scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Developing a high-resolution gridded rainfall product for Bangladesh during 1901–2018

Ashraf Dewan[1], Shamsuddin Shahid[2], Md. Hanif Bhuian[3], Shaikh M. Jobayed Hossain[4], Mohamed Salem Nashwan[5], Eun-Sung Chung[6]✉, Quazi K. Hassan[7] & Md Asaduzzaman[8]

A high-resolution (1 km × 1 km) monthly gridded rainfall data product during 1901–2018, named Bangladesh Gridded Rainfall (BDGR), was developed in this study. *In-situ* rainfall observations retrieved from a number of sources, including national organizations and undigitized data from the colonial era, were used. Leave-one-out cross-validation was used to assess product's ability to capture spatial and temporal variability. The results revealed spatial variability of the percentage bias (PBIAS) in the range of −2 to 2%, normalized root mean square error (NRMSE) <20%, and correlation coefficient (R^2) >0.88 at most of the locations. The temporal variability in mean PBIAS for 1901–2018 was in the range of −4.5 to 4.3%, NRMSE between 9 and 19% and R^2 in the range of 0.87 to 0.95. The BDGR also showed its capability in replicating temporal patterns and trends of observed rainfall with greater accuracy. The product can provide reliable insights regarding various hydrometeorological issues, including historical floods, droughts, and groundwater recharge for a well-recognized global climate hotspot, Bangladesh.

Background

Located in a tropical monsoonal climatic region, Bangladesh receives nearly 2200 mm of rainfall every year, which supports agriculture, the environment and livelihood activities since time immemorial[1–3]. However, the country is susceptible to hydrometeorological hazards due to its flat topography and high rainfall seasonality[4,5]. For instance, nearly 70% of total rainfall occurs in the monsoon season (June to September), and <3% takes place in the dry season (December to February)[6]. Any change in rainfall patterns in terms of deficit/surplus or even a subtle shift can lead to climatic extremes such as drought and floods. Ozaki[7] showed that the country experiences economic damage of US$ 2.2 billion, equivalent to 1.5% of gross domestic product (GDP) during an abnormal flood year. It experiences a reduction in crop production of 20 to 30% in a drought year[8]. Nearly 40% of total crops in the country are rain-fed, and the rest depend on groundwater irrigation[9], meaning that rainfall variability, especially monsoonal amount (1500 mm), severely affects rain-fed agriculture, which has implications for food security due to its large population. It is also a primary source of groundwater replenishment. For instance, a moderate rainfall deficit causes a decline in groundwater in subsequent years[10], causing an increase in irrigation costs, reducing farmers' profit, and overwhelming a vast majority of the rural population[11]. Since large river networks of the country are fed by Himalayan snowmelt and local rainfall, high rainfall evidently affects riverbank erosion[12], while low rainfall reduces freshwater flow towards coastal areas, causing salinity ingress further inland[13] and affecting the livelihoods of millions.

With the increased concern of anthropogenic climate change, the demand for high-resolution gridded datasets of climate variables such as rainfall has increased over the last few decades[14]. Consequently, a number of daily and monthly gridded products have been developed at global[15–17], regional[14,18,19] and national scales[20–23]

[1]Spatial Sciences discipline, Curtin University, Bentley 6102, Perth, Western Australia. [2]School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), 81310, Johor Bahru, Malaysia. [3]Department of Geography and Environment, Jagannath University, Dhaka, 1100, Bangladesh. [4]Meiji University, Chiyoda city, Tokyo, 101-8301, Japan. [5]Construction and Building Engineering Department, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), 2033, Cairo, Egypt. [6]Department of Civil Engineering, Seoul National University of Science and Technology, Nowon-gu, 01811, Seoul, South Korea. [7]Department of Geomatics Engineering, University of Calgary, 2500 University Drive NW, T2N 1N4, Calgary, Canada. [8]Department of Engineering, School of Digital, Technologies and Arts, Staffordshire University, Stoke-on-Trent, UK. ✉e-mail: eschung@seoultech.ac.kr

with spatial resolutions varying from 0.01 to 1.0°. Several interpolation methods, such as kriging[24], angular distance weighting[25], minimum surface curvature[26], thin plate splines[27], and inverse distance weighting[28], are utilized. Although global and regional gridded rainfall products have enhanced our knowledge of extreme events, determining their space–time variability and pattern requires high-resolution datasets[21]. Thus, existing global- or regional-scale products may not support the evaluation of local-scale extreme events and spatiotemporal variability[22]. Fine spatial resolution (~0.01°) is, therefore, a prerequisite to capture environmental variability[17]. Furthermore, high-resolution data can be instrumental to examine climate–impact studies on an area[23].

While increases in near-surface air temperature have been well documented for Bangladesh[6], lack of long–term high–resolution rainfall data was an important constraint to determine changes in spatiotemporal rainfall pattern. A dense rainfall monitoring network is needed to assess the spatiotemporal variability of hydrometeorological hazards for developing mitigation measures. It is also important to examine how increased anthropogenic activities resulting from rapid urban expansion, forest loss associated with land use/land cover changes, alteration of fluvial morphology and ever-increasing populations influence regional climate. However, such studies in the country are few and far between. The unavailability of long-term rainfall data is thought to have contributed to the nonexistence of a detailed regional-scale analysis of climatic conditions such as the spatiotemporal pattern of rainfall variability.

The Bangladesh Meteorological Department (BMD) has 42 rainfall observation stations, but their distribution is not uniform across the country (http://live.bmd.gov.bd/). The majority of these stations were installed in the 1960s, and therefore, data are available only since 1960s. The number of stations is also far below (one station per 3428 km^2) the recommended number by the World Meteorological Organization (WMO), which is one station per 10−20 km^2 for climatic studies[29]. However, almost all studies employ BMD data and appear to provide coarse spatial detail regarding rainfall climatology and their trend for the country. Compared to BMD, the Bangladesh Water Development Board (BWDB) maintains an extensive rainfall monitoring station. It currently has 293 rain gauges over 130,170 km^2 of land area (www.ffwc.gov.bd). Combining rainfall records of BWBD and BMD provides good coverage over the country, but the inhomogeneous distribution of stations and unavailable/missing data during the last century inhibit long-term regional assessment.

Then, East Bengal (currently Bangladesh) used to experience frequent floods, droughts and cyclones; thus, rainfall data were indispensable to better prepare against these hazards and to reduce the loss of lives and property during the British Raj. The Public Works Department (PWD) during the colonial government therefore installed rainfall monitoring stations in the region. The monthly rainfall records of those stations were later published by the East Pakistan (present Bangladesh) Water and Power Development Authority in two separate volumes of water supply papers (WSP)[30,31]. The undigitized rainfall information from 1901 to 1959 is of high value and is a reliable source to examine the rainfall distribution of the region since 1901. However, rainfall data during that period have not been utilized in any of the studies due to limited access and/or their unavailability in digital form. This work was motivated to harness the power of digital technologies by which more than a hundred years of rainfall data were encoded to generate a high-resolution gridded product.

It is obviously a major challenge to develop a rainfall product, especially when the data were recorded by different instruments and maintained by various origins. To make them useable, proper quality control and homogenization are essential. Furthermore, selecting a suitable interpolation method to generate gridded dataset with station data is also important. This study presented the development of a fine-resolution rainfall product by using multiple sources of data. It is expected that the product would be of great value to advance climatic and hydrological studies of Bangladesh, a global hotspot of anthropogenic climate change.

Geography and precipitation climatology of Bangladesh. Bangladesh, located between 20°34′–26°38′ N and 88°01′–92°41′ E, covers an area of 147,570 km^2. Approximately 90% of the land is a deltaic plain, known as the lower Gangetic plain. The elevation of the country varies from 0 m in the south to approximately 110 m in the extreme north[32]. Low hills and highlands in the southeast and northeast cover only 10% of the total land. The maximum elevation of the hilly regions reaches 1010 m. The geographical position of the country in the three mighty rivers, viz., the Ganges, Brahmaputra and Meghna (Fig. 1), made it highly prone to recurrent floods.

Rainfall of the country is controlled by differential heating of land and sea surfaces. Indian landmass heated rapidly compared to its surrounding oceans, introducing a strong thermal contrast between land and sea[33]. Therefore, air flows from the oceans towards heated land. The moist air from the sea enters Bangladesh from the south and causes high monsoon rainfall from June to September. A reveres situation occurs during the winter season from December to February, and air flows from land to the sea[34,35]. The air flowing over the landmass is dry and thus experiences almost no rainfall during winter. Between summer and winter, there are two transitional periods, post-monsoon (October–December) and premonsoon (March–May). While the former is relatively cool, the latter is hot with occasional downpour.

The monsoon of Bangladesh flows in two branches, one of which strikes western India, and the other travels up to the Bay of Bengal before entering Bangladesh via the southeast. The monsoon from the Bay of Bengal crosses the plain to the north and northeast before being turned to the west and northwest by the foothills of the Himalayas[6,36]. As monsoon air moves farther inland, its moisture content decreases, resulting in decreased rainfall towards the northwest and west of the country[37]. However, the additional uplifting effect of the Meghalaya Plateau increases rainfall to the northeast. The premonsoon is a transitional season between the northerly circulation of winter and the southerly circulation of the monsoon. Thunderstorms are a major source of premonsoon rainfall[38,39]. The precipitation mechanism indicates that topography and distance to coast can also have a large influence on its spatial distribution and seasonality.

