

# Developing High-resolution Gridded Rainfall and Temperature Data for Bangladesh: the ENACTS-BMD dataset

Nachiketa Acharya<sup>1</sup>, Rija Faniriantsoa<sup>1</sup>, Bazlur Rashid<sup>2</sup>, Razia Sultana<sup>2</sup>, Carlo Montes<sup>3</sup>, Tufa Dinku<sup>1</sup>, and S.M.Q. Hassan<sup>2</sup>

<sup>1</sup> International Research Institute for Climate and Society (IRI), Columbia University, Palisades, NY, USA

<sup>2</sup> Bangladesh Meteorological Department (BMD), Dhaka, Bangladesh

<sup>3</sup> International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

\*Corresponding author: [nachiketa@iri.columbia.edu](mailto:nachiketa@iri.columbia.edu)

**Abstract :** This manuscript describes the construction and validation of high resolution daily gridded ( $0.05^\circ \times 0.05^\circ$ ) rainfall and maximum and minimum temperature data for Bangladesh : the Enhancing National Climate Services for Bangladesh Meteorological Department (ENACTS-BMD) dataset. The dataset was generated by merging data from weather stations, satellite products (for rainfall) and reanalysis (for temperature). ENACTS-BMD is the first high-resolution gridded surface meteorological dataset developed specifically for studies of surface climate processes in Bangladesh. Its record begins in January 1981 and is updated in real-time monthly and outputs have daily, decadal and monthly time resolution. The Climate Data Tools (CDT), developed by the International Research Institute for Climate and Society (IRI), Columbia University, is used to generate the dataset. This data processing includes the collection of weather and gridded data, quality control of stations data, downscaling of the reanalysis for temperature, bias correction of both satellite rainfall and downscaled reanalysis of temperature, and the combination of station and bias-corrected gridded data. The ENACTS-BMD dataset is available as an open-access product at BMD's official website, allowing the enhancement of the provision of services, overcoming the challenges of data quality, availability, and access, promoting at the same time the engagement and use by stakeholders.

**Keywords:** climate data, gridded product, data merging.

## 1. INTRODUCTION

Meteorological observations data (historical records and near-real-time) are the backbone of National Weather Services (NWS). Long-term historical data allow the assessment of climate risks or long-term trends in the rainfall and temperature, while near-real-time data are important for monitoring weather-related hazards, allowing the development of timely and actionable early warning systems by governments and agencies (Maidment et al.,2017). Traditionally, records from weather stations networks (in-situ) maintained by NWS provide the most reliable means to obtain accurate local information about weather and climate. However, stations networks are often spatially sparse and missing records issues pose a challenge in terms of information gaps. Therefore, high-quality datasets with an appropriate spatial and temporal coverage are highly demanded.

For Bangladesh, long-term rainfall records are available for multiple weather stations that are distributed all over the country from the Bangladesh Meteorological Department (BMD) (Fig.1). However, some of the stations have long periods of missing data, and although the spatial coverage is adequate considering the size of the country, over some regions they are sparsely distributed, especially in the Northeast, the rainiest region of the country. Consequently, large areas of Bangladesh have an inadequate coverage which, for example, limits the monitoring and analysis of small-scale and short life span convective systems, making rain-gauge measurements to be used as representative of the several-square kilometers area surrounding the station. Similar, sparse temperature records can be representative of a small area when factors such as topography modify the spatial patterns, limiting the use of neighboring stations for data filling. Moreover, since Bangladesh is very prone to extreme weather events leading to flash floods, drought or heatwaves, high quality rainfall and temperature data for their monitoring are of great importance, and also for other applications such as services for agriculture and water resources management in remote areas (Nashwan et al. 2019). However, simple statistical interpolation approaches of station

data are a partial solution as the associated uncertainty can be high when the commonly used co-variables explain a small proportion of the observed variance. Limitations associated with station observations can be more important over regions with complex topography where station measurements are generally sparse or nonexistent. Notwithstanding, satellite rainfall estimates or reanalysis data which are easily and freely available provide an alternative for obtaining climate data over regions where ground observations are limited or unavailable. However, without direct reference to ground measurements, satellite rainfall estimates and reanalysis are subject to systematic bias. In principle, a suitable fusion of weather station data with satellite and reanalysis products is expected to provide sufficiently accurate spatial and temporal rainfall and temperature estimates on an operational basis for a wide range of data users.

