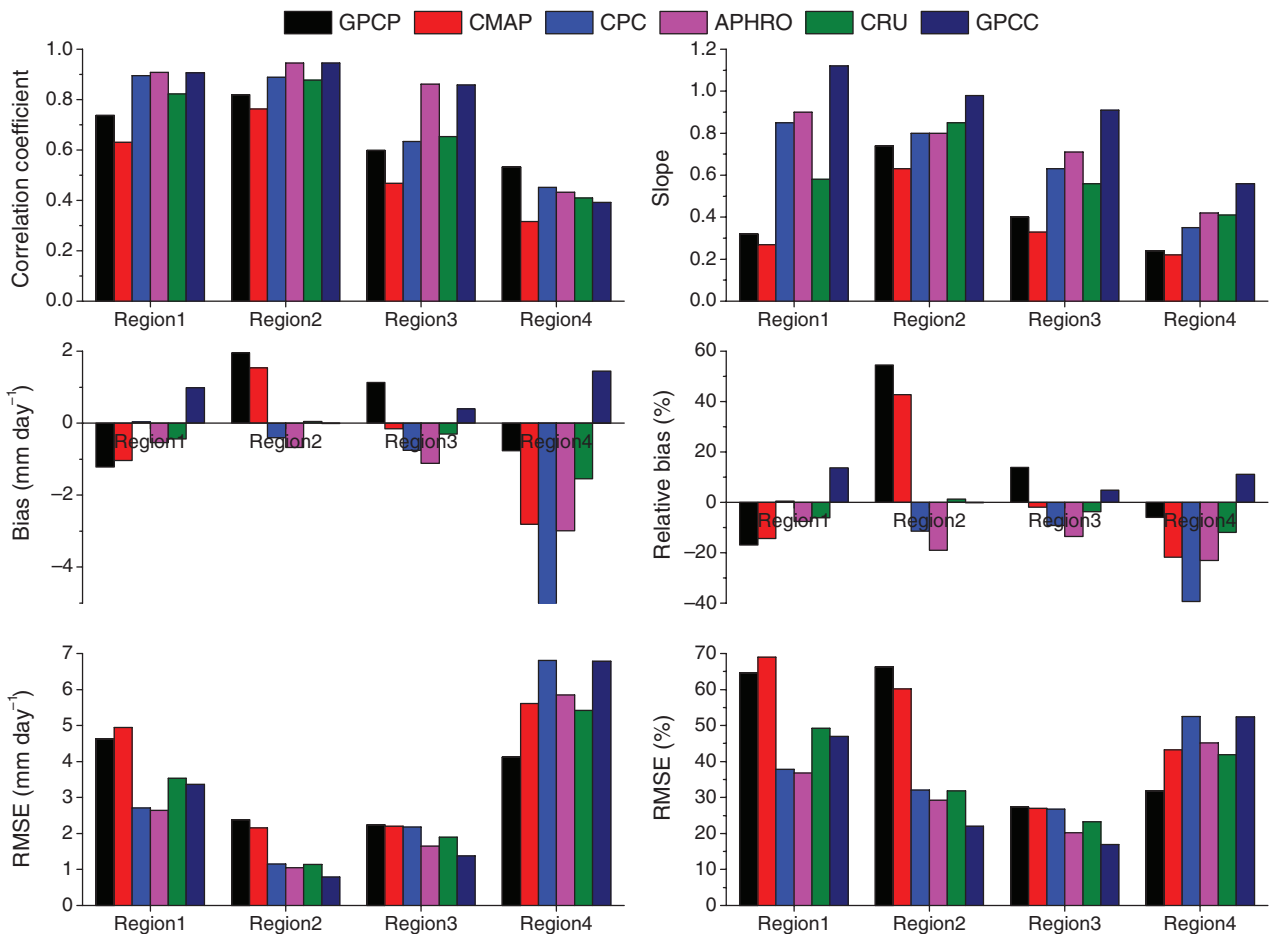


Figure 12. Regional variations of JJAS correlation coefficient, slope, bias and RMSE of the seasonal monsoon rainfall for the period of 1979–2005.