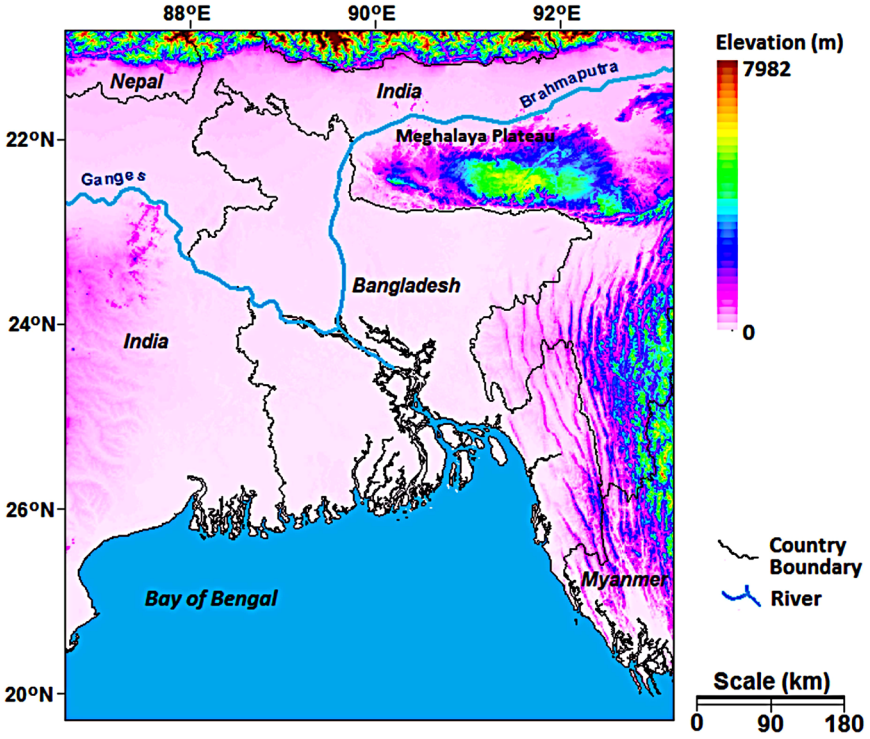

Fig. 1 Location and topography of Bangladesh. Bangladesh, located between 20°34′–26°38′ N and 88°01′–92°41′ E, covers an area of 147,570 km^2. The elevation of the country varies from 0 m in the south to approximately 110 m in the extreme north. Low hills and highlands in the southeast and northeast cover only 10% of the total land.

Data acquisition. Monthly rainfall records were acquired from three sources. They were:

(i) BMD: daily rainfall records of 33 stations were collected. Data for nearly 52% stations were available after the 1960s, although only three stations had data since 1901.
(ii) BWDB: daily rainfall records of 293 stations between 1961 and 2018 were obtained and utilized in this work. Note that data were available for 1961–2018 at 52% stations, while others provided data during different temporal windows of 1961–2018.
(iii) WSP: The WSP rainfall records were available for 124 locations from 1901 to 1959[30]. These records were available in printed form, and the digitizing process includes editing and proofreading before publication, as in other government gazetteer reports. They were encoded with caution and cross-checked several times to ensure high fidelity. However, only 45% of stations contain records from 1901 to 1959. The rest of the stations have data for different temporal resolutions of 1901–1959.

Attributes of 124 stations (e.g., name and locations) are common for both BWDB and WSP, meaning that 124 stations have data for different temporal windows of 1901 to 2018. Compilation of a database using rainfall data from all sources provides records of a total of 326 stations, crisscrossing over Bangladesh (Fig. 2). Furthermore, a more or less homogeneous distribution of rain gauges can be seen (Fig. 2), except for southwestern dense and southeastern forested mountainous regions.

Daily data were aggregated to generate monthly rainfall time series. The general rules in aggregating daily rainfall to monthly rainfall differ for monsoon and nonmonsoon months. The monsoon month's rainfall was considered unavailable if rainfall in a single day was missing, while rainfall of a nonmonsoon month was considered missing if records for three consecutive days were missing[40]. WSP data were available only at the monthly scale, whereas BMD and BWBD had daily data. To bring them in a common temporal resolution, data were aggregated to monthly time series using the rule–based procedures noted above.

Stations having observations for at least 25 years with random missing values were included in the present study. Short-term records (<25 years) were removed to avoid noise. Few duplicates were identified due to changes in station names or locations in the government gazetteers. In addition, few stations were adjacent to each other. This was due to a slight change in the installation locations after decommissioning the previous station. Duplicates were removed, and stations in close proximity (<500 m) were averaged. In the case of a locational change, the station was given a separate name. Thus, the new and old stations were considered as a single station. The station's new location is used for processing data recorded after locational shift.

Figure 3 shows the availability of data from different sources during 1901–2018. Missing data over the study period range between 3.2% and 76.1% for different stations. The percent of missing records at different stations is shown in Supplementary Figure S1. The number of stations with different percentages of missing data during

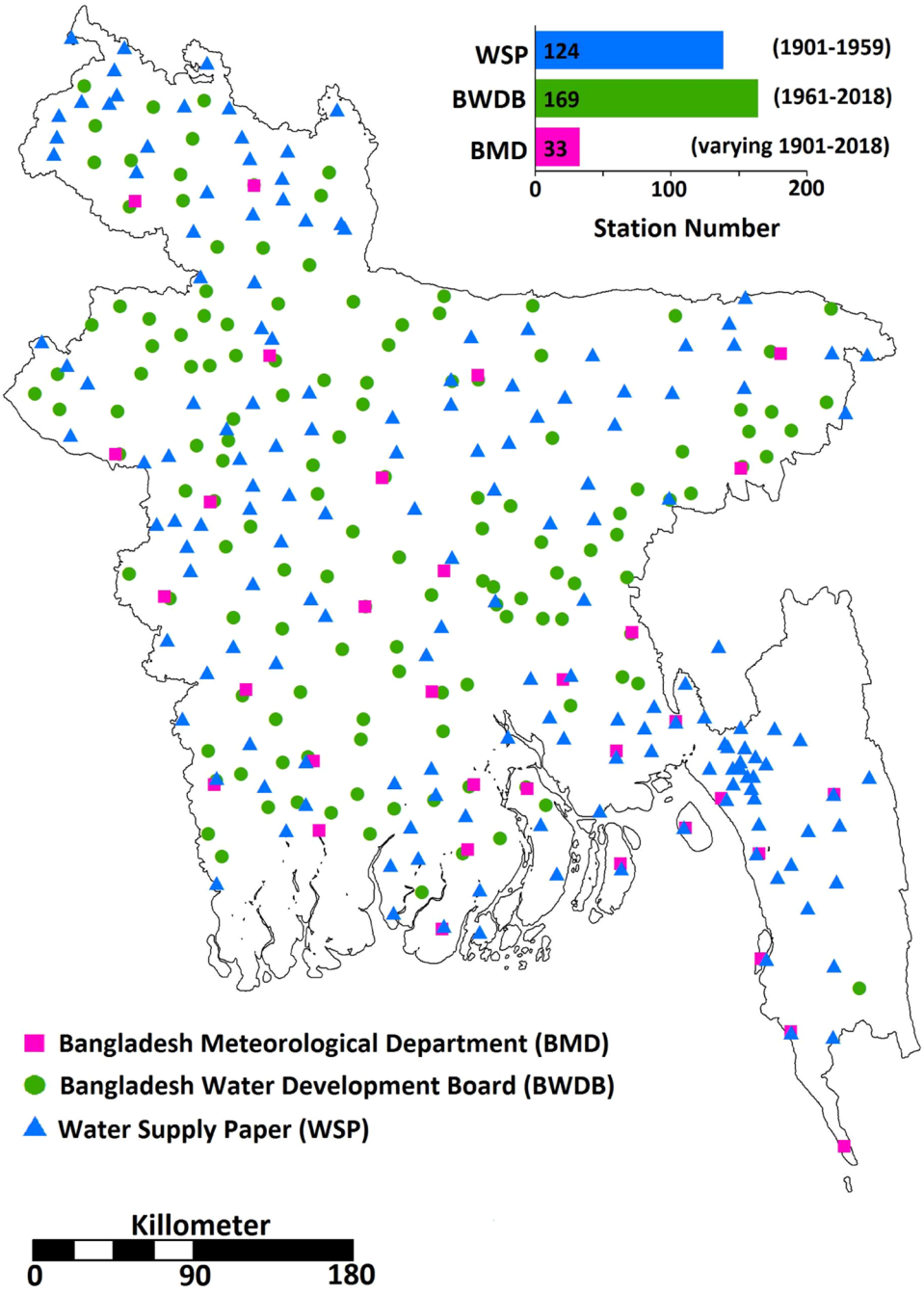

Fig. 2 Distribution of rainfall stations and their records. Monthly rainfall records were acquired from three sources: 33 stations from BMD for different periods, 293 stations from BWDB for different windows between 1961 and 2018, 124 stations from WSP for different windows between 1901 and 1959. Attributes of 124 stations (e.g., name and locations) are common for both BWDB and WSP, meaning that 124 stations have data for different temporal windows from 1901 to 2018. Compilation of a database using rainfall data from all sources provides records of a total of 326 stations crisscrossing over Bangladesh.

1901–2018 is shown in Fig. 4, which shows that the highest number of stations (138) having missing rainfall is in the range of 50 to 60%, and 66 stations had more than 60% missing records.

The Precipitation Bias CORrection (PBCOR)[41] data was used to evaluate the rainfall climatology of the BDGR product. PBCOR is a global high-resolution (0.05°) precipitation climatology bias-corrected using streamflow. It consists of bias-corrected precipitation climatologies of WorldClim V2 for the period 1970–2000[42], Climatologies at High Resolution for the Earth's Land Surface Areas (CHELSA V1.2) for 1979–2013[43], and Climate Hazards Group Precipitation Climatology (CHPclim V1) for 1980–2009[44]. These WorldClim, CHELSA and CHPclim climatologies underestimate precipitation in the broader Himalayas region, including

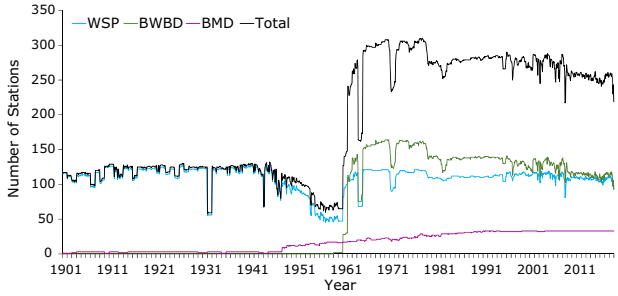

Fig. 3 Availability of data from different sources. Number of available meteorological stations per year from different sources during 1901–2018.