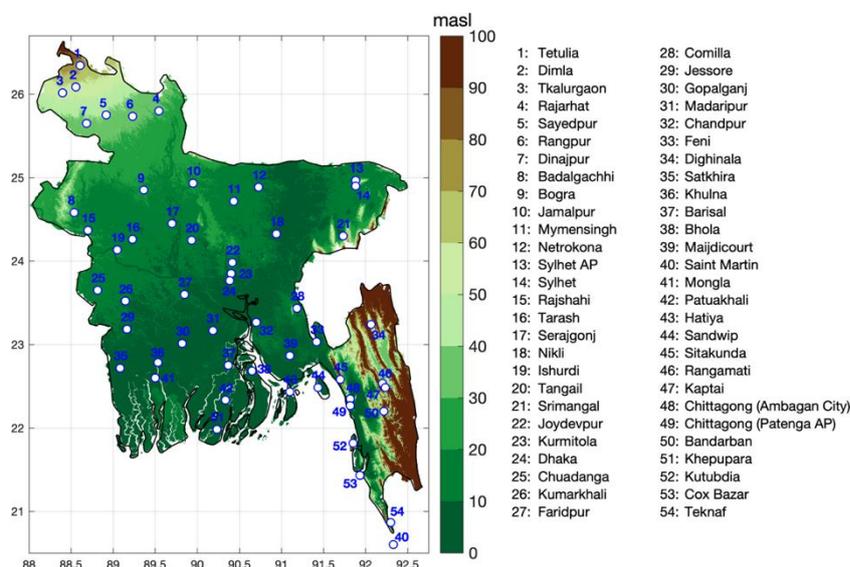


Figure 1. Topographic map of Bangladesh showing the location of the weather stations from the Bangladesh Meteorological Department (BMD). Elevation data obtained from the Shuttle Radar Topography Mission SRTM90 digital elevation model.

To overcome the problem of unavailability of long-term quality-assured climate data, a high-resolution gridded rainfall and temperature dataset, the ENACTS-BMD product, being developed as part of the “Enhancing National Climate Services” (ENACTS) initiative under the Columbia University’s World Project “Adapting Agriculture to Climate Today, for Tomorrow” (ACToday; <https://iri.columbia.edu/actoday/>) with a close collaboration between and Society (IRI), Columbia University. IRI has a long history generating high-resolution meteorological data for African countries under the ENACTS program (Dinku et al., 2017). The main processing steps to generate ENACTS-BMD include spatial downscaling and bias correction of the temperature reanalysis and satellite rainfall estimates, and finally the procedure to merge the station data with the bias-corrected gridded ones. In this study, we describe the construction and validation of the ENACTS-BMD dataset.

## 2. Development of ENACTS-BMD database

The development of the ENACTS-BD dataset considers several steps, including the data collection, quality control of stations data, spatial downscaling of reanalysis for temperature, bias correction for both rainfall and temperature gridded data, and finally the combination of station and bias-corrected gridded data. Each of the above steps is performed using the Climate Data Tools (CDT), an open-source R package based on a set of utility functions for meteorological data quality control, homogenization and merging station data with satellite and others proxies such as reanalysis. This tool is developed and maintained by the International Research Institute for Climate and Society (IRI), Columbia University. All functions in CDT are available at <https://github.com/rjfaf-iri/CDT> in graphical user interface (GUI) mode.

### 2.1 Station data availability

The main data source used to generate the ENACTS-BMD dataset are daily observations provided by BMD. However, prior 1979, data are known to be not of high quality and contain a large number of missing records. Although this data could be also considered for improvement, high-resolution satellite precipitation products are

typically available since 1980. Therefore, we decided to make ENACTS-BMD begins in January 1981. The percentage of available (non-missing) data 1981-2019 for rainfall, minimum and maximum temperature are presented in figure 2. For rainfall, 35 out of 54 available stations have more than 80% whereas most of them (54) report 90-100% of temperature data.

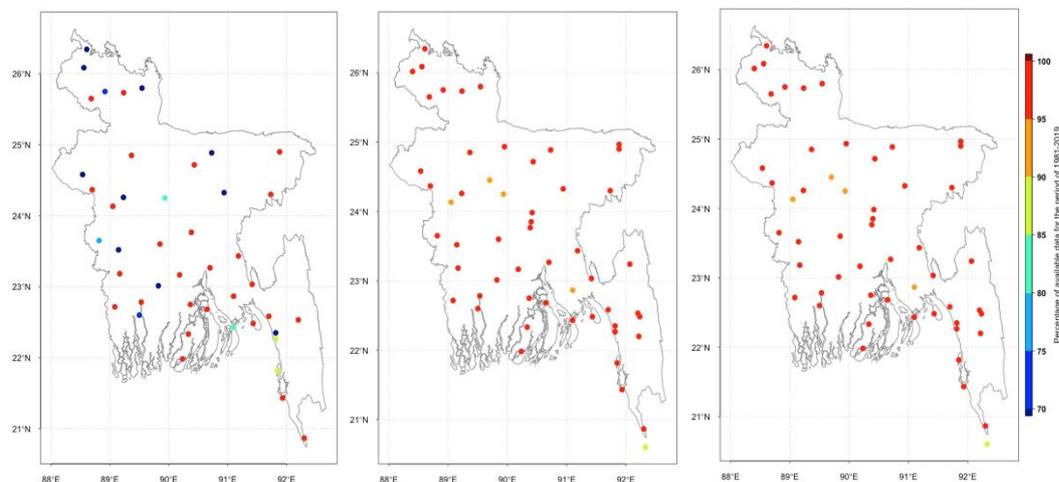


Figure 2. Percentage of available daily data of (a) rainfall (b) minimum and (c) maximum temperature data for the period 1981-2019.