different rainfall datasets at larger spatial resolution ( $1^\circ$  latitude/longitude), the contrasting rainfall regions of Southeast Peninsular India are taken as one subregion. In region 2, Northwest India receives comparatively lower rainfall during the southwest monsoon season as it is a semi-arid region (containing the Thar desert) but with high rainfall variability. Region 3 contains the northern plains of the Ganges basin and receives reasonably high rainfall due to the presence of the monsoon trough and passage of low-pressure systems and monsoon depressions originating in the Bay of Bengal. Region 4 (northeast India) is a hilly region and receives very heavy rainfall during the summer monsoon period. The different error statistics including CC, slope, bias and RMSE in the monsoon rainfall over the above four subregions are shown in Figure 12. The lowest CC and the largest RMSE of the six rainfall datasets when compared with the IMD gridded data are found over the region 4, suggesting that users of these products need to exercise the most caution when assessing the monsoon rainfall over Northeast India. However, GPCP shows higher CC and slope than CMAP over all four subregions. Among the four gauge-only products, GPCC and APHRODITE show similar and higher CC than the other datasets over regions 1, 2 and 3. However, APHRODITE shows higher CC than GPCC over region 4. The change of sign seen in biases for regions 2 and 3

is again a reason of concern on the choice of a particular dataset.

Figure 13 presents the standardized anomaly (from their respective means) of the domain-mean seasonal monsoon rainfall for the four subregions. The regions of high and low seasonal rainfall are well captured by all the rainfall datasets for regions 1, 2, and 3, whereas there is notable disagreement among the rainfall products over region 4 in Northeast India. Region 4 is known for very high convective rainfall associated with complex orography and rainfall processes of the monsoon system. In terms of the monsoon flow, both at lower and upper levels, this region is significant, containing deep convective clouds. Satellite estimates of rainfall in this environment have less reliability due to ambient atmospheric conditions (Adler *et al.*, 2003). Additionally, it is also true that region 4 has fewer rain gauges due to the difficult terrain. All these factors combined may be contributing to the higher uncertainty in region 4.

Figure 14 shows the biases and correlations in different datasets with respect to varying orography. It is seen that the biases in rainfall increase significantly in regions with orography exceeding 1000-m height. The interannual variations of differences (bias) in the six rainfall datasets with respect to the IMD gridded data over the four subregions are presented in Figure 15. For all regions, we notice