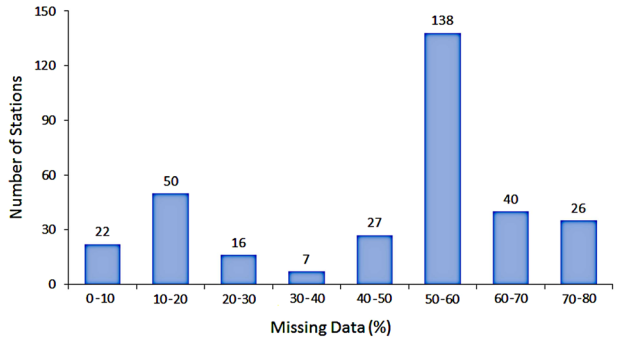

Fig. 4 Missing records (%) during 1901–2018. The number of stations with different percentages of missing data during 1901–2018. The highest number of stations (138) had missing rainfall in the range of 50 to 60%, and 66 stations had more than 60% missing records.

some parts of Bangladesh[45]. PBCOR corrected this bias to develop PBCOR WorldClim, PBCOR CHELSA and PBCOR CHPclim to provide better estimates of the precipitation climatologies.

Methodology. *Gap filling.* In this study, multivariate imputation by chained equations (MICE) was used to fill in missing rainfall data. The MICE was implemented considering missing data at all stations in the study area. The MICE method estimates missing values of rainfall data from a station using rainfall data from its nearby stations. In the first step, the MICE replaces missing values for each station of the set with the mean rainfall of the station as initial estimates. It then sets back the first estimated missing value as missing while keeping the other imputed values as original values and regresses them. The rest of the imputed missing values were regressed in the same way. The regressed values were then subtracted from the initial set of data, which gives errors for missing value estimates. Further steps generate errors in the same way by setting imputed values as missing one-by-one, regressing them and subtracting them from the previously generated sets of data. The process was repeated until the change in estimated error in predicting missing data is not significant. In the present study, the process was repeated five times, as no noticeable improvement in the error matrix was found when multiple imputations were more than five times. The preliminary analysis of filling in missing data of a station was performed using data from 4 to 10 nearby stations. The results showed no improvement in prediction error if more than six nearby stations were used. Therefore, rainfall data from six nearby stations were used to fill in the missing data of each station.

The mice package[46] of R program was used in this study to fill in missing data. Different classical models, for instance, linear regression, logistic, Poisson, polytomous regression, and random forest, can be used in MICE for imputation model development. This study evaluated the performance of different methods available in the mice package in R and found the best performance using 'random forest' for the number of trees (ntree) equal to fifty.

The performance of MICE models was evaluated for filling both continuous and random missing data. The results showed poor performance of the method in filling in continuous missing data for more than four years. It should be noted that no data were available for 184 stations before 1961. Some of these stations also contain continuous missing data for four years after 1961. Rainfall data with random missing values for these stations were available for 29 to 56 years. Filling nearly 60 years of missing rainfall data with the model developed using 29 to 56 years of data would lead to large uncertainty. Therefore, these 184 stations' data were not filled during 1901–1960. However, random missing data from these stations after 1961 were filled. The missing data at the remaining 142 stations during 1901–2018 were between 3.2 and 54.4%. Continuous missing rainfall data at 70 out of these 142 were noticed for 4 to 26 years, which were also not filled. After leaving the continuous missing rainfall for more than four years, the random missing rainfall data at all stations ranged from 3.2% to 28.8%.

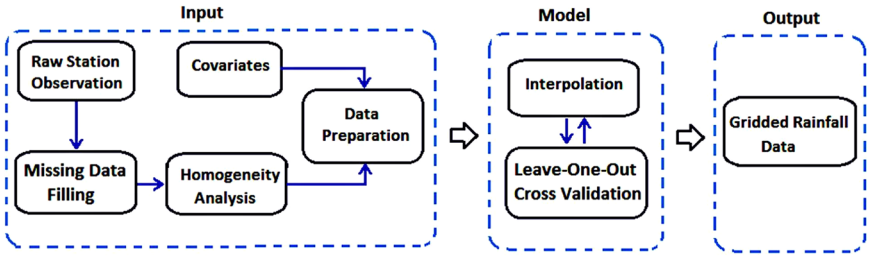

Fig. 5 Flowchart showing the procedures. Flowchart showing the procedures used for generating the BDGR product. Leave-one-out cross-validation (LOOCV) was employed, where only a single observation was used for validation and the remaining observation was used for model training. The process was repeated for N (number of observations) time to cover all observations as a validation sample.

However, at nearly 90% of the stations, the random missing data were below 20%. These random missing data were filled with the MICE model.

Homogenization of the rainfall series. Longer period rainfall series are significantly affected by inhomogeneities because of upgrading the instruments, relocation of stations and measurement error. These factors often result in unexpected jumps or shifts, reflecting outliers in the data series. Therefore, the detection of inhomogeneities and homogenization of climate data is essential[47,48]. In this study, Climatol[49], available in the statistical package R, was used to homogenize the data series. Coll, *et al.*[48] compared the break frequency results of different homogenization methods and reported superior performance of Climatol than the other methods. Skrynyk, *et al.*[50] evaluated the uncertainty associated with Climatol's adjustment algorithm and reported a better capacity of Climatol to remove systematic errors related to jumps in the means. Domonkos, *et al.*[51] also showed Climatol to be effective for homogenizing climate data. Therefore, it has been widely used as an effective climate data homogenization tool in recent years[52,53]. Details of Climatol can be found in Guijarro[54].

Climatol uses a standard normal homogeneity test (SNHT) both on overlapping windows and on the whole rainfall series to detect inhomogeneities. It first uses the whole series to detect the break and then splits the series based on breakpoints to reveal any other breaks in its subseries. Multiple checks provided a better capability of Climatol to manage multiple inhomogeneities. Finally, it fills all missing values using the weighted ratio of data from neighbouring stations to make the series homogeneous by using the reduced major axis model[55], a kind of orthogonal regression, $\hat{y}_i = x_i$, where x_i is the mean (or the weighted mean) rainfall (mm) of nearby stations, and $\hat{y}_i$ is the data (mm) containing missing values. Both x_i and $\hat{y}_i$ are standardized based on available data before developing the regression method.

It should be noted that Climatol fills the missing data with spatial interpolation. However, these filled values were not used in generating BDGR data. Only the random missing data were filled using MICE before homogenization using Climatol. Missing data were filled with Climatol and replaced with NA before using the data for generating BDGR and cross validation.

Product development procedure. Leave-one-out cross-validation (LOOCV) was employed, where only a single observation was used for validation and the remaining observation was used for model training (Fig. 5). This process was repeated for N (number of observations) time to cover all observations as a validation sample. The LOOCV approach required a large computation time. However, it provided error estimation for each station, thus, quantifying errors both spatially and temporarily.

The process was repeated for all months for all years between 1901 and 2018. In each case, outliers or abnormal values were examined using residuals. Nearby station data were checked to examine whether there are any outliers or abnormal values. This was done for every station considered in this work. If the difference in rainfall from all nearby stations was greater than 50% of the mean rainfall of a station, the data are outliers, possibly caused by mechanical faults or human error, and were considered missing.

Thin-plate spline (TPS) interpolation. Observed rainfall data were interpolated using the TPS method[56]. It is a robust technique and is used for developing gridded air temperatures of China[57], 'climate surfaces' of global land areas[17,42], and high–resolution daily precipitation datasets of Europe[58]. Details about TSP can be found in Hutchinson and Xu[56]. In TPS, N rainfall data are used to fit the spline model, and z_i (mm) takes the following formula[56]:

$$z_i = f(x_i) + b^T y_i + e_i \quad (1)$$

where i = 1 … N, f is a smooth function of the spline independent variables x_i (mm), b is a vector of linear coefficients for independent covariates, and y_i and e_i are independent, zero-mean error terms[56]. The models were developed using different combinations of covariates, and the model producing the least error was finally used for interpolation.

Rainfall data for each month of all years between 1901 and 2018 were interpolated separately to generate gridded rainfall surfaces for individual months. All interpolated surfaces are then integrated to produce gridded monthly rainfall time series for the study period over Bangladesh.

Metric Formula	Range	Unit	Optimal Value
$NRMSE = 100 * \frac{\sqrt{\frac{1}{N} * \sum_{i=1}^{N}(S_i - O_i)^2}}{\mu_o}$	0–∞	%	0
$PBIAS = 100 * \frac{\sum_{i=1}^{N}(S_i - O_i)}{O_i}$	−∞–∞	%	0
$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - \mu_o)^2}$	0–1	—	1
$md = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)}{\sum_{i=1}^{n}(\mid S_i - \mu_o \mid + \mid O_i - \mu_o \mid)}$	0–1	—	1
$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\mu_s}{\mu_o} - 1\right)^2 + \left(\frac{\sigma_s / \mu_s}{\sigma_o / \mu_o} - 1\right)^2}$	−1–∞	—	−1
$rSD = \sigma_o / \sigma_s$	−∞–∞	—	1
$SS = \sum_{i=1}^{n} \min(f_s, f_o)$	0–1	—	1

Table 1. Description of statistical metrics used for evaluating data product.

Error metrics. The BDGR product, generated through the TPS model, was validated using the normalized root mean squared error (NRMSE) in %, bias (PBIAS) in %, coefficient of determination (R^2), Kling-Gupta efficiency (KGE), Willmott's modified coefficient of agreement (MD) and Perkins skill score (PSS). Besides, the ratio of standard deviation (rSD) was used to assess the missing rainfall data estimation model. The equation of the metrics, their range and optimum values are provided in Table 1. The NRMSE and PBIAS were widely used metrics to estimate errors in the data product, while R^2 is a universal metric for showing an association between two time series, i.e., observed versus gridded data. The KGE was an integrated metric that measures association, similarity in the variance and mean between two variables[59]. The MD provides appropriate weight to error and differences between two variables for measuring agreement[60]. Therefore, it provided a better estimate of association in the case of extreme values. It was expected that gridded data should replicate the probability distribution of observed rainfall. Therefore, PSS was used, which measures how the probability distribution function (PDF) of two variables matches each other[61]. In Table 1, r is Pearson's correlation; μ and σ represent the mean and standard deviation of gridded (S) and observed (O) rainfall (mm); N is the sample size; and f_o and f_s are PDFs of the observed and gridded rainfall (mm).