Figure 3. shows the annual (1981-2019) average number of stations reporting rainfall and temperature observations. Red bars in Figure 3 represent the yearly minimum and maximum number of days with observations. It can be seen that the average number of stations with rainfall data gradually increases until 2000 and then stabilized until 2018, after which 8 more stations were added in 2019. For temperature data, all 54 stations report regularly data until 2018, but the total number decreased in 2019.

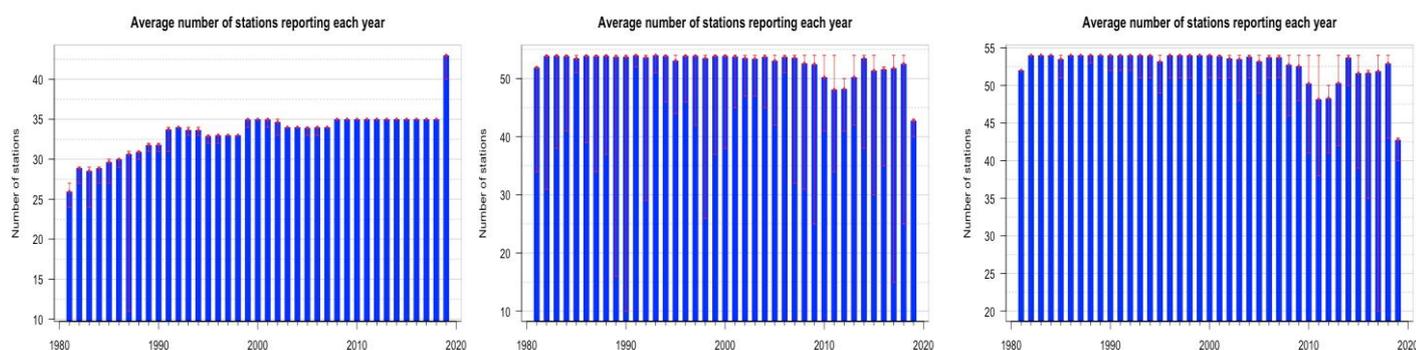


Figure 3. Annual average number of stations reporting observations of rainfall (a) rainfall (b) minimum and (c) maximum temperature for the period 1981-2019 Red bars represent the minimum and the maximum number of stations with observations for each year.

## 2.2 Quality control

As it was mentioned, the data quality control was performed using CDT. The typical workflow for quality control is explained as follows:

- Verification of in-situ station's geographical coordinates.

- For daily rainfall data, the second step corresponds to checking the false zeros for a given month. A false zero is defined as an abnormal observed zero values for the whole month during the rainy season. The main source of the false zeros is the ambiguity in coding the days with no observations (missing values) and days without rain (zero values) when digitizing the paper records. In these cases, data entry operators leave empty the input for both no observations and no rain.
- The next step is to check the presence of outliers in the time series. Outliers can be checked for both temporal and spatial cases.
  - Temporal check is performed for each month to ensure that each observed value is consistent with the climatology for each station. All values detected by the test are flagged as suspicious or outliers and must be checked very carefully.
  - A spatial check is performed by comparing each value of a given station with the values of neighboring stations for the same date.
  - Internal consistency checks for temperature data. In other words, checking if minimum temperature is greater than maximum temperature.
- Erroneous values and outliers were consequently identified. However, undetected observation errors are always possible even using objective methods. Since we did not have the original observations from paper version for a manual verification, the erroneous values were replaced by missing values. All days with precipitation greater than 600 mm were eliminated. For temperature data, whenever possible the outliers were replaced with the estimated data from neighbors' stations. Removing valid extreme values can cause errors in the merged data as easily as keeping erroneous extreme values.

An example of the spatial check results for rainfall and maximum temperature is presented in figure 4. The station in red reported extremely high daily rainfall and relatively low maximum temperature, which are very different compared to values recorded by the neighbor stations.

