RESEARCH ARTICLE

International Journal of Climatology · RMetS

Calibrated multi-model ensemble seasonal prediction of Bangladesh summer monsoon rainfall

Nachiketa Acharya[1,2] | **Carlo Montes**[3] | **S. M. Q. Hassan**[4] | **Razia Sultana**[4] |
Md. Bazlur Rashid[4] | **Md. Abdul Mannan**[4] | **Timothy J. Krupnik**[5]

[1]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, Boulder, Colorado, USA

[2]National Oceanic and Atmospheric Administration (NOAA), Physical Sciences Laboratory, Boulder, Colorado, USA

[3]International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

[4]Bangladesh Meteorological Department (BMD), Dhaka, Bangladesh

[5]International Maize and Wheat Improvement Center (CIMMYT), Dhaka, Bangladesh

Correspondence
Nachiketa Acharya, Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, Boulder, CO, USA.
Email: dr.nachiketaacharya@gmail.com

Abstract

Bangladesh summer monsoon rainfall (BSMR), typically from June through September (JJAS), represents the main source of water for multiple sectors. However, its high spatial and interannual variability makes the seasonal prediction of BSMR crucial for building resilience to natural disasters and for food security in a climate-risk-prone country. This study describes the development and implementation of an objective system for the seasonal forecasting of BSMR, recently adopted by the Bangladesh Meteorological Department (BMD). The approach is based on the use of a calibrated multi-model ensemble (CMME) of seven state-of-the-art general circulation models (GCMs) from the North American Multi-Model Ensemble project. The lead-1 (initial conditions of May for forecasting JJAS total rainfall) hindcasts (spanning 1982–2010) and forecasts (spanning 2011–2018) of seasonal total rainfall for the JJAS season from these seven GCMs were used. A canonical correlation analysis (CCA) regression is used to calibrate the raw GCMs outputs against observations, which are then combined with equal weight to generate final CMME predictions. Results show, compared to individual calibrated GCMs and uncalibrated MME, that the CCA-based calibration generates significant improvements over individual raw GCM in terms of the magnitude of systematic errors, Spearman's correlation coefficients, and generalised discrimination scores over most of Bangladesh areas, especially in the northern part of the country. Since October 2019, the BMD has been issuing real-time seasonal rainfall forecasts using this new forecast system.

KEYWORDS
climate services, multi-model ensemble, North American multi-model ensemble project, seasonal forecasting

1 | INTRODUCTION

Located in sub-tropical South Asia, Bangladesh is one of the world's most densely populated countries. Bangladesh is characterised by a tropical monsoon-type climate, with a warm and rainy summer, and a pronounced dry season in winter, features that make it highly vulnerable to the effects of interannual climate variability (Rahman &

Lateh, 2015) and change (Huq, 2001). The country experiences an unimodal rainfall distribution, with most of the rainfall typically concentrated from June through September (JJAS). During this period, Bangladesh receives about 70% of the total annual rainfall, with coefficient of variability that has been quantified around 12% (Ahasan et al., 2010). The pattern of the Bangladesh's summer monsoon rainfall (BSMR) is highly variable spatially, exhibiting a general west–east climatological gradient in annual rainfall ranging from 1500 to 4400 mm (Montes, Acharya, Hassan, & Krupnik, 2021; Nashwan et al., 2019). This pattern of variability strongly shapes human livelihoods, especially in agriculture, which is a mainstay of the country's economy. For instance, crop management decisions and production losses often occur in Bangladesh as a consequence of early or later arrival of rains, along with excess or deficient monsoon rainfall amounts (Nahar et al., 2018). Consequently, reliable BSMR forecasting at actionable time scales could potentially play a significant role in the planning and management of agriculture and other activities such as flood management, urban planning, water-resource management and optimal operation of irrigation systems (Hansen et al., 2006; Montes, Acharya, Stiller-Reeve, Kelley, & Hassan, 2021).