Technical Validation

Performance of gap filling. The MICE models were developed by randomly selecting 70 to 95% of the available data and then validating the remaining 5 to 30% per data at each station. The model's performance was evaluated using statistical metrics at all stations and is presented in Fig. 6. The performance was evaluated for 5 to 30% of missing data to show the robustness of the MICE model developed in this study in filling different amounts of missing values. The results showed that the mean bias in 5 to 30% missing rainfall was 0.1 to 0.9, and the NRMSE was between 14.5 and 29.1%. For all cases, the mean R^2 was above 0.86, rSD was near 1 and KGE was above 0.92. In particular, the metric values showed that the performance of the MICE model in missing rainfall of less than 20% was near ideal values. This indicates perfect filling of random missing values of less than 20% at nearly 90% of stations.

Homogenization of rainfall series. Data homogenization requires determination of optimum SNHT values for overlapping windows and for the entire series. In this study, SNHT values were determined through visualization of histograms. Figure 7(a,b) shows the varying SNHT values of 326 stations. Careful observation of the histograms reveals an apparent minimum SNHT value after 28. Therefore, the SNHT value for both the overlapping windows and the whole series was 28.

Figure 7(c) shows a histogram of normalized anomalies used for outlier detection. Anomalous values beyond ±5 standard deviations (SDs) were considered outliers. A total of 972 outliers were observed. All outliers were manually checked and compared with neighbouring station data. Few outliers resulted from data entry; however, most of the values seemed to have legitimate values. After correcting outliers associated with data entry, no values are found beyond ±18 SDs. Therefore, a threshold of ±18 SDs was chosen to incorporate all extremes.

Figure 7(d) shows a correlogram of monthly rainfall between stations. The correlation for stations within a 50 km radius was >0.75 and decreases rapidly after 100 km. A high correlation with nearby stations and gradually a low relationship with increasing distance was expected due to synoptic–scale moist air circulation that influences rainfall in Bangladesh. Overall, the correlogram indicates good quality of the observed data.

Climatol detected a total of 132 breaks in the whole rainfall series (i.e., 326). Some of the breaks can be explained with the aid of literature. For example, Shahid[39] showed abrupt changes in rainfall at some stations between 1973 and 1975. However, most of them could not be explained due to unavailability of metadata. All the breaks were visually inspected and compared with nearby stations. Highly abnormal breaks were discarded, while others were accepted. This study finally detected homogeneity of 322 out of 326 time series. These 322 series were finally employed to develop a high-resolution gridded product for Bangladesh.

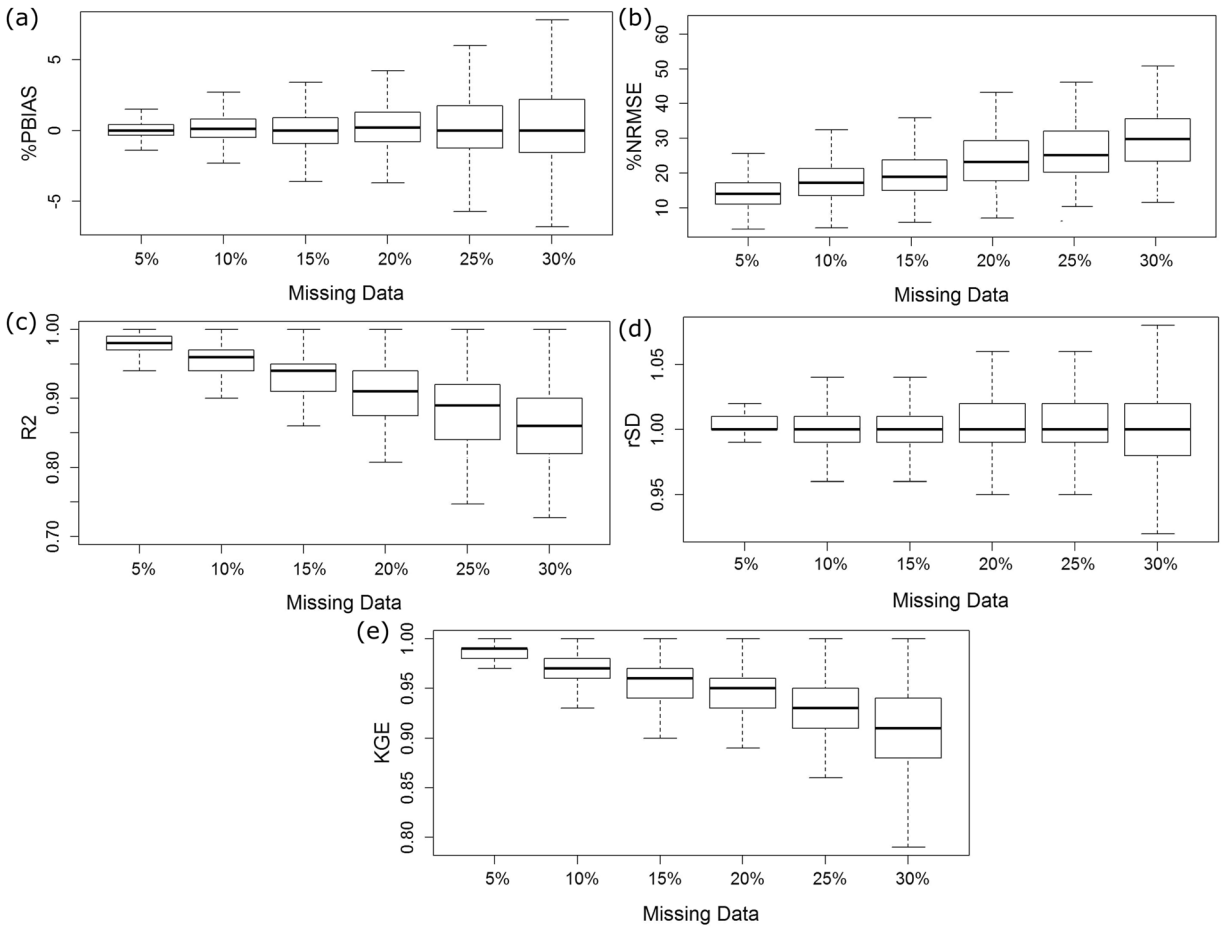

Fig. 6 Performance in filling missing values. Performance in estimating missing rainfall values in terms of: (**a**) bias in % (PBIAS); (**b**) normalized root mean squared error (NRMSE) in %,; (**c**) coefficient of determination (R^2); (**d**) ratio standard deviation (rSD); and (**e**) Kling-Gupta efficiency (KGE).

Development of gridded product. The initial assessment was conducted to select the covariates and interpolation method. Data from 80% of stations were used for model development and the rest for validation in this stage. The present study evaluated similar covariates (topography and distance to coast) used in developing WorldClim[42] data to improve the performance of interpolation. Previous studies[33,39] indicated the influence of topography and distance to coast on rainfall in the country. High rainfall is generally recorded at elevated locations[40]. In addition, air moisture usually decreases from the coast to inland[62]. These two covariates, i.e., elevation and distance to coast, are accounted for in this study. The TPS models of the work were therefore developed: (a) without covariates; (b) each of the covariates included separately; and (c) employing both covariates simultaneously. In addition, models are developed considering covariates as independent and linear. Since there is a non–linear effect of elevation and distance to coast on rainfall in the country, data are transformed using log and square root prior to inclusion in the TPS models. Prediction of multiple models developed with different combinations of covariates, considering their linear and nonlinear effects. However, the results show no improvement, and therefore, the model was developed without any covariates.

Kriging was one of the most efficient and widely used interpolation methods[25]. Therefore, the performance of TPS was compared with kriging to justify the use of TPS in generating gridded products. A similar process as the TPS was followed to validate the output of the kriging model. The results are presented in Figure S2 (supplementary materials), which show a nearly 2 to 10% improvement in the median of the evaluation metrics using the TPS method.

Performance of gridding method. The performance of TPS models without considering any covariates at all stations is presented in Fig. 8. The model's performance for the whole period (1901–2018), based on NRMSE, PBIAS, R^2, KGE, MD and PSS, is shown in Fig. 8(a). The medians of KGE, MD and PSS were 0.87, 0.90 and 0.98, respectively. The minimum values of the indices were 0.66, 0.83 and 0.96. The ranges of NRMSE and PBAIS were 9.3 to 20.2 and −9.7 to 11.7, and their medians were 17.2 and −0.9%, respectively. The median of rSD was 1.04. The results indicate a good performance of the TPS model in reconstructing rainfall time series of all rainfall stations (i.e., 322). The product was also evaluated by checking any abnormality in the generated surfaces compared to known monthly rainfall patterns or small circular patterns by potential outliers. No abnormal rainfall pattern was noticed for any month examined, indicating good homogenization of the series. LOOCV was used