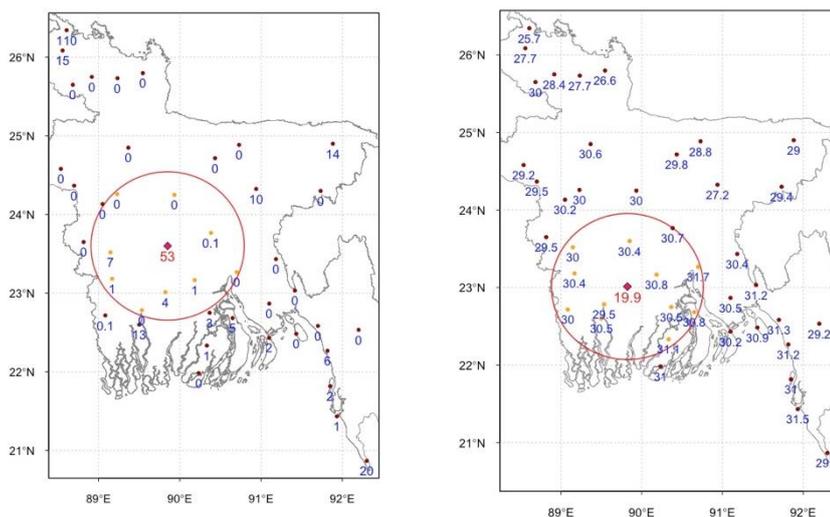


Figure 4: Example of spatial checking of (a) rainfall for July 21, 2019 and (b) maximum temperature for December 06, 2019.

### 2.3 Merging stations with gridded data

For constructing the gridded rainfall product, BMD stations data merged using satellite rainfall estimates from the Climate Hazard Group InfraRed Precipitation (CHIRP), developed by the Santa Barbara Climate Hazards Group at the University of California in association the U.S. Geological Survey Earth Resources Observation and Science Center. CHIRP is a last-generation satellite-only high temporal (daily, pentad and dekadal) and spatial resolution ( $0.05^\circ \times 0.05^\circ$ ) rainfall product with a quasi-global coverage ( $50^\circ\text{S}$ – $50^\circ\text{N}$ ), with rainfall timeseries available since 1981 to the present. CHIRP precipitation is generated from satellite thermal infrared (TIR) measurements with mean bias removed using a satellite-improved station-based climatology. The detailed procedure of the generation of CHIRP can be found in Funk et al (2014) and data are available at <ftp://chg-ftpout.geog.ucsb.edu/pub/org/chg/products/CHIRP/>. Similar, the main source of temperature data in Bangladesh are BMD stations. However, since the stations network is sparse over some regions, temperature values for non-covered locations have to be computed based on the neighboring stations when necessary. Therefore, the use of

reanalysis data as a proxy can potentially solve this problem. Among the main advantages of reanalysis products it can be mentioned: (1) they provide a multivariate, spatially complete, and uniform record of multiple atmospheric variables; (2) they are produced using a single version of a data assimilation system and therefore they are not affected by changes in methods, which makes them suitable for studies of longer-term climate variability. In the present work, the JRA-55 reanalysis (Japanese 55-year Reanalysis) developed by Japan Meteorological Agency is used to generate the temperature datasets, which is among the most recent long-term reanalysis projects. It adopts a relatively high-resolution (~55 km) atmospheric model and uses state-of-the-art assimilation techniques. The dataset extends from 1958 to the present with 3 hourly temporal resolution. Both the data and the documentation are available at [https://jra.kishou.go.jp/JRA-55/index\\_en.html](https://jra.kishou.go.jp/JRA-55/index_en.html). These products were selected because it spans a long time period, over 30 years, their spatial and temporal resolution, and their availability in near-real-time.

The steps considered for merging stations and gridded data include: 1) spatial downscaling of the reanalysis for the temperature data, 2) bias correction of downscaled temperature and satellite rainfall estimates, and 3) finally merging the station data with the bias-corrected gridded data. There are many ways to correct specific errors and artifacts in the gridded data. They can be corrected regardless of the source of the error using ground truth measurements such as station data. In this sense, CDT has 4 methods of bias correction which are Multiplicative Bias Time Step Variable (MBTSV), Multiplicative Bias for Each Month (MBEM), Quantile Mapping with Fitted (QMF) Distribution and Quantile Mapping with Empirical (QME) Distribution and 4 methods of merging, namely the Simple Bias Adjustment (SBA), Regression Kriging (RK), Cressman Scheme (CSc) and Barnes Scheme (BSc). Similar, CDT has 4 interpolation methods: Inverse Distance Weighted (IDW), Ordinary Kriging (Kriging), Modified Shepard (Shepard) and Spheremap. We performed several runs using all possible combinations of the merging and interpolation methods to choose the best method of merging and interpolation in order to generate ENACTS-BMD.