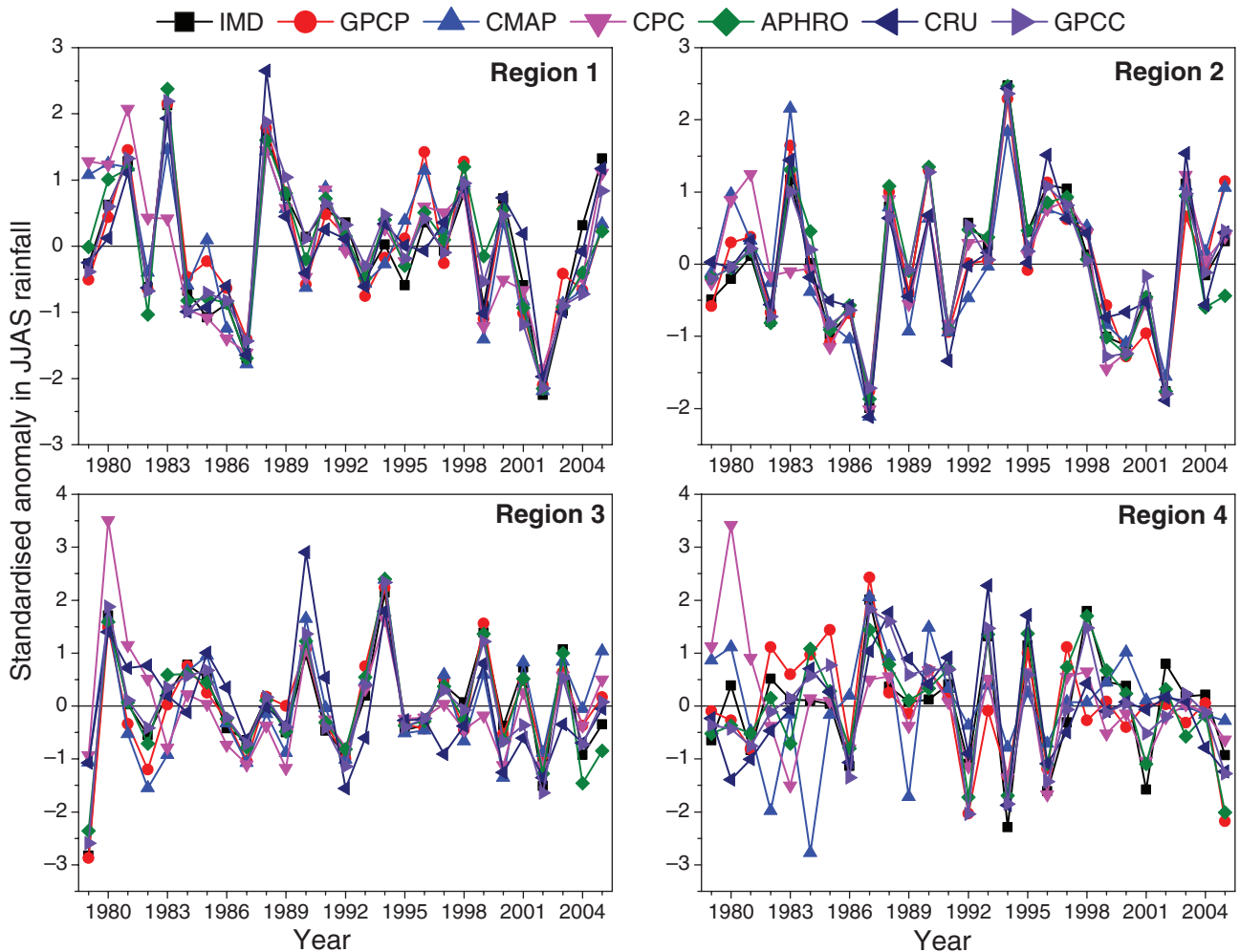


Figure 13. Interannual variations of standardized anomaly of seasonal rainfall during the JJAS southwest monsoon.

a wide range of biases over the study period. It is also noticed that the sign of bias varies from one dataset to the other over different regions during certain years. Among the merged datasets, GPCP has the least bias overall for the four regions. Even though GPCC and APHRODITE show smaller biases, their signs of bias are seen to be opposite in regions 1, 3 and 4.

#### 4. Conclusions

Indian summer monsoon rainfall and its variability is a scientifically interesting and crucial parameter for various hydrometeorological applications, evaluation of numerical models and their development and for studies of climate change. In this study, two merged gridded rainfall products (GPCP and CMAP) and four gauge-based land-only rainfall products (CPC, APHRODITE, CRU and GPCC) were compared with the IMD  $0.5^\circ$  gridded rainfall dataset based on a dense network of gauges all at a common  $1^\circ$  latitude/longitude resolution. Various skill metrics including bias, CC, ACC and RMSE, and subjective comparisons were carried out over the Indian land region for 27 summer monsoon seasons (1979–2005). In the merged category,

in general, GPCP was seen to be better than CMAP for representing the Indian monsoon rainfall whereas in gauge-only category, APHRODITE and GPCC performed better compared to other datasets. These results rely on our assumptions that the IMD gridded data most accurately characterizes the Indian monsoon rainfall. APHRODITE and GPCC were highly correlated with each other over the Indian monsoon region. However, most of the datasets had difficulties in representing orographic rainfall particularly over the Western Ghats Mountains, in Northeast India and in the Himalayan foothills. Additional gauges over these difficult mountainous regions will be highly beneficial. Additionally, the wide range of skill metrics seen among the datasets and even the change of sign of the bias found in some years are causes of concern in terms of uncertainty in choosing an appropriate dataset for inputs to models or model evaluation. The comparison at subregional scales showed that even though all the rainfall datasets were able to capture the interannual variability reasonably well, there were considerable regional differences among them in terms of bias and other skill metrics. Each of the datasets studied has its own strengths and deficiencies. The skill of merged products could possibly be enhanced by putting more number of gauges.