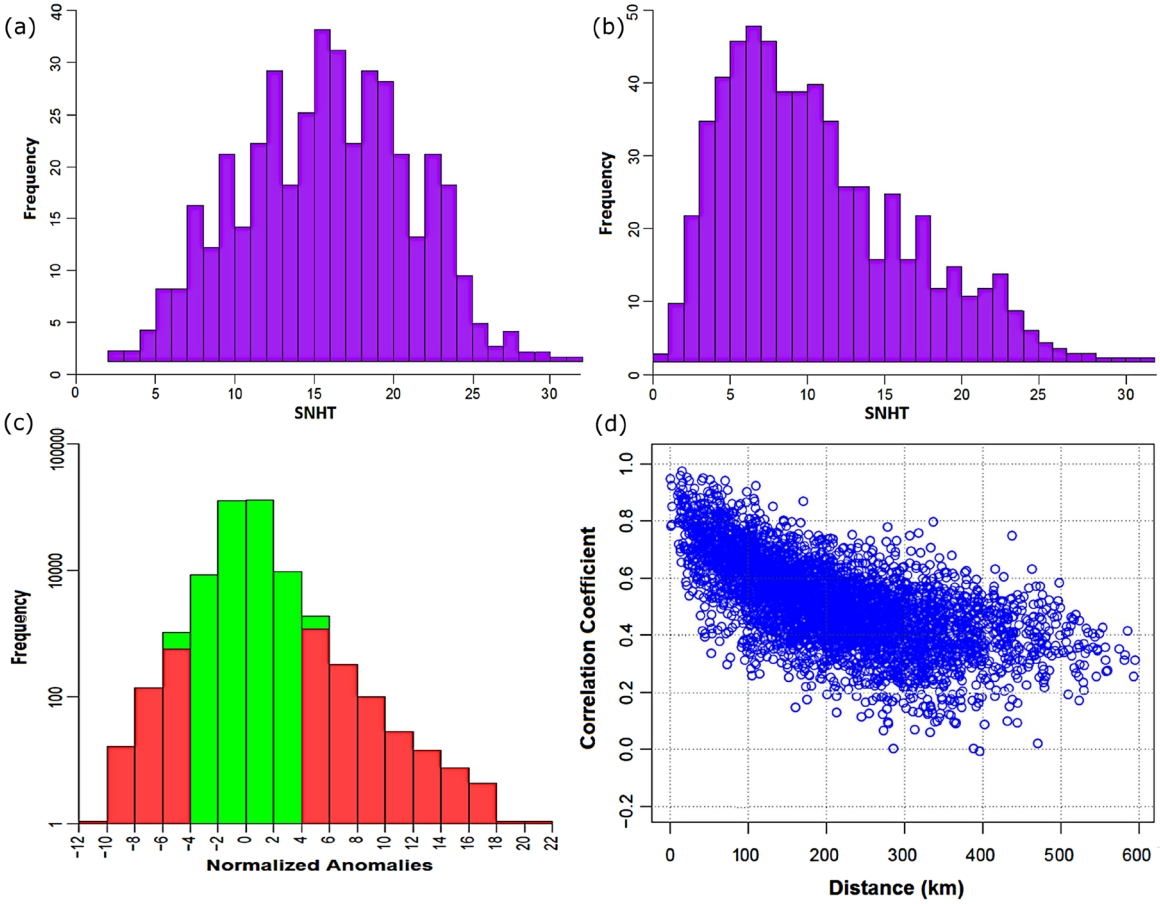

Fig. 7 Standard normal homogeneity test (SNHT) results. Optimum SNHT values were determined through visualization of histograms. (**a**,**b**) shows the varying maximum SNHT values of 326 stations for overlapping windows and whole series, respectively; (**c**) normalized anomalies during exploratory run; and (**d**) correlogram of the entire rainfall series.

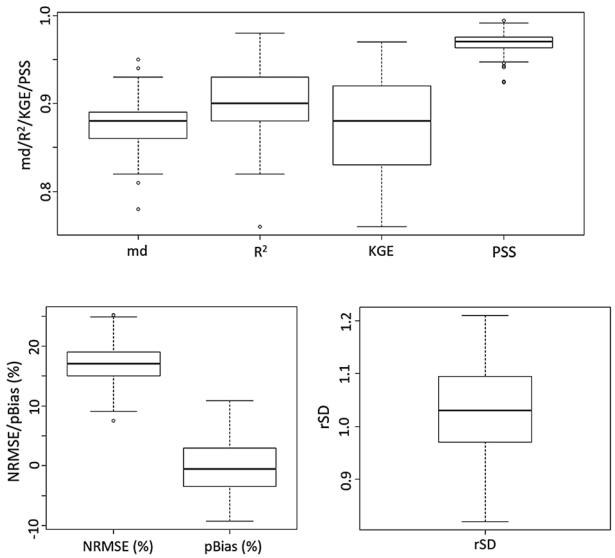

Fig. 8 Performance of the thin-plate spline model. Performance of the thin-plate spline model in reconstructing observed rainfall during 1901–2018. The boxplots show the model's performance for the whole period (1901–2018), based on normalized root mean squared error (NRMSE) in %, bias (PBIAS) in %, coefficient of determination (R^2), Kling-Gupta efficiency (KGE), Willmott's modified coefficient of agreement (MD) and Perkins skill score (PSS).

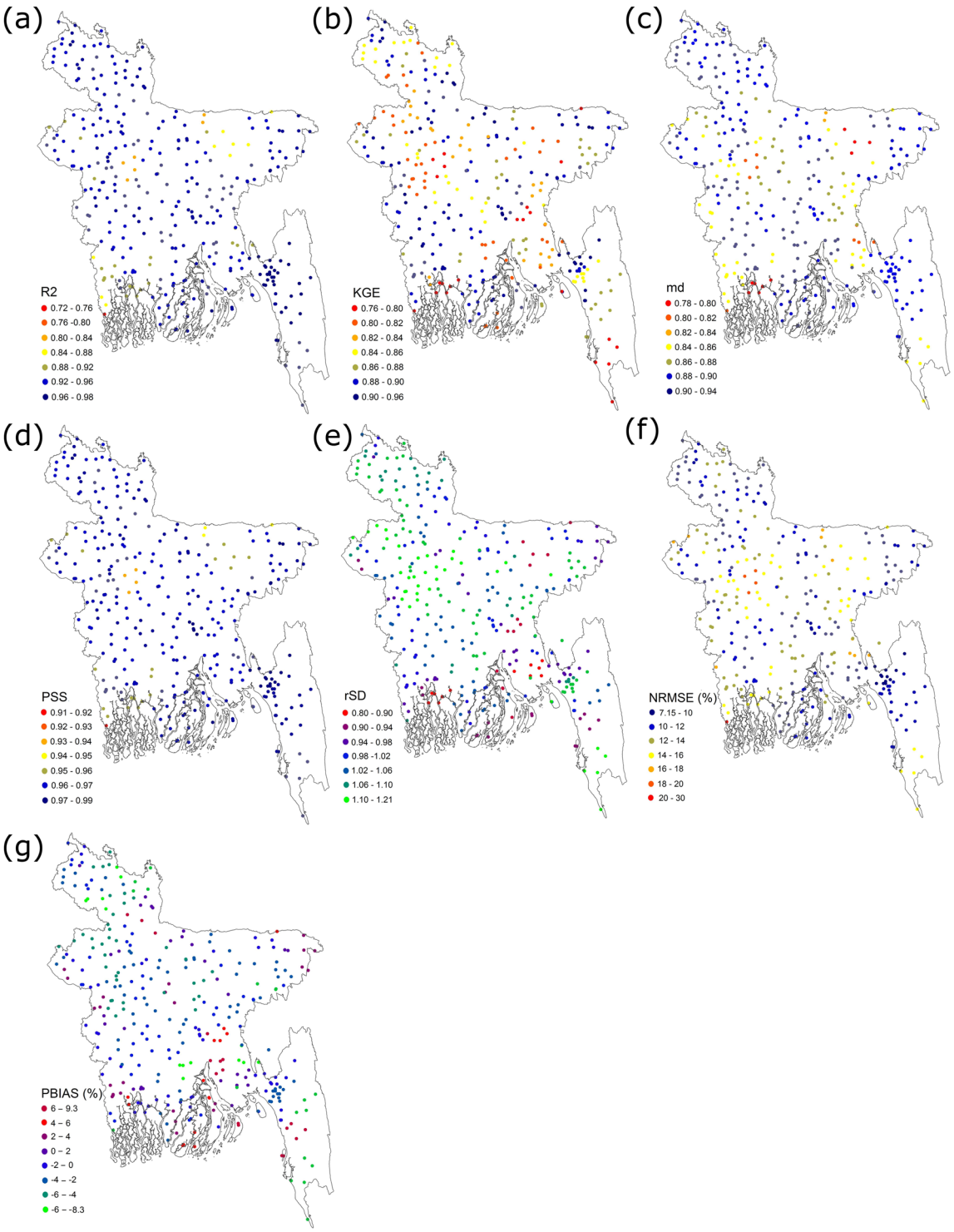

Fig. 9 The spatial distribution of BDGR performance. The spatial distribution of BDGR performance in terms of (**a**) coefficient of determination (R^2); (**b**) Kling-Gupta efficiency (KGE); (**c**) Willmott's modified coefficient of agreement (MD); (**d**) Perkins skill score (PSS); (**e**) ratio of standard deviation (rSD); (**f**) root mean squared error (NRMSE) in %; and (**g**) bias (PBIAS) in %.

to estimate the performance of interpolated rainfall at each station to show the spatial and temporal variability of the metrics.