The workflow used to generate the final gridded products for rainfall and temperature time series is as follows:

- **Downscale:** Since reanalysis data from JRA-55 have a spatial resolution of approximately 55 km, it is therefore necessary to downscale the temperatures data from reanalysis to meet the rainfall grid ( $0.05^\circ \times 0.05^\circ$ ). As the temperature depends on altitude, a constant lapse rate (vertical temperature gradient) is used for each month using station temperature data and elevation data from a digital elevation model. Then these coefficients are applied to the reanalysis data using a linear model. A bilinear method is used for the interpolation.
- **Bias correction:** The bias coefficients are computed and interpolated into the grid of the satellite rainfall estimates by using an inverse distance weighted interpolation. To interpolate a grid point, at least 3 and at most 9 stations within a radius of influence of 1.0 degree for rainfall and 1.5 degree for temperature are used. To correct the bias, the satellite rainfall estimates and downscaled reanalysis temperature data are multiplied by the interpolated bias coefficients.
- **Merging:** To merge the station data with the bias-corrected satellite rainfall estimates and downscaled reanalysis temperature, a nested merging method was used. The "Regression Kriging" merging method using Modified Shepard interpolation was used to generate the merged data for the daily and dekadal time scale. The merged data are generated using three passes, the bias-corrected data are used as the first guess.

Figure 5 and 6 summarizes the steps of the merging process for rainfall data and temperature, respectively.

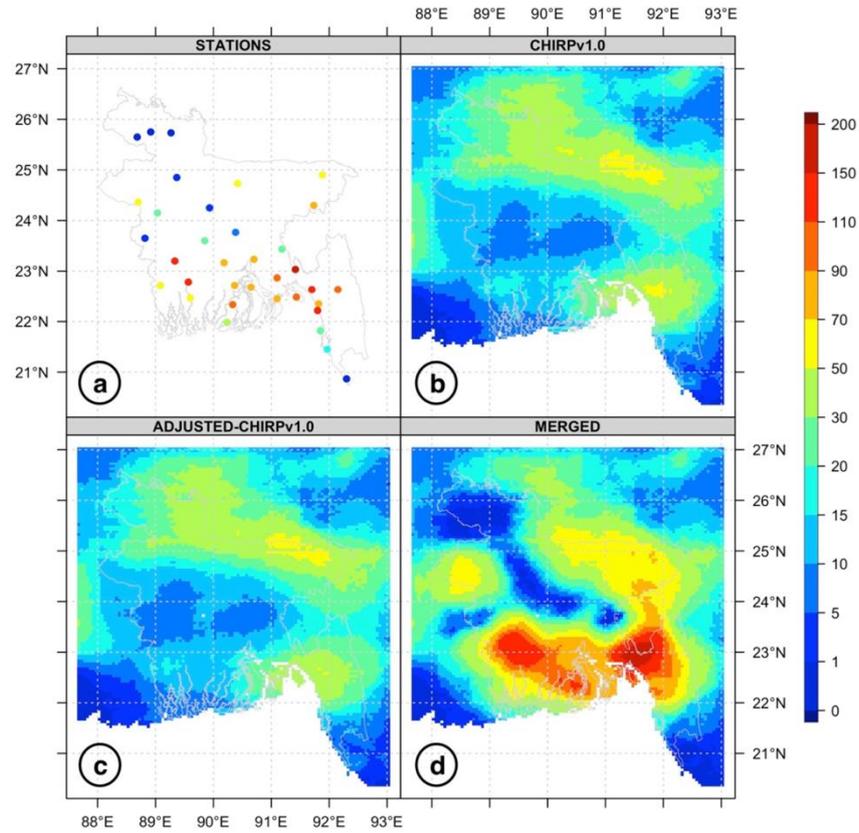


Figure 5: Steps of the merging process for rainfall data: (a) station data, (b) satellite rainfall estimates CHIRP v1.0, (c) bias-corrected CHIRP, and (d) merged data: a combination of (a) and (c).