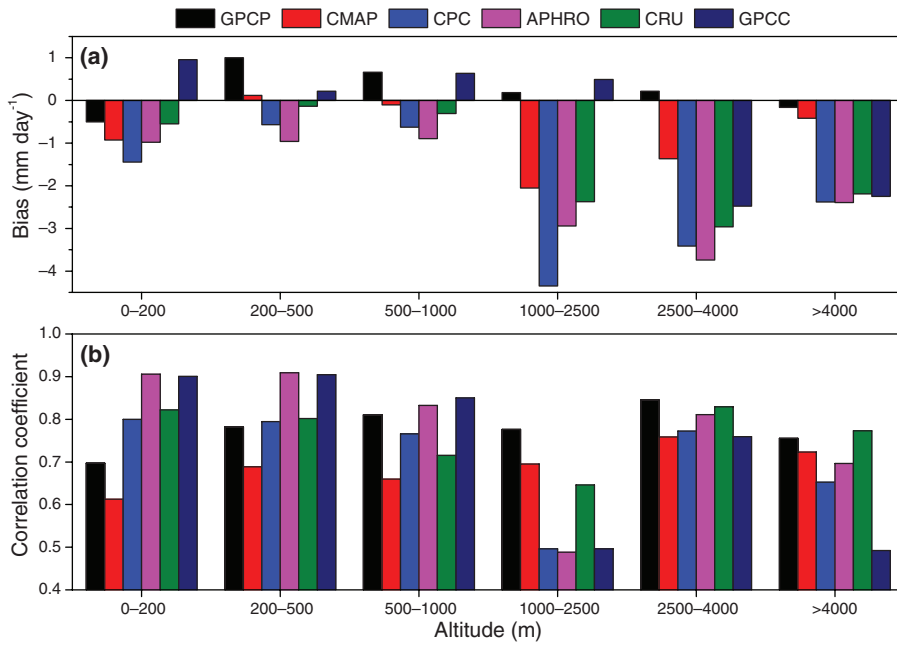


Figure 14. (a) Bias and (b) correlation coefficient of monsoon rainfall from six datasets with respect to IMD gauge-based product for different altitude ranges.