Performance of BDGR. *Statistical evaluation.* The LOOCV approach provided error estimation for each station, thus quantifying errors in estimated rainfall both spatially and temporarily. The performance of rainfall at

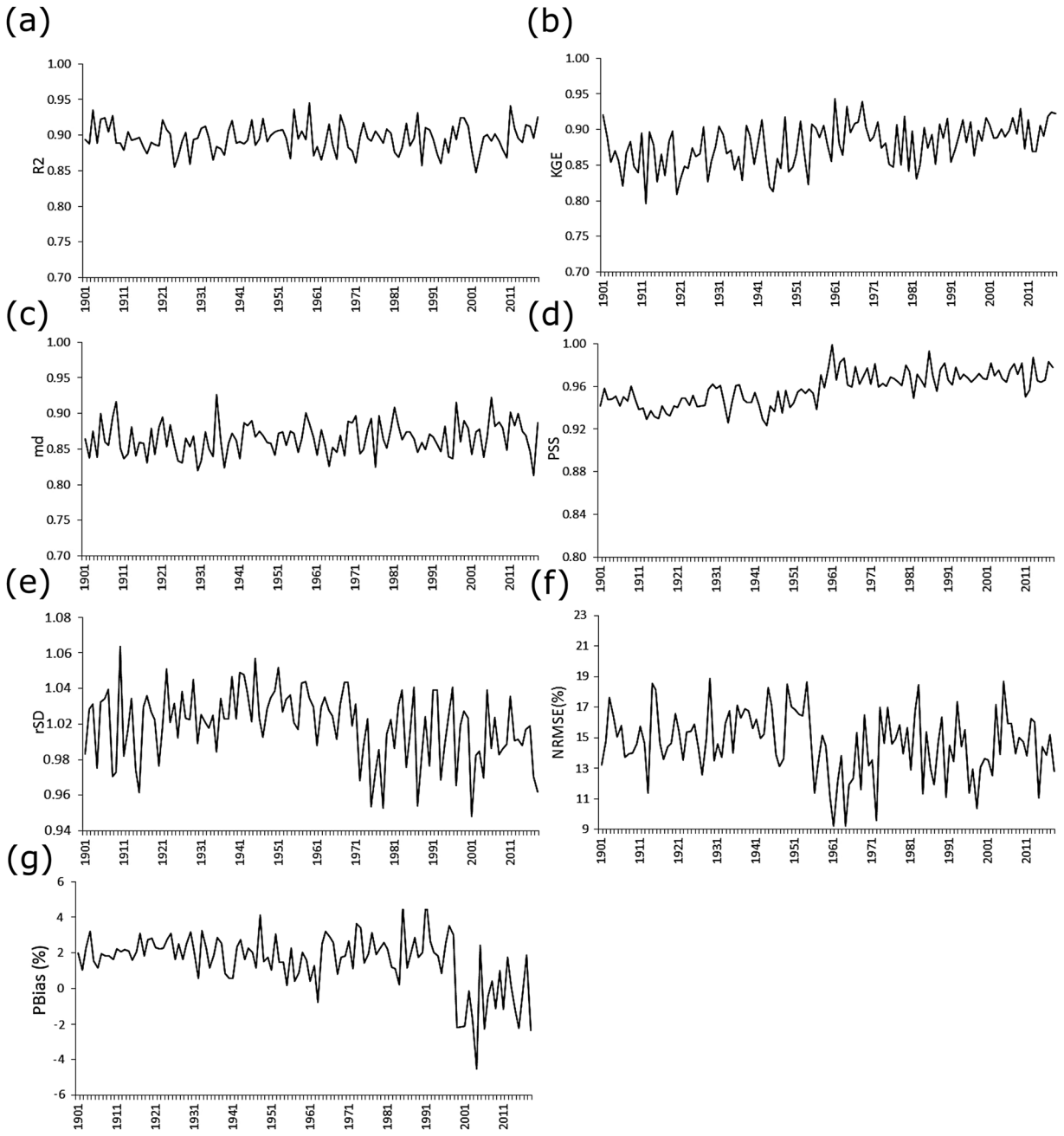

Fig. 10 The mean performance of the BDGR. The temporal variability of the mean performance of the BDGR over Bangladesh in terms of (**a**) coefficient of determination (R^2); (**b**) Kling-Gupta efficiency (KGE); (**c**) Willmott's modified coefficient of agreement (MD); (**d**) Perkins skill score (PSS); (**e**) ratio of standard deviation (rSD); (**f**) root mean squared error (NRMSE) in %; and (**g**) bias (PBIAS) in %.

each station was used to prepare the maps to show the spatial distribution of the performance metrics of BDGR (Fig. 9). The association metrics, R^2 and MD, revealed high correlations of observed and interpolated rainfall time series at most stations. The R^2 values were more than 0.88 at 79% of the stations and between 0.72 and 0.76 at only 2.6% of the stations. The MD was more than 0.86 at 88% of stations and between 0.79 and 0.82 at only three stations. The KGE was more than 0.8 at 98.5% of the stations. The PSS was above 0.96 at 93% of stations, while rSD was between 0.94 and 1.06 at 83% of stations. Among the two error metrics, NRMSE showed less than 14% error at 57% of the stations while between 20 and 30% only at 7% of stations. The PBIAS was between −2 and 2 at 78% of stations. It was out of the range of −6 to 6 only at nearly 8% of stations. The spatial distribution of the metrics showed a random distribution of high and low values over most of the country, except for the southwest and southeast corners. This was due to the low density of observation data at these locations. The northwest region is covered by dense forest, while the northwest corner is mountainous and less populated.

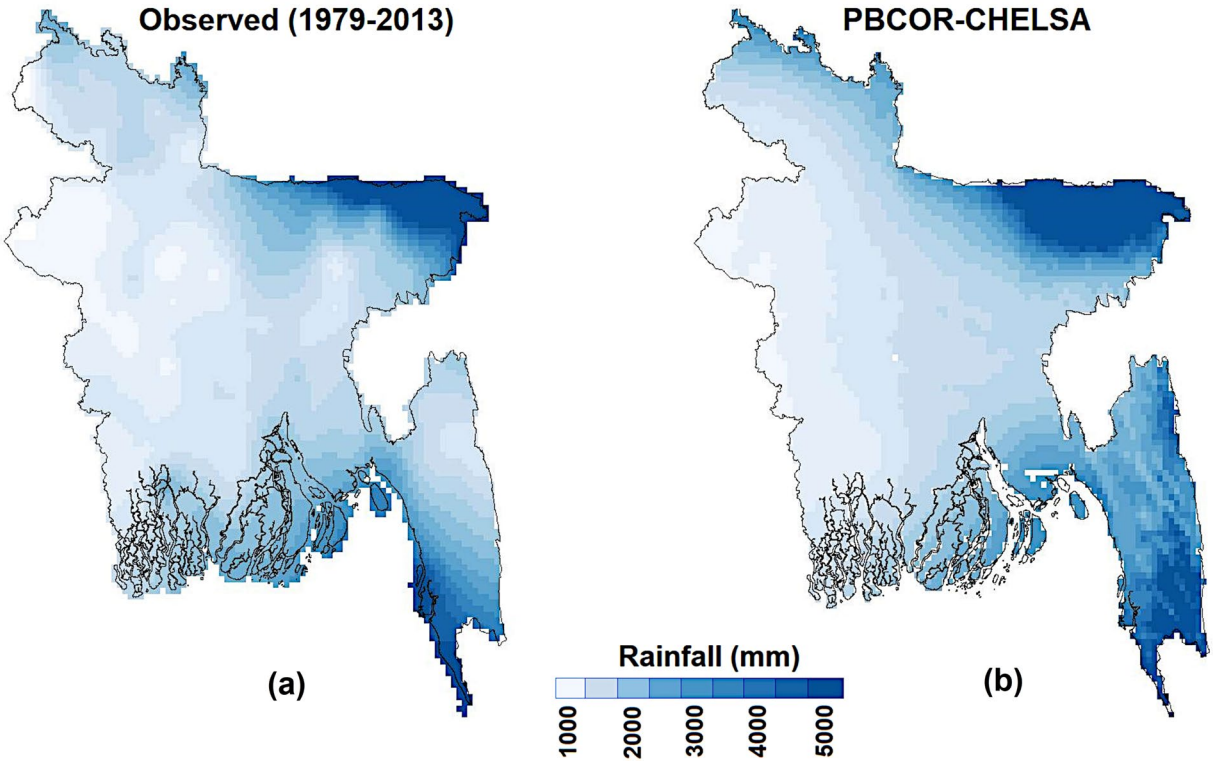

Fig. 11 Spatial distribution of annual rainfall. Spatial distribution of annual rainfall: (**a**) BDGR product; (**b**) PBCOR–CHELSA. The 0.01° resolution BDGR data are aggregated to the resolution of PBCOR–CHELSA (0.05°) for comparison. The results show similarity between BDGR and PBCOR–CHELSA annual rainfall climatology. The high and low regions and the transition of rainfall from the high to low zone from west to the east are well captured by the BDGR product.

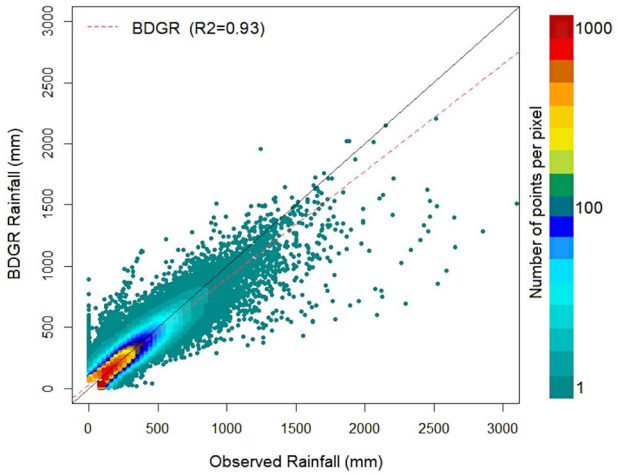

Fig. 12 Density scatter plot. Density scatter plot showing the relative performance of the BDGR compared to the observed monthly rainfall (n = 255952). The observed and nearby grid rainfall data for the whole period (1901–2018) are used for comparison. The red colour in the plot indicates dense data, and green shows low density. The best-fit line of the gridded products was presented using a dashed red line. A close correspondence of the regression line to the diagonal (solid black line) indicates better performance of the BDGR product.

The temporal variability in the performance of the BDGR over Bangladesh is shown in Fig. 10. The performance metrics estimated at 322 stations for different years were averaged to show the temporal variable of BDGR performance over Bangladesh for the period 1901–2018. The results showed the temporal variability of R^2 in the range of 0.87 to 0.95 and md between 0.84 and 0.93 for the period 1901 to 2018. The KGE showed an increase in performance with time. The mean KGE was 0.85 in the early period (1901–1960) and increased to nearly 0.89 after 1960. Similar improvements were observed in all other metrics after 1960. This is due to the

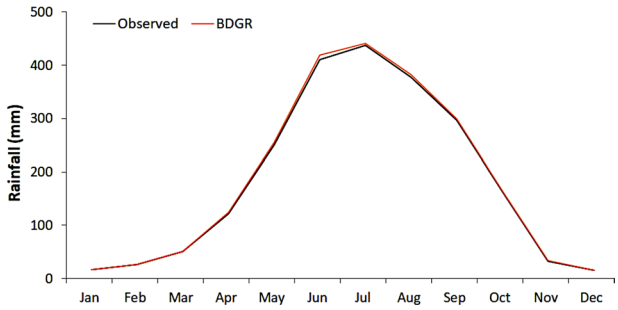

Fig. 13 Performance in reconstructing seasonal rainfall variability. Performance of the BDGR in reconstructing seasonal rainfall variability, 1979−2010. All observations were averaged and compared with the average value of all grids of BDGR. The BDGR monthly average rainfall line completely overlapped with the observed data, indicating its ability to reconstruct the seasonal variation in rainfall for the country.