Seasonal climate-prediction efforts in Bangladesh have been based mostly on statistical and empirical forecasting methods using Auto-Regressive Integrated Moving Average (ARIMA) models for rainfall and temperature prediction (Bari et al., 2015; Mahmud et al., 2017; Mohsin et al., 2012; Rahman & Lateh, 2015) or regression models of the teleconnections between rainfall and various predictors such as sea-surface temperature (SST; Hossain et al., 2019; Mannan et al., 2015; Rahman et al., 2013a). ARIMA models have been used to predict rainfall with lead times of up to 12 months (Mahmud et al., 2017), but the lack of statistical significance of year-to-year autocorrelation can lead to limited forecasting skills (Dahale & Singh, 1993). A more widely-used approach has been the use of empirical relationships between observed BSMR and predictors such as sea-surface temperature, surface air temperature and pressure gradients (Hossain et al., 2019; Rahman et al., 2013b). For instance, the prediction of the monthly and seasonal frequency of rainy days and heavy rainfall events have been attempted using SST as a predictor (Mannan et al., 2015), and skill is higher than for the monsoon seasonal total amount, consistent with results from other parts of the world (Robertson et al., 2009). Nevertheless, the relatively weak teleconnection between sources of seasonal predictability such as El Niño-Southern Oscillation (ENSO) and seasonal climate in Bangladesh strongly limits the skill of these rainfall forecasts compared to other parts of the globe (Ahmed et al., 2017; Cash et al., 2017; Hossain et al., 2019). Due to the complexity of the diverse climate interactions in the vicinity of Bangladesh, non-linear and data-driven forecasting methods, such as artificial neural networks, adaptive neuro-fuzzy inference systems (ANFIS) and genetic algorithms, may have some advantages over linear methods (Banik et al., 2009) if sufficiently long time-series are available.

State-of-the-art general circulation models (GCMs) that represent atmospheric processes provide an alternative non-linear physically-based approach to statistical modelling (Kang et al., 2004; Kang and Shukla, 2005). This approach may produce more accurate and reliable climate predictions compared to statistical models based on empirical relationships (mostly linear) from observational data (Barnston & Tippett, 2017). However, predictions from GCMs often require correction due to their inherent systematic biases (Acharya et al., 2013; Tippett et al., 2007; Wilks, 2002). Calibration methods can be used to modify the amplitudes of large-scale patterns, and also to refine the details of anomaly patterns for local downscaling (Acharya et al., 2021; Barnston & Tippett, 2017; Doblas-Reyes et al., 2005; Tippett et al., 2008; Wilks, 2017). In this sense, multiple efforts have been carried out in order to quantify the improvements in skill from GCMs after calibration over different regions worldwide. However, in Bangladesh, these efforts have focused on single-location approaches but not at the country level (e.g., Montes et al., 2022).

Officially, the Bangladesh Meteorological Department (BMD) is responsible for providing operational seasonal and monthly monsoon climate predictions to climate information users. BMD has used a subjective consensus approach based on meteorologists' experience to generate products using all available Global Producing Center's forecasts and other available information. This subjectively-based forecasting approach, however, has been found to be a poor fit for many decision makers interested in more reliable and objective forecasts. There is an increasing demand for high-resolution seasonal forecasts over Bangladesh at sufficient lead times to allow response planning from users in agriculture, hydrology, disaster management, energy, health, and other sectors. This demand has prompted the research for the development of an objective seasonal forecast system following the World Meteorological Organization's (WMO) recently published seasonal-forecast guidance (World Meteorological Organization (WMO), 2020). The guidance advocates the use of an objective seasonal forecast procedure, defined as a traceable, reproducible, and well-documented set of steps that allows the quantification of forecast quality. The WMO has started to promote the adoption of such objective-based forecasting methods at Regional Climate Outlook Forums (WMO, 2017, 2020) and by National Meteorological and Hydrological Services. In response, an objective forecasting system was developed for seasonal forecasting for Bangladesh, similar to others

recently developed around the world (Acharya, Dinku, et al., 2020; Acharya et al., 2021; IRI, 2020). This advanced forecast system enables calibration, combination, and verification of objective climate forecasts from the state-of-the-art GCMs of the North American Multi-Model Ensemble (NMME) project, and positions BMD to generate and deliver targeted climate information products that could be made relevant to the needs of decision-makers. Although multi-model-based methods have been explored for the Indian subcontinent (Acharya, Kar, et al., 2011; Kar et al., 2012; Rajeevan et al., 2012), this is the first time, to our knowledge, that they have been used for Bangladesh at the country level, aligned with BMD needs. As of October 2019, this new forecast system is used in real-time by the BMD (http://live.bmd.gov.bd/p/ThreeMonth283/). Therefore, from an operational perspective, the potential benefits of this new forecasting system need to be assessed in terms of hindcast skill assessment.

In this article, we describe the development and performance of an objective forecasting system which is based on calibrated multi-model ensemble (CMME) system in the seasonal prediction of BSMR and compare its performance with uncalibrated GCMs. The paper is organised as follows: in Section 2, we briefly describe the data used in this study, including NMME GCMs and the observational reference; in Section 3, we explain the procedures of the proposed canonical correlation analysis (CCA)-based calibration methods and illustrate how the methods are employed in practice to make CMME-based forecasts. In Section 4, we examine the performance of calibrated individual model outputs compared to that of uncalibrated outputs, along with validation of the CMME system compared; in Section 5, we provide a brief discussion and draw conclusions.