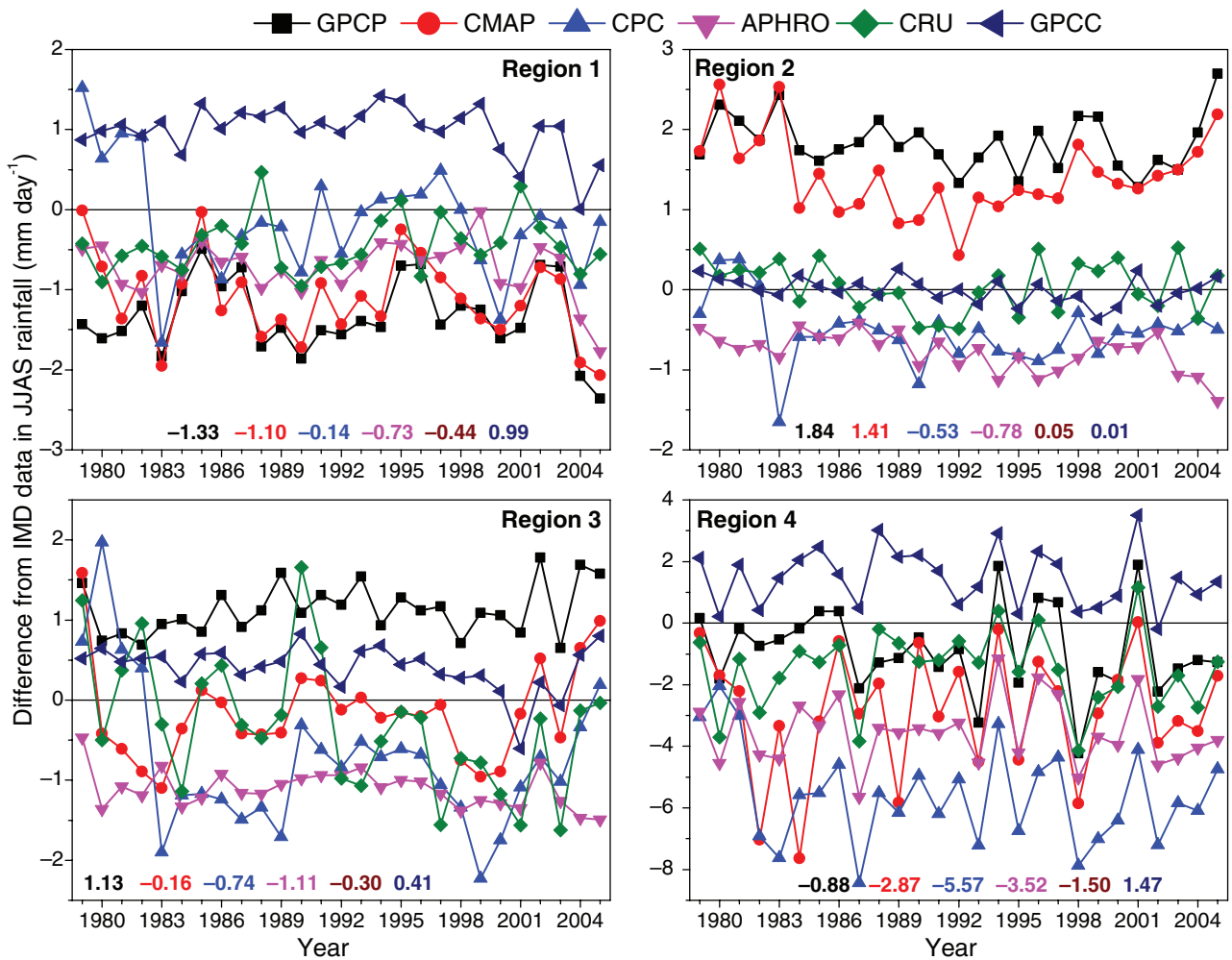


Figure 15. Interannual variations of bias with respect to IMD rain gauge data during the JJAS southwest monsoon for the period 1979–2005. Domain-mean values are given in each plot.

These results will help monsoon data users to select an appropriate dataset depending on their application. This analysis suggests the importance of continued evaluation and production of improved rainfall datasets for the monsoon (Mitra *et al.*, 2013b). It may be possible to combine multisatellite estimates and gauge information from the South Asia, like IMD and APHRODITE/GPCC, to prepare a comprehensive dataset at monthly/seasonal scale. A suitable combination of available datasets for a multianalysis-based consensus dataset could also be superior to any individual dataset (Adler *et al.*, 2009; Tian and Peters-Lidard, 2010). We hope that the uncertainty for orographic regions will be taken care in upcoming datasets by recent satellite estimates from the Global Precipitation Mission (GPM) and use of more gauge data.

### Acknowledgements

Rainfall datasets used here (GPCP, CMAP, CPC, APHRODITE, CRU, GPCC) were obtained from respective project sites via ftp. IMD gridded rainfall data was made available from IMD, Pune. This work is carried out to meet the goals of the MoES-NERC Changing Water Cycle (South Asia) project SAPRISE (NE/I022841/1 and MoES/NERC/16/02/10PC-II). A. G. Turner was funded under a NERC Fellowship (grant number NE/H015655/1). Comments from three anonymous reviewers helped greatly to improve the quality of the paper.