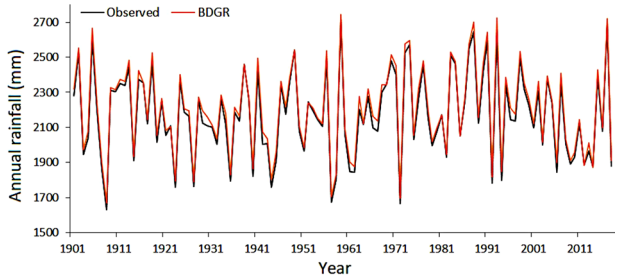

Fig. 14 Performance in replicating annual rainfall. Performance of the BDGR product in replicating rainfall in Bangladesh, 1901−2018. The annual rainfall line of the BDGR overlaps with the observed rainfall for the whole study period (i.e., 1901−2018), indicating the ability of newly generated gridded data to estimate the long-term climate variability of the country.

high number of stations used for interpolation after 1960 compared to the early period. It has been mentioned earlier that no data were available at 184 stations until 1960. However, the overall performance of the interpolated rainfall was sufficiently high for the whole period. The PSS was in the range of 0.93 to 0.99, rSD between 0.96 and 1.066, NRMSE between 9 and 19% and PBIAS in the range of −4.5 to 4.3%.

Boxplots of each error metric for different periods were prepared following Lucas, *et al.*[63] to show the error spread in different periods. Metric values were grouped every five years and presented for 1901–2018 in Supplementary Figure S1. The figure shows no apparent discontinuities in any of the error series.

Spatial variability. The annual and monthly BDGR climatologies of BDGR are compared with PBCOR rainfall climatologies to show the performance of BDGR relative to PBCOR. The BDGR and PBCOR CHELSA annual rainfall distributions for 1979–2013 are shown in Fig. 11, and WC and PBCOR CHPclim annual rainfall for the periods 1970–2000 and 1980–2009 are shown in Supplementary Figure S4. In all cases, the 0.01° resolution BDGR data are aggregated to the resolution of PBCOR datasets (0.05°) for comparison. The reason for comparing different datasets is that they all together cover a longer period (1970–2013). The results show that the BDGR annual rainfall climatology was correlated with PBCOR CHELSA by 0.91, WC by 0.89 and CHPclim by 0.88. The high and low regions and the transition of rainfall from the high to low zone from west to the east are well captured by the BDGR product. The PBCOR rainfall climatology is developed with limited station data, and therefore, local heterogeneity in the distribution of rainfall is not visible. However, regional and local heterogeneity was well captured by the BDGR product, indicating its ability to represent the spatial distribution of annual rainfall.

A similar analysis with PBCOR CHELSA was conducted for all months, and their spatial distribution is shown in Supplementary Figure S5. The results showed correlations of BDGR climatology with PBCOR CHELSA between 0.74 and 0.99 for different months.

BDGR and observed rainfall. The performance of the BDGR was also evaluated to (i) reconstruct temporal patterns, (ii) reproduce seasonal variability, and (ii) estimate trends in different quantiles of observed rainfall. These evaluations were conducted to show the reliability of the BDGR as a substitute for observed rainfall.

The relative performance of the BDGR compared to the observed monthly rainfall is shown using a scatter plot in Fig. 12. The observed and nearby grid rainfall data for the whole period (1901–2018) are used for comparison. Red colour in the plot indicates dense data, and green shows low density. The best-fit line of the gridded products was presented using a dashed red line in the plot. A close correspondence of the regression line to the diagonal (solid black line) indicates better performance of the BDGR product. Figure 9 shows the regression line

was closer to the diagonal line. The R^2 for the BDGR was 0.93, indicating better correspondance of the generated product in replicating rainfall time series across Bangladesh.

The seasonal performance of the BDGR product is shown in Fig. 13. All observations were averaged and compared with the average value of all grids of BDGR. The comparison was made for each month to demonstrate BDGR's ability in replicating observed variability of seasonal rainfall. The BDGR monthly average rainfall line completely overlapped with the observed data, indicating its ability to reconstruct the seasonal variation in rainfall for the country.

The annual rainfall series provides an estimation of the wet and dry years. In doing so, all observations were first averaged and then converted to annual time series. Likewise, rainfall data of all BDGR grids were average and converted to annual series. The observed and BDGR annual rainfall for the period 1901−2018 are presented in Fig. 14, illustrating that the annual rainfall line of the BDGR overlaps with the observed rainfall for the whole study period (i.e., 1901−2018). Furthermore, this study indicates the ability of newly generated gridded data to estimate the long-term climate variability of the country.

Data Records

High resolution (0.01°) monthly total rainfall, in mm, for Bangladesh from January 1901 to December 2018, are freely available at https://doi.org/10.6084/m9.figshare.16607912 in NetCDF[64]. The data records spatially, covering land area of Bangladesh. The records will be updated frequently in the upcoming years when more rainfall records will be available. Temporal coverage of the data may be extended in the future when recent observations are accessible.

Usage Notes

The BDGR product can be used for many applications at various temporal resolutions. The datasets can be used to estimate spatial distribution of rainfall, temporal pattern, seasonality, and the trends more accurately than any other datasets, presently available for the country. Furthermore, this high resolution (e.g., 0.01°) data can be used for localized changes in long–term climate, including the changes in rainfall, and groundwater recharge, along with other scientific and social benefits.

Code availability

A code is written using statistical package (R.4.1) to process data. The code is available online[64] (https://doi.org/10.6084/m9.figshare.16607912).

Received: 1 October 2021; Accepted: 13 July 2022;
Published online: 03 August 2022

References

1. Islam, M. R. Climate Change, Natural Disasters and Socioeconomic Livelihood Vulnerabilities: Migration Decision Among the Char Land People in Bangladesh. *Social Indicators Research* **136**, 575–593, https://doi.org/10.1007/s11205-017-1563-y (2018).
2. Carrico, A. R. & Donato, K. Extreme weather and migration: evidence from Bangladesh. *Population and Environment* **41**, 1–31, https://doi.org/10.1007/s11111-019-00322-9 (2019).
3. Ghose, B. *et al.* Rice yield responses in Bangladesh to large-scale atmospheric oscillation using multifactorial model. *Theoretical and Applied Climatology* **146**, 29–44, https://doi.org/10.1007/s00704-021-03725-7 (2021).
4. Ali, A. Climate change impacts and adaptation assessment in Bangladesh. *Climate Research* **12**, 109–116 (1999).
5. Huq, S. Climate Change and Bangladesh. *Science* **294**, 1617–1617, https://doi.org/10.1126/science.294.5547.1617 (2001).
6. Shahid, S. Recent trends in the climate of Bangladesh. *Climate Research* **42**, 185–193, https://doi.org/10.3354/cr00889 (2010).
7. Ozaki, M. in *ADB South Asia Working Paper Series* Vol. 46 (Asian Development Bank, 2016).
8. Rahman, R. & Salehin, M. in *Disaster Risk Reduction Approaches in Bangladesh* (eds R., Shaw, F., Mallick, & A., Islam) 65–90 (Springer Japan, 2013).
9. Mohsenipour, M., Shahid, S., Chung, E. S. & Wang, X. J. Changing Pattern of Droughts during Cropping Seasons of Bangladesh. *Water Resources Management* **32**, 1555–1568, https://doi.org/10.1007/s11269-017-1890-4 (2018).
10. Shahid, S. & Hazarika, M. K. Groundwater drought in the northwestern districts of Bangladesh. *Water Resources Management* **24**, 1989–2006, https://doi.org/10.1007/s11269-009-9534-y (2010).
11. Salem, G. S. A., Kazama, S., Shahid, S. & Dey, N. C. Groundwater-dependent irrigation costs and benefits for adaptation to global change. *Mitigation and Adaptation Strategies for Global Change* **23**, 953–979, https://doi.org/10.1007/s11027-017-9767-7 (2018).
12. Chowdhury, N. T. Water management in Bangladesh: an analytical review. *Water Policy* **12**, 32–51, https://doi.org/10.2166/wp.2009.112 (2009).
13. Mahmuduzzaman, M., Uddin, Z., Nuruzzaman, A., Rabbi, F. & Ahmed, S. Causes of Salinity Intrusion in Coastal Belt of Bangladesh. *International Journal of Plant Research* **2014**, 8–13, https://doi.org/10.5923/s.plant.201401.02 (2014).
14. Perry, M. & Hollis, D. The generation of monthly gridded datasets for a range of climatic variables over the UK. *International Journal of Climatology* **25**, 1041–1054, https://doi.org/10.1002/joc.1161 (2005).
15. Adler, R. F. *et al.* The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation. *Atmosphere* **9**, 138, https://doi.org/10.3390/atmos9040138 (2018).
16. Harris, G. R., Sexton, D. M. H., Booth, B. B. B., Collins, M. & Murphy, J. M. Probabilistic projections of transient climate change. *Climate Dynamics* **40**, 2937–2972, https://doi.org/10.1007/s00382-012-1647-y (2013).
17. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965–1978, https://doi.org/10.1002/joc.1276 (2005).
18. Yatagai, A. *et al.* APHRODITE Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges. *Bulletin of the American Meteorological Society* **93**, 1401–1415, https://doi.org/10.1175/Bams-D-11-00122.1 (2012).
19. Maurer, E. P., Wood, A., Adam, J., Lettenmaier, D. P. & Nijssen, B. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal of climate* **15**, 3237–3251, 10.1175/1520-0442(2002)015<3237:Althbd>2.0.Co;2 (2002).
20. Srivastava, A., Rajeevan, M. & Kshirsagar, S. Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region. *Atmospheric Science Letters* **10**, 249–254, https://doi.org/10.1002/asl.232 (2009).
21. Herrera, S. *et al.* Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02). *International Journal of Climatology* **32**, 74–85, https://doi.org/10.1002/joc.2256 (2012).