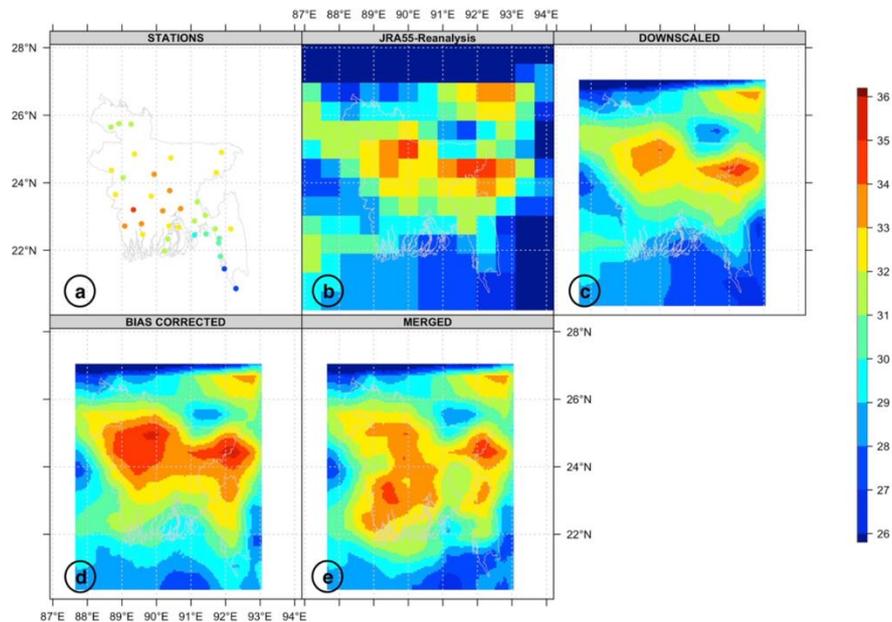


Figure 6: Steps of the merging process for temperature data: (a) station data, (b) JRA55 reanalysis, (c) downscaled JRA55, (d) bias-corrected data, and (e) merged data: a combination of (a) and (d).

### 3. Technical Validation of ENACTS-BMD and comparison with other products

A meticulous validation of ENACTS-BMD data against observations from BMD stations was carried out using CDT's leave-one-out cross-validation (LOOCV). For this, LOOCV takes the total  $n$  observations and one of them is set aside for validation, and then the merging is performed with the remaining  $n-1$  observations. In a second step, the merged value at the discarded station is extracted. This is repeated  $n$  times, leaving out each observation at the end of the procedure. The evaluation is performed daily and dekadal data. For daily data, time series from 2005 to 2018 are used both for rainfall and temperature data and from 1991 to 2018 for dekadal data. A point-to-grid analysis was performed to compare and validate the performance of the gridded data in relation to station observations. Grid values of the gridded data containing the station locations were extracted.

The scatter plots of rainfall, minimum and maximum temperature between stations and ENACTS-BMD results extracted at the station locations for daily and dekadal are shown in figures 7 and 8, respectively. It is observed that in general, the association between ENACTS-BMD rainfall (temperature) and stations data is moderate-to-strong for both daily and dekadal data.

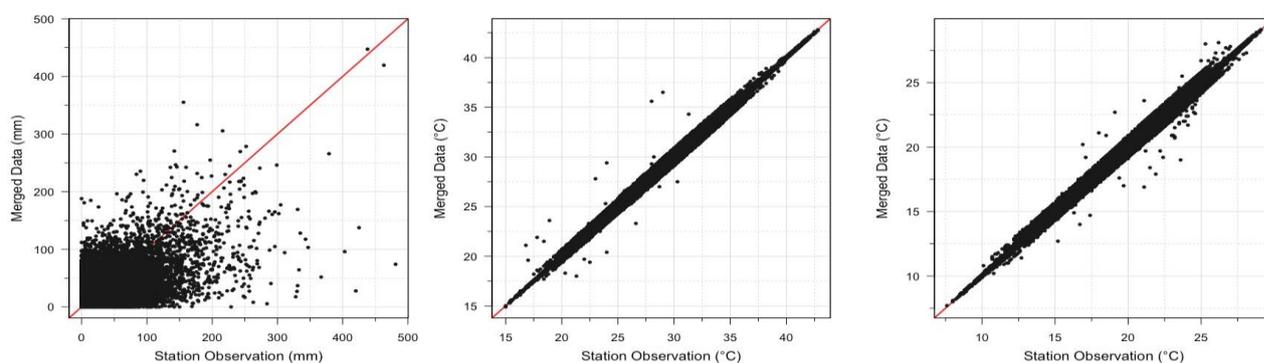


Figure 7: Scatter plots of (a) daily rainfall, (b) daily minimum (c) maximum temperature from stations data versus ENACTS-BMD for the period 2005 to 2018.