2 | DATASETS

2.1 | Observational reference

Developed by the Columbia University's International Research Institute for Climate and Society (IRI) and BMD, the latest Enhancing National Climate Services for Bangladesh Meteorological Department (ENACTS-BMD) dataset (Acharya, Faniriantsoa, et al., 2020) version is used in this study. The ENACTS-BMD dataset is a high-resolution ($0.05° \times 0.05°$) daily gridded rainfall and temperature dataset constructed by blending data from BMD weather stations, satellite products (for rainfall) and reanalysis data (for temperature). Since February 2020, BMD archives and maintains this dataset. Its record begins in January 1981 and is ongoing (updated every month in real-time) at daily, decadal and monthly temporal resolutions. For constructing gridded rainfall, BMD

data from almost entire country's weather stations (54) are merged with rainfall estimates from the Climate Hazards Group InfraRed Precipitation (CHIRP; Funk et al., 2015). Compared with other available gridded precipitation products, ENACTS-BMD performs better in terms of monsoon total rainfall (Montes, Acharya, & Hassan, 2021). In this study, seasonal total rainfall for the period June through September (JJAS) are accumulated from daily data for the years 1982 to 2018. Figure 1 presents the climatology, interannual standard deviation and first empirical orthogonal function which (explains 44% of total variance) of total JJAS rainfall from ENACTS-BMD product during the study period.

2.2 | GCM hindcasts and forecasts

Hindcasts and forecasts from seven GCMs belonging to the NMME project phase 2 (Kirtman et al., 2014) were used in this study (details of each model can be found in the corresponding reference in Table 1). The NMME project coordinates intra-seasonal to interannual climate predictions from climate-modelling centres in the United States and Environment Canada. The NMME products provide opportunities to characterise and quantify the uncertainty associated with model structure and initial conditions using a large number of contributing models, each consisting of several ensemble members. The lead-1 (initial conditions of May for forecasting JJAS total rainfall) hindcasts (spanning 1982–2010) and forecasts (spanning 2011–2018) of seasonal total rainfall for the JJAS season from these seven GCMs were used. As the statistical post-processing process, especially CCA, required longer training sample, we have combined hindcast (29 years) and forecast runs (8 years; altogether 37 years) from these models, under the assumption that the hindcasts and forecasts are consistent with each other. These models have different number of ensemble members that were averaged to generate an ensemble mean and having a common 1° resolution spatial grid. These NMME monthly hindcast and forecast datasets were obtained from the Columbia University's International Research Institute's data library (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/).

3 | METHODOLOGY

As described in Section 1, we used a calibrated multi-model ensemble (CMME) approach. This approach involves calibrating individual GCMs using canonical correlation analysis (CCA) based regression and assessing their skill against raw GCM outputs. The calibrated GCMs are averaged (equal weighting) to make a final CMME time series. The CMME-based forecast is
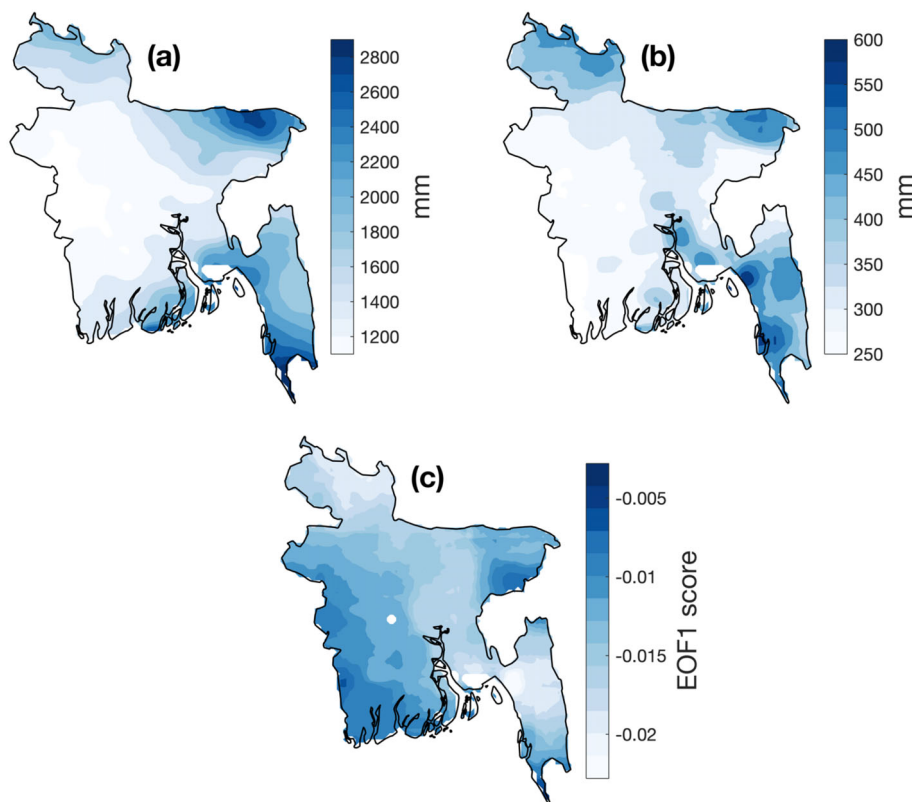
FIGURE 1   Maps of (a) climatology (mean), (b) interannual variability (standard deviation) and the first empirical orthogonal function (EOF1) of total June–September rainfall from ENACTS-BMD product. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1   List of the seven NMME models used, the responsible institutions, number of ensemble members and reference.