22. Nashwan, M. S., Shahid, S. & Chung, E.-S. Development of high-resolution daily gridded temperature datasets for the central north region of Egypt. *Scientific Data* **6**, 138, https://doi.org/10.1038/s41597-019-0144-0 (2019).
23. Aryee, J. N. A. *et al*. Development of high spatial resolution rainfall data for Ghana. *International Journal of Climatology* **38**, 1201–1215, https://doi.org/10.1002/joc.5238 (2018).
24. Li, J. & Heap, A. Spatial interpolation methods: a review for environmental scientists. *Geoscience Australia, Record. Geoscience Australia, Canberra* (2008).
25. Chiles, J.-P. & Delfiner, P. *Geostatistics: modeling spatial uncertainty*. Vol. 497 (John Wiley & Sons, 2009).
26. Smith, W. H. F. & Wessel, P. Gridding with continuous curvature splines in tension. *GEOPHYSICS* **55**, 293–305, https://doi.org/10.1190/1.1442837 (1990).
27. Wahba, G. *Spline models for observational data*. (SIAM, 1990).
28. Hagan, P. S. & West, G. Interpolation Methods for Curve Construction. *Applied Mathematical Finance* **13**, 89–129, https://doi.org/10.1080/13504860500396032 (2006).
29. Chacon-Hurtado, J. C., Alfonso, L. & Solomatine, D. P. Rainfall and streamflow sensor network design: a review of applications, classification, and a proposed framework. *Hydrol. Earth Syst. Sci.* **21**, 3071–3091, https://doi.org/10.5194/hess-21-3071-2017 (2017).
30. EPWAPDA. in *Water Supply Paper* Vol. 8, 314 (East Pakistan Water and Power Development Authority (EPWAPDA), 1960).
31. EPWAPDA. in *Water Supply Paper* Vol. 9, 292 (East Pakistan Water and Power Development Authority (EPWAPDA), 1960).
32. Chowdhury, A. K., Kar, K. K., Shahid, S., Chowdhury, R. & Rashid, M. M. Evaluation of spatio-temporal rainfall variability and performance of a stochastic rainfall model in Bangladesh. *International Journal of Climatology* **39**, 4256–4273, https://doi.org/10.1002/joc.6071 (2019).
33. Wahiduzzaman, M. *et al*. Trends and Variabilities of Thunderstorm Days over Bangladesh on the ENSO and IOD Timescales. *Atmosphere* **11**, 1176 (2020).
34. Kelley, C. *et al*. Exploring the predictability of within-season rainfall statistics of the Bangladesh monsoon using North American Multimodel Ensemble outputs. *Theoretical and Applied Climatology* **141**, 495–508, https://doi.org/10.1007/s00704-020-03202-7 (2020).
35. Hatsuzuka, D. & Fujinami, H. Effects of the South Asian Monsoon Intraseasonal Modes on Genesis of Low Pressure Systems over Bangladesh. *Journal of Climate* **30**, 2481–2499, https://doi.org/10.1175/JCLI-D-16-0360.1 (2017).
36. Nashwan, M. S., Shahid, S. & Wang, X. J. Uncertainty in Estimated Trends Using Gridded Rainfall Data: A Case Study of Bangladesh. *Water* **11**, 349, https://doi.org/10.3390/w11020349 (2019).
37. Ahmed, R. & Kim, I.-K. Patterns of Daily Rainfall in Bangladesh During the Summer Monsoon Season: Case Studies at Three Stations. *Physical Geography* **24**, 295–318, https://doi.org/10.2747/0272-3646.24.4.295 (2003).
38. Dewan, A., Ongee, E. T., Rafiuddin, M., Rahman, M. M. & Mahmood, R. Lightning activity associated with precipitation and CAPE over Bangladesh. *International Journal of Climatology* **38**, 1649–1660, https://doi.org/10.1002/joc.5286 (2018).
39. Shahid, S. Trends in extreme rainfall events of Bangladesh. *Theoretical and Applied Climatology* **104**, 489–499, https://doi.org/10.1007/s00704-010-0363-y (2011).
40. Shahid, S. Rainfall variability and the trends of wet and dry periods in Bangladesh. *International Journal of Climatology* **30**, 2299–2313, https://doi.org/10.1002/joc.2053 (2010).
41. Beck, H. E. *et al*. Bias Correction of Global High-Resolution Precipitation Climatologies Using Streamflow Observations from 9372 Catchments. *Journal of Climate* **33**, 1299–1315, https://doi.org/10.1175/JCLI-D-19-0332.1 (2020).
42. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302–4315, https://doi.org/10.1002/joc.5086 (2017).
43. Karger, D. N. *et al*. Climatologies at high resolution for the earth's land surface areas. *Sci Data* **4**, 170122, https://doi.org/10.1038/sdata.2017.122 (2017).
44. Funk, C. *et al*. The climate hazards infrared precipitation with stations–a new environmental record for monitoring extremes. *Sci Data* **2**, 150066, https://doi.org/10.1038/sdata.2015.66 (2015).
45. Malanson, G. P. Research resource review. *Progress in Physical Geography: Earth and Environment* **44**, 978–981, https://doi.org/10.1177/0309133320956946 (2020).
46. Van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *Journal of statistical software* **45**, 1–67 (2011).
47. Freitas, L., Pereira, M. G., Caramelo, L., Mendes, M. T. & Nunes, L. F. Homogeneity of monthly air temperature in Portugal with HOMER and MASH. Idojaras, **117**(1), 69–90 (2013)
48. Coll, J. *et al*. Application of homogenization methods for Ireland's monthly precipitation records: Comparison of break detection results. *International Journal of Climatology* **40**, 6169–6188, https://doi.org/10.1002/joc.6575 (2020).
49. Pastor *et al*. 98–106 (WMO Climate Data and Monitoring WCDMP).
50. Skrynyk, O., Aguilar, E., Guijarro, J., Randriamarolaza, L. Y. A. & Bubin, S. Uncertainty evaluation of Climatol's adjustment algorithm applied to daily air temperature time series. *International Journal of Climatology* **41**, E2395–E2419, https://doi.org/10.1002/joc.6854 (2021).
51. Domonkos, P., Guijarro, J. A., Venema, V., Brunet, M. & Sigró, J. Efficiency of Time Series Homogenization: Method Comparison with 12 Monthly Temperature Test Datasets. *Journal of Climate* **34**, 2877–2891, https://doi.org/10.1175/JCLI-D-20-0611.1 (2021).
52. Dumitrescu, A., Cheval, S. & Guijarro, J. A. Homogenization of a combined hourly air temperature dataset over Romania. *International Journal of Climatology* **40**, 2599–2608, https://doi.org/10.1002/joc.6353 (2020).
53. Montero-Martínez, M. J., Pita-Díaz, O. & Andrade-Velázquez, M. Potential Influence of the Atlantic Multidecadal Oscillation in the Recent Climate of a Small Basin in Central Mexico. *Atmosphere* **13**, 339 (2022).
54. Guijarro, J. A. Homogenization of climatic series with Climatol. *Reporte técnico State Meteorological Agency (AEMET), Balearic Islands Office, Spain* (2018).
55. Clarke, M. R. B. The reduced major axis of a bivariate sample. *Biometrika* **67**, 441–446, https://doi.org/10.1093/biomet/67.2.441 (1980).
56. Hutchinson, M. F. & Xu, T. Anusplin version 4.2 user guide. *Centre for Resource and Environmental Studies. The Australian National University. Canberra* **5** (2004).
57. Xu, Y. *et al*. Development and preliminary application of a gridded surface air temperature homogenized dataset for China. *Theoretical and Applied Climatology* **139**, 505–516, https://doi.org/10.1007/s00704-019-02972-z (2020).
58. Hofstra, N., Haylock, M., New, M. & Jones, P. D. Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. *Journal of Geophysical Research: Atmospheres* **114**, https://doi.org/10.1029/2009JD011799 (2009).
59. Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* **377**, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003 (2009).
60. Willmott, C. J. On the Validation of Models. *Physical Geography* **2**, 184–194, https://doi.org/10.1080/02723646.1981.10642213 (2013).
61. Perkins, S. E., Pitman, A. J., Holbrook, N. J. & McAneney, J. Evaluation of the AR4 Climate Models' Simulated Daily Maximum Temperature, Minimum Temperature, and Precipitation over Australia Using Probability Density Functions. *Journal of Climate* **20**, 4356–4376, https://doi.org/10.1175/jcli4253.1 (2007).
62. Ziarh, G. F., Shahid, S., Ismail, T. B., Asaduzzaman, M. & Dewan, A. Correcting bias of satellite rainfall data using physical empirical model. *Atmospheric Research* **251**, 105430, j.atmosres.2020.105430 (2021).

63. Lucas, M. P. *et al*. Optimizing Automated Kriging to Improve Spatial Interpolation of Monthly Rainfall over Complex Terrain. *Journal of Hydrometeorology* **23**, 561–572, https://doi.org/10.1175/JHM-D-21-0171.1 (2022).
64. Dewan, A. *et al*. Long–term High–resolution Gridded Rainfall Product for Bangladesh (BDGR). *figshare* https://doi.org/10.6084/m9.figshare.16607912.v6 (2021).

Acknowledgements

The authors would like to acknowledge Bangladesh Meteorological Department and Bangladesh Water Development Board for giving access to their data archive. We also grateful to Dr Dewan Abdul Quadir for his support at various stages of this study, especially for giving access to WSP books.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-022-01568-z.

Correspondence and requests for materials should be addressed to E.-S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.