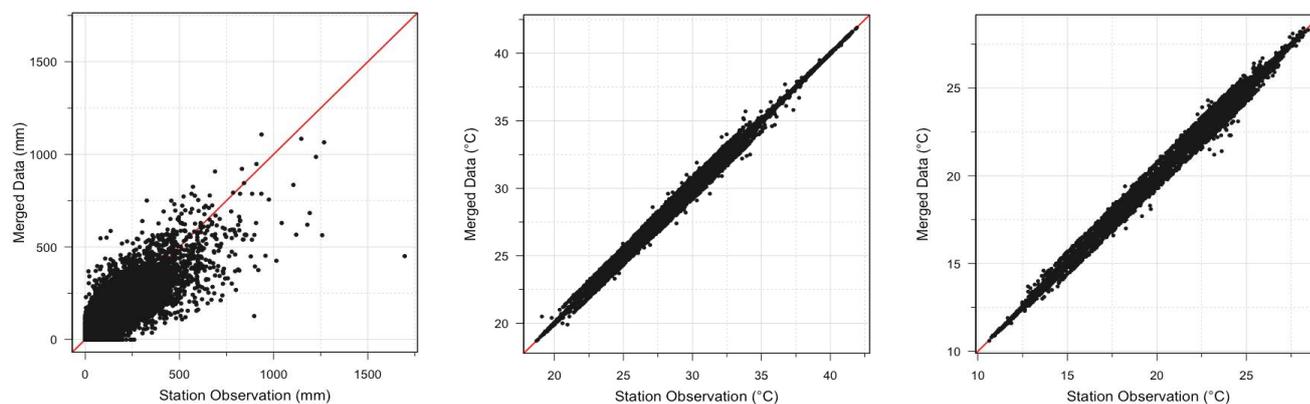


Figure 8: Same as figure 7 but for dekadal rainfall for the period 1991-2018.

The performance of ENACTS-BMD was also assessed by computing a set of skill scores. These scores are described on <https://www.cawcr.gov.au/projects/verification/>, and implemented within CDT. The performance of rainfall results was evaluated using categorical and continuous verification metrics considering the fact that rainfall has a mixed nature as both a categorical (occurrence/non-occurrence) and continuous (amount) process. The evaluation method includes pairwise comparison continuous statistics to assess the performance of satellite rainfall estimates in terms of amount, as well as categorical statistics used to assess rain or no rain events detection capabilities. For categorical statistics, a threshold value of 1.0 mm is used both for daily and dekadal data. For the ENACTS-BMD temperature data, the evaluation is performed using only continuous statistical indicators. Although CDT has an option of a long list of skill scores for performance assessment, for the brevity we only

discuss seven of them: five continuous and two categorical for daily and dekadal rainfall and temperature data. The five continuous skill scores include (1) Pearson's correlation coefficient (CORR), varies between plus-minus one, a value of 1 is the perfect score; (2) Nash-Sutcliffe efficiency coefficient (NSE), varies from minus infinity to one, a value of 1 is the perfect score; (3) multiplicative bias (BIAS), a value greater than 1 indicates an overestimation while a value less than 1 indicates an underestimation; (4) the mean absolute error (MAE), a value of 0 as a perfect score, MAE is less sensitive to outliers; (5) RMSE measures the average magnitude of the error, a low RMSE value indicates larger central trends and a smaller extreme errors. The two categorical skill scores include the probability of detection (POD), and false alarm ratio (FAR), which are computed based on a contingency table in order to assess the successes and failures of detecting the occurrence of rain. POD indicates the fraction of rainy events from station observations detected correctly by the merged data, it varies between 0 (no detection) and 1 (perfect detection) whereas FAR corresponds to the portion of events identified by the merged data but not confirmed by stations observations, a value of 0 is the perfect score.

Results of the above-presented skill scores for ENACTS-BMD dataset with respect to the BMD-station data for all the stations at daily and dekadal scale are presented in Table 1. The spatial distribution of each skill score also calculated for each station location. The Bias and CORR for rainfall and minimum temperature are shown in figure 9 and 10 respectively.

	Daily			Dekadal		
	Rainfall	Min.T	Max.T	Rainfall	Min.T	Max.T
CORR	0.681	0.999	0.999	0.867	0.999	0.999
NSE	0.455	0.999	0.999	0.75	0.998	0.998
BIAS	0.92	1	1	1.004	1	1
MAE	5.308	0.049	0.049	27.844	0.067	0.065
RMSE	13.976	0.145	0.145	52.1317	0.148	0.147
POD	0.803	NA	NA	0.955	NA	NA
FAR	0.327	NA	NA	0.134	NA	NA

Table1. Skills score for the ENACTS-BMD with respect to the station dataset for all the station's location. As POD and FAR are categorical skill score, it is Not Applicable (NA) for temperature data.