Model acronym	Institution	Number of ensemble members	References
CanSIPSv2	Canadian Centre for Meteorological and Environmental Prediction (CCMEP)	20	Lin et al. (2020)
COLA-RSMAS-CCSM4	The Center for Ocean-Land-Atmosphere Studies	10	Kirtman et al. (2014)
NASA-GEOS-S2S-2	National Aeronautics and Space Administration (NASA), Goddard Space Flight Center	4 (10[a])	Borovikov et al. (2017)
GFDL-CM2p1-aer04	Geophysical Fluid Dynamics Laboratory	10	Delworth et al. (2006), Zhang et al. (2007)
GFDL-CM2p5-FLOR-A06	Geophysical Fluid Dynamics Laboratory	12	Vecchi et al. (2014)
GFDL-CM2p5-FLOR-B01	Geophysical Fluid Dynamics Laboratory	12	Vecchi et al. (2014)
NCEP-CFSv2	NOAA's Centers for Environmental Prediction	24 (28[a])	Saha et al. (2019)

Note: The lead-1 (initial conditions of May for forecasting JJAS total rainfall) hindcasts (spanning 1982–2010) and forecasts (spanning 2011–2018) of seasonal total rainfall for the JJAS season from these seven GCMs were used.
[a]The value in parentheses shows the ensemble size for the forecast period.

subsequently compared against observations to assess its performance in relation to the uncalibrated forecasts. The processing chain is summarised in the flow chart presented in Figure 2.

3.1 | CCA-based calibration

CCA is widely used for calibration of forecasts from GCMs, for which the spatio-temporal patterns of GCM rainfall are projected onto the observed patterns (Barnston & Tippett, 2017; Tippett et al., 2007, 2008). CCA is basically a multivariate linear regression method allowing the identification of a sequence of pairs of patterns in two multivariate data sets, to then construct a set of transformed variables by projecting the original data onto these patterns. Correlations between the pairs of canonical variates, which are the transformed variables generated from truncated empirical orthogonal functions (EOF) or principal components (PC) of anomalies of
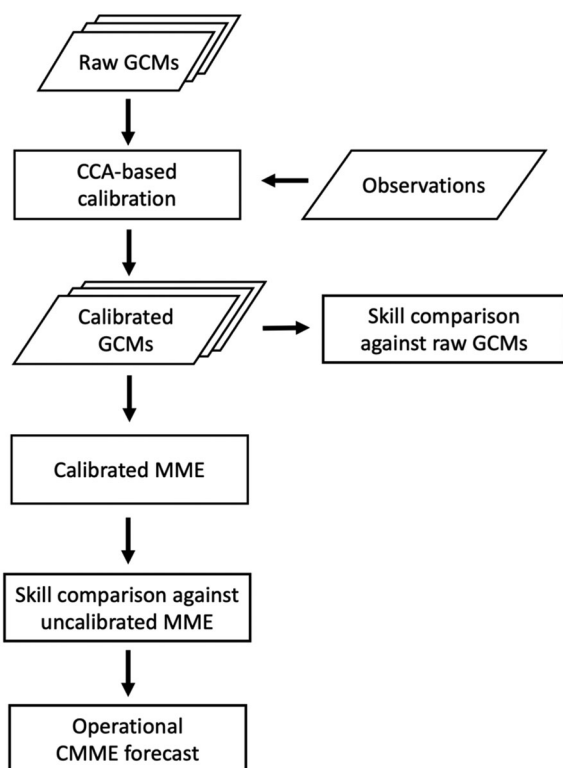
FIGURE 2 Flow chart illustrating the steps of generation of seasonal forecasts using the calibrated multi-model ensemble approach.

predictor and predictand data, are called canonical correlations. Linear regression between predictand-predictor canonical variates is used for the forecast. Finally, the predicted values are recovered by EOF synthesis and reconstructed from the predictand means and standard deviations. More details of CCA method can be found in Wilks (2020).

The CCA-based calibration has been carried out separately for ensemble mean of each GCM prior to producing multi-model ensembles. The full procedure consists of the following sequential steps:

- At the outset, observed rainfall was transformed to Gaussian by fitting a Gamma distribution. From estimates of the shape and scale parameters, the mean and variance of the corresponding Gaussian distribution are given in closed form.
- As pre-orthogonalisation, CCA requires truncation of the EOF or PC expansions of the GCM (the predictor) and on the corresponding observations (the predictand). To avoid overfitting due to small sample size to train CCA, we have pre-selected five PCs for GCM and observation. The total variance explained by 5 PC is 92% for observation and for GCMs it is on average 85% (as there are different GCMs).

- In CCA, the predictor domain is usually designed to be larger than the predictand domain, so that relevant features outside of the predictand domain can be used for better model calibration (Barnston & Tippett, 2017). Therefore, the spatial domains for the GCM predictor fields were taken to be 15° N–35° N, 80° E–100° E, and all the ENACTS-BMD grid points within Bangladesh (Figure 3) were considered as our predictand (Figure 3).
- The CCA model was trained using a leave-5-out cross-validation in the 37 years of dataset (1982–2018) in which 5 consecutive years are retained from both the pre-EOF and the CCA training sample from GCM and observation, and the middle year of the 5 is predicted. The years withheld progress from the earliest 5 to the latest 5 in which the first and the last years are also predicted so that each year has a cross-validated prediction.
- Finally, the cross-validated series for the predictand variable is generated for 37 years and then validated against the observed rainfall data using skill scores mentioned in the next section.

## 3.2 | Calibrated multi-model ensemble

Previous works have shown that the use of multi-model ensemble (MME) approaches improves the forecast skills from individual GCM (Acharya, Kar, et al., 2011; Acharya et al., 2014; Kar et al., 2012; Krishnamurti et al., 2009). In general, an MME can be generated by combining equally weighted ensemble members or weighted according to their prior performance (Acharya, Kar, et al., 2011; Kar et al., 2012; Wang et al., 2019; Weigel et al., 2008; Weigel et al., 2010). Studies shown that performance-based weighting does not bring significant differences compared to the equal weighting to make MME based on calibrated GCMs (Wang et al., 2019; Weigel et al., 2008). In this work, equally weighted calibrated GCMs were used to generate the MME following Acharya et al. (2021).

3.3 | Verification metrics

To examine the skill of uncalibrated, calibrated GCM and MME forecasts, two commonly used forecast verification metrics, that is, the Root Mean Square Error (RMSE), which corresponds to the average squared difference between the forecast and observation pairs, and the Spearman rank correlation coefficient, which is the Pearson's product–moment correlation on the ranked values for each variable. In Spearman's rank correlation, a monotonic relationship between two variables is an important underlying assumption and is less restrictive
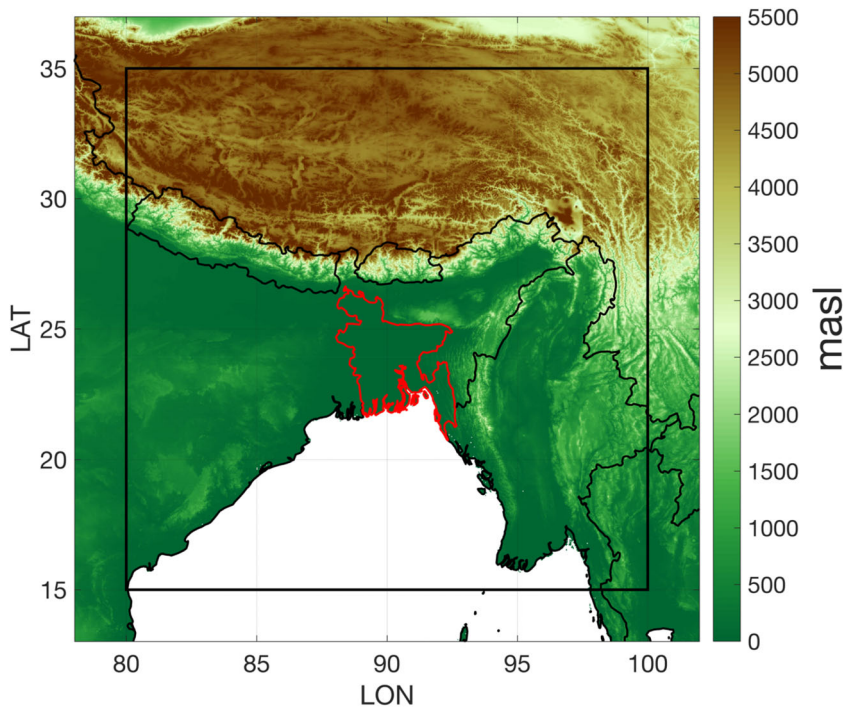
FIGURE 3 Topographic map showing the spatial domain used as predictor for general circulation models (black square), and Bangladesh (predictand area). Elevation data obtained from the Shuttle Radar Topography Mission SRTM90 digital elevation model. [Colour figure can be viewed at wileyonlinelibrary.com]
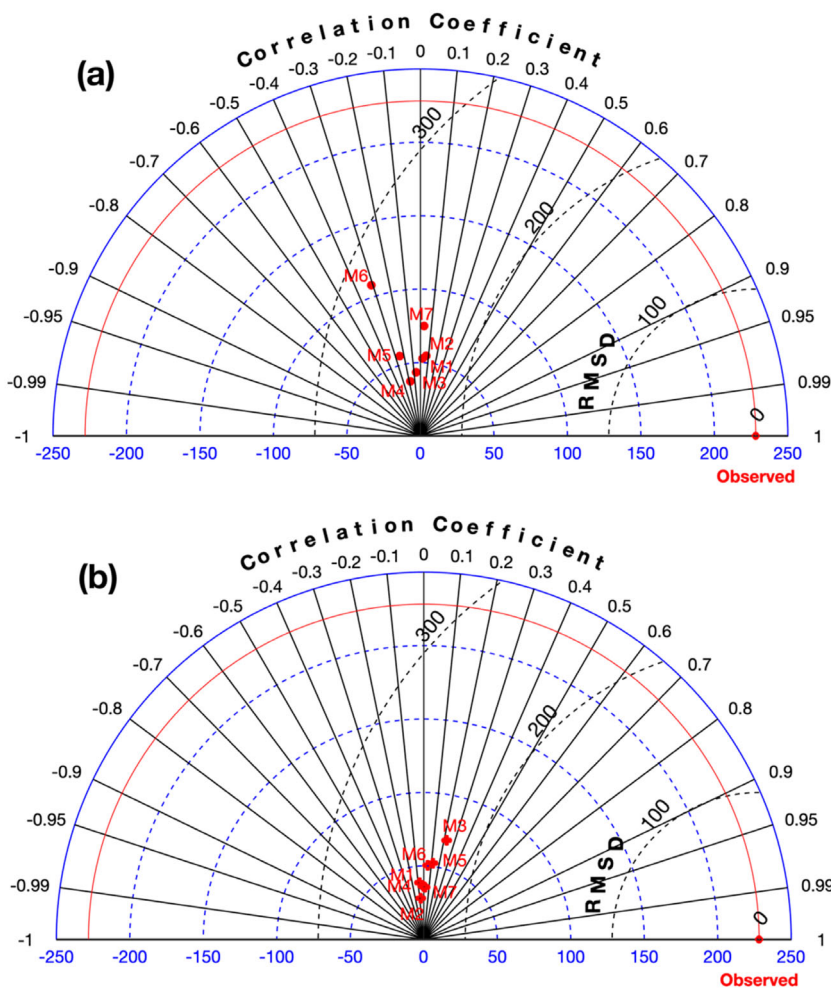
FIGURE 4 Taylor diagram for prediction skill of country-averaged time series of total June–September rainfall of (a) uncalibrated and (b) calibrated GCMs: CanSIPSv2 (M1), COLA-RSMAS-CCSM4 (M2), GFDL-CM2p5-FLOR-A06 (M3), GFDL-CM2p5-FLOR-B01 (M4), GFDL-CM2p1-aer04 (M5), NASA-GEOSS2S (M6), and NCEP-CFSv2 (M7). Blue dashed line represents the standard deviation, and black dashed line the root mean square difference (RMSD). The red line represents the standard deviation of the country-averaged time series of observed total precipitation. [Colour figure can be viewed at wileyonlinelibrary.com]

than that of a linear relationship, which must be met by Pearson's correlation. We also employed the 'generalised discrimination score', also known as 'two alternative forced-choice score' (2AFC score; Mason & Weigel, 2009). The 2AFC score measures the proportion or probability of a correct decision of all available pairs of observations of a differing category whose forecasts are discriminated in the correct direction. The score ranges between 0% and 100% while any value higher than 50% implies that the forecast is able to discriminate beyond random guessing. These verification measures are used in this study as they are recommended by WMO standardised verification system for long-range forecasts for skill assessment (WMO, 2018).

4 | RESULTS AND DISCUSSION

4.1 | Skill of uncalibrated GCM predictions

Before assessing the skill of the CMME-based prediction, the performance of each individual GCM was analysed.

A Taylor diagram (Taylor, 2001) summarising the country-averaged performance of total JJAS rainfall predicted by each GCM is displayed in Figure 4a. Each of the models does not perform well in terms of correlation with the observations, which varies between −0.3 and close to zero. Observed standard deviations are largely underestimated by the GCMs, which range from around 50 mm to 100 mm, with root mean square differences between 230 and 300 mm. In general, these models performed poorly in reproducing the observed variability in JJAS rainfall over Bangladesh. This performance is in agreement with a recent study by Kelley et al. (2020), where the skill of NMME models were examined in the context of sub-seasonal metrics prediction, and which described low-to-modest skill in predicting seasonal rainfall in Bangladesh.

These differences may be related to the model's coarser spatial resolution ($1° \times 1°$ grid) compared to higher resolution observed data ($0.05° \times 0.05°$ grid). Although large-scale anomalies can be predicted at such coarse resolution, details on rainfall heterogeneity over Bangladesh could not be resolved, which suggests that the downscaling of GCM
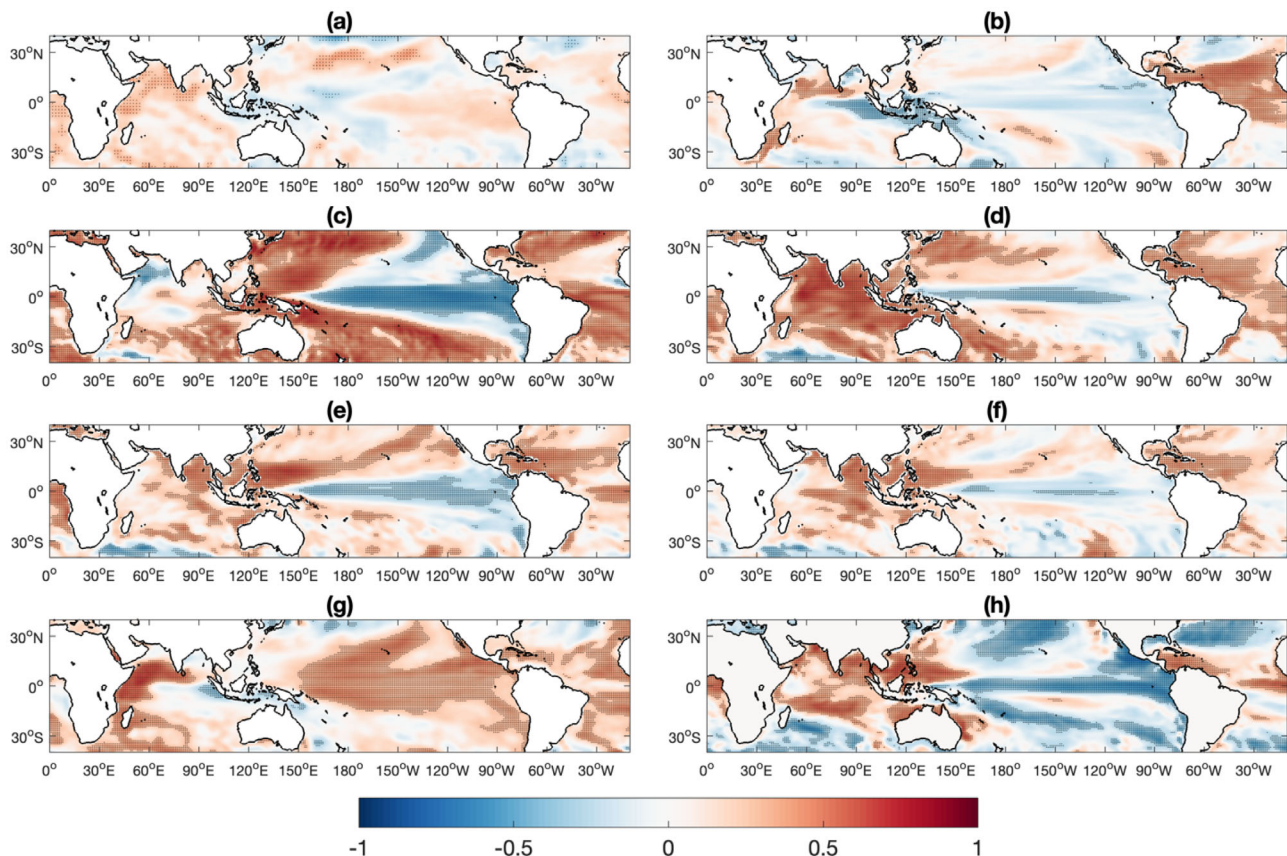

FIGURE 5 Correlation coefficient between (a) observed area averaged BSMR and global SST for JJAS and (b–h) model predicted rainfall with model SSTs. (b) CanSIPSv2, (c) NASA-GEOSS2S, (d) GFDL-CM2p1-aer04, (e) GFDL-CM2p5-FLOR-A06, (f) GFDL-CM2p5-FLOR-B01, (g) COLA-RSMAS-CCSM4, and (h) NCEP-CFSv2 over the period of 1982–2018. Dotted areas denote statistically significant correlations. [Colour figure can be viewed at wileyonlinelibrary.com]

outputs can be highly important. A possible hypothesis for this poor performance by GCMs that has been described as driving bias in GCMs forecasting is the over-sensitivity of GCMs to El Niño-Southern Oscillation (ENSO)-rainfall teleconnections (Acharya, Kar, et al., 2011; Pillai et al., 2018; Singh et al., 2019). To investigate this possibility, Pearson's correlation coefficients between area-averaged seasonal total rainfall over Bangladesh and global sea surface temperature (SST) have been computed for observed and predicted rain and SSTs in each model (Figure 5). In observation, the ENSO-rainfall teleconnection is found to be positive although it is not statistically significant. Rahman et al. (2013b) found the similar positive ENSO-rainfall teleconnection using observations from 1985 to 2008. In contrast, the ENSO-rainfall teleconnections in most of the GCMs indicate a strongly negative relationship, indicating that GCMs are unable to reproduce the observed teleconnections satisfactorily, even of an opposite sign. Although the CCSM4 model showed the similar signal (positively correlated) of teleconnection pattern as observed teleconnection pattern, the magnitude of correlation is highly positive and statistically significant. Previous studies evaluating NMME models for Indian monsoon also found that the ENSO-rainfall teleconnections in the GCMs are stronger than in the observation which is a potential reason for GCM's poor performances to simulate monsoon rainfall (Pillai et al., 2018; Singh et al., 2019). Additionally, studies also shown that the seasonal prediction of north-eastern Indian region including Bangladesh is very challenging due its positive ENSO (out-of-phase) teleconnection whereas the major part of Indian subcontinent has a negative relationship with ENSO (Choudhury et al., 2019; Saha et al., 2019). However, most of the GCM's cannot distinguish the out-of-phase relationship and having negative teleconnection with ENSO for the monsoon over entire Indian subcontinent. Other hypotheses of poor predictability by GCM can be drawn from the potential predictability (PP) analysis. Although there is a myriad of possible ways to estimate PP, we consider signal-to-noise ratio (SNR) to evaluate the predictive power of the models where the individual ensemble members from each of the models are taken into consideration (Figure 6). The SNR is used in several studies for the quantification of the predictive power of GCMs (Attada et al., 2022; Kang et al., 2004; Nair et al., 2013; Singh et al., 2012) for the Indian summer monsoon season. The SNR is defined as the ratio of external and internal variability where the external component is obtained as the variance of the ensemble mean and the internal component can be evaluated as the variance of noise (deviation of members from the ensemble mean). This implies that the larger the SNR, the better the predictive power. It can be noticed from Figure 6 that the except for the NASA-GEOSS2S, most of the GCMs

(CanSIPSv2, GFDL-CM2p5-FLOR-A06, GFDL-CM2p5-FLOR-B01, COLA-RSMAS-CCSM4, and NCEP-CFSv2) has SNR within 0–0.2 range which represents a very weak predictability (external variance is ~0%–4%). These lower SNR values explain the predictability limit for each GCM. This inability underscores the importance of calibration