### References

- Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P. 2003. The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–Present). *J. Hydrometeorol.* **4**: 1147–1167.
- Adler RF, Wang JJ, Gu G, Huffman GJ. 2009. A ten-year tropical rainfall climatology based on a composite of TRMM products. *J. Meteorol. Soc. Jpn.* **87A**: 281–293.
- Collins M, AchutaRao K, Ashok K, Bhandari S, Mitra AK, Prakash S, Srivasatva R, Turner A. 2013. Observational challenges in evaluating climate models. *Nat. Clim. Change* **3**(11): 940–941, doi: 10.1038/nclimate2012.
- Ebert EE, Janowiak JE, Kidd C. 2007. Comparison of near real-time precipitation estimates from satellite observations and numerical models. *Bull. Am. Meteorol. Soc.* **88**: 47–64.
- Harris I, Jones PD, Osborn TJ, Lister DH. 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol.* **34**(3): 623–642, doi: 10.1002/joc.3711.
- Hoskins B. 2013. The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Q. J. R. Meteorol. Soc.* **139**: 573–584, doi: 10.1002/qj.1991.
- Kidd C, Levizzani V. 2011. Status of satellite precipitation retrievals. *Hydrol. Earth Syst. Sci.* **15**: 1109–1116, doi: 10.5194/hess-15-1109-2011.
- Krishnamurthy V, Ajayamohan RS. 2010. Composite structure of monsoon low pressure systems and its relation to Indian rainfall. *J. Clim.* **23**: 4285–4305.
- Kucera PA, Ebert EE, Turk FJ, Levizzani V, Kirschbaum D, Tapiador FJ, Loew A, Borsche M. 2013. Precipitation from space: advancing earth system science. *Bull. Am. Meteorol. Soc.* **94**: 365–375, doi: 10.1175/BAMS-D-11-00171.1.
- Mitra AK, Bohra AK, Rajeevan MN, Krishnamurti TN. 2009. Daily Indian precipitation analyses formed from a merge of rain-gauge with TRMM TMPA satellite derived rainfall estimates. *J. Meteorol. Soc. Jpn.* **87A**: 265–279.
- Mitra AK, Rajagopal EN, Iyengar GR, Mohapatra DK, Momin IM, Gera A, Sharma K, George JP, Ashrit R, Dasgupta M, Mohandas S, Prasad VS, Basu S, Arribas A, Milton SF, Martin GM, Barker D, Martin M. 2013a. Prediction of monsoon using a seamless coupled modelling system. *Curr. Sci.* **104**(10): 1369–1379.
- Mitra AK, Momin IM, Rajagopal EN, Basu S, Rajeevan MN, Krishnamurti TN. 2013b. Gridded daily Indian monsoon rainfall for 14 seasons: merged TRMM and IMD gauge analyzed values. *J. Earth Syst. Sci.* **122**: 1173–1182, doi: 10.1007/s12040-013-0338-3.
- Rajeevan M, Bhatte J. 2009. A high resolution daily gridded rainfall dataset (1971–2005) for mesoscale meteorological studies. *Curr. Sci.* **96**(4): 558–562.
- Schneider U, Becker A, Finger P, Meyer-Christoffer A, Ziese M, Rudolf B. 2014. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.* **115**: 15–40, doi: 10.1007/s00704-013-0860-x.
- Slingo J, Palmer T. 2011. Uncertainty in weather and climate prediction. *Philos. Trans. R. Soc. A* **369**: 4751–4767, doi: 10.1098/rsta.2011.0161.
- Sorooshian S, AghaKouchak A, Arkin P, Eytan J, Fofoula-Georgiou E, Harmon R, Hendrickx JMH, Imam B, Kuligowski R, Skahill B, Skofronick-Jackson G. 2011. Advancing concepts of remote sensing of precipitation at multiple scales. *Bull. Am. Meteorol. Soc.* **92**(10): 1353–1357, doi: 10.1175/2011BAMS3158.1.
- Sperber KR, Annamalai H, Kang I-S, Kitoh A, Moise A, Turner A, Wang B, Zhou T. 2013. The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Clim. Dyn.* **41**: 2711–2744, doi: 10.1007/s00382-012-1607-6.
- Stanski HR, Wilson LJ, Burrows WR. 1989. Survey of common verification methods in meteorology. WWC Technical Report No. 8, WMO/TD No. 358, WMO, Geneva, Switzerland, 114 pp.
- Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **106**(D7): 7183–7192.
- Tian Y, Peters-Lidard CD. 2010. A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.* **37**: L24407, doi: 10.1029/2010GL046008.
- Trenberth KE, Stepaniak DP, Caron JM. 2000. The global monsoon as seen through the divergent atmospheric circulation. *J. Clim.* **13**: 3969–3993.
- Turner AG, Annamalai H. 2012. Climate change and the south Asian summer monsoon. *Nat. Clim. Change* **2**: 587–595, doi: 10.1038/nclimate1495.
- Turner A, Sperber K, Slingo J, Meehl G, Mechoso CR, Kimoto M, Giannini A. 2011. Modelling monsoons: understanding and predicting current and future behavior. In *The Global Monsoon System: Research and Forecast*. World Scientific Series on Asia-Pacific Weather and Climate, Vol. 5, 2nd edn, Chang CP, Ding Y, Lau GNC, Johnson RH, Wang B, Yasunari T (eds). World Scientific Publishing Company: Singapore, 421–454.
- Xie P, Arkin PA. 1996. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Am. Meteorol. Soc.* **78**: 2539–2558.
- Xie P, Yatagai A, Chen M, Hayakawa T, Fukushima Y, Liu C, Yang S. 2007. A gauge-based analysis of daily precipitation over East Asia. *J. Hydrometeorol.* **8**: 607–626.
- Yatagai A, Kamiguchi K, Arakawa O, Hamada A, Yasutomi N, Kitoh A. 2012. APHRODITE: constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Am. Meteorol. Soc.* **93**(9): 1401–1415, doi: 10.1175/BAMS-D-11-00122.1.
- Yin X, Gruber A, Arkin P. 2004. Comparison of the GPCP and CMAP merged gauge-satellite monthly precipitation products for the period 1979–2001. *J. Hydrometeorol.* **5**: 1207–1222, doi: 10.1175/JHM-392.1.