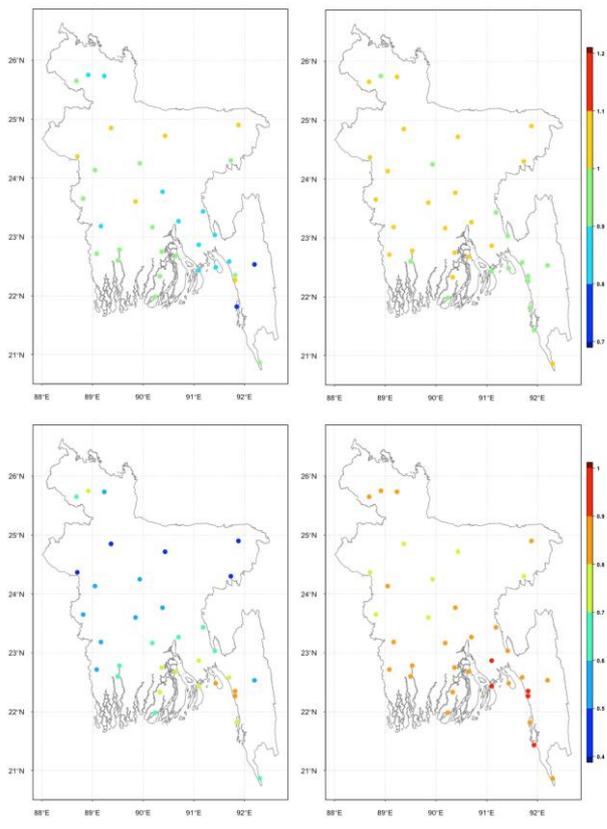


Figure 9: Spatial distribution of the bias of (a;b) and the correlation coefficients (c;d) for the daily and dekadal rainfall respectively.

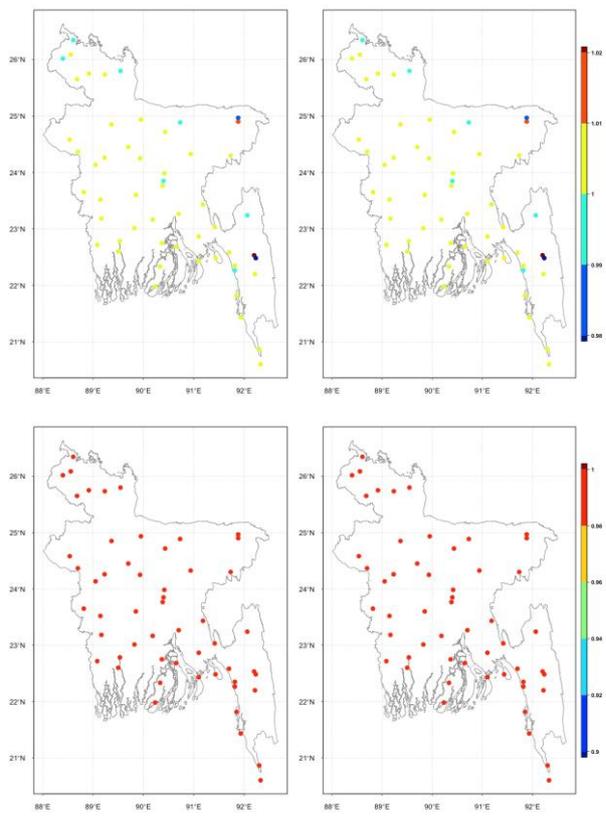


Figure 10: Same as figure 9 but for minimum temperature.

#### 4. Data archive

Since February 2020, BMD hosted the ENACTS-BMD dataset for rainfall and maximum and minimum temperature at  $0.05^\circ \times 0.05^\circ$  resolution on their website <http://datalibrary.bmd.gov.bd/>. Its record begins in January 1981 and is ongoing with a temporal resolution of daily, decadal and monthly. ENACTS-BMD updates every month for rainfall and every other six months for temperature. Users can easily download this dataset in various data formats such as NetCDF, CSV, GrADS format, among others.

#### Acknowledgments

This research was supported and funded by ACToday, the first of Columbia University's World Projects. We acknowledge Santa Barbara Climate Hazards Group at the University of California and the Japan Meteorological Agency for creating, updating, and maintaining the satellite rainfall product (CHIRP) and reanalysis temperature product (JRA-55).

#### References

1. Maidment, R., Grimes, D., Black, E. *et al.* 2017: A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. *Sci Data* **4**, 170063 <https://doi.org/10.1038/sdata.2017.63>.
2. Nashwan, M.S., S. Shahid and X. Wang, 2019: Uncertainty in estimated trends using gridded rainfall data: a case study of Bangladesh. *Water*, **11**, 349.
3. Dinku, T., M. C. Thomson, R. Cousin, J. del Corral, P. Ceccato, J. Hansen and S. J. Connor, 2017: Enhancing National Climate Services (ENACTS) for development in Africa. *Clim. Develop.*, DOI: 10.1080/17565529.2017.1405784.
4. C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, J. Rowland, B. Romero, G. Husak, J. Michaelsen, A. Verdin. A quasi-global precipitation time series for drought monitoring Technical Report, U.S Geological Survey Data Series, 832 (2014), p. 4.
5. Kobayashi, S., Y. Ota, Y. Harada, A. Ebata, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, 93, 5-48, doi:10.2151/jmsj.2015-001.
6. Harada, Y., H. Kamahori, C. Kobayashi, H. Endo, S. Kobayashi, Y. Ota, H. Onoda, K. Onogi, K. Miyaoka, and K. Takahashi, 2016: The JRA-55 Reanalysis: Representation of atmospheric circulation and climate variability, *J. Meteor. Soc. Japan*, 94, 269-302, doi:10.2151/jmsj.2016-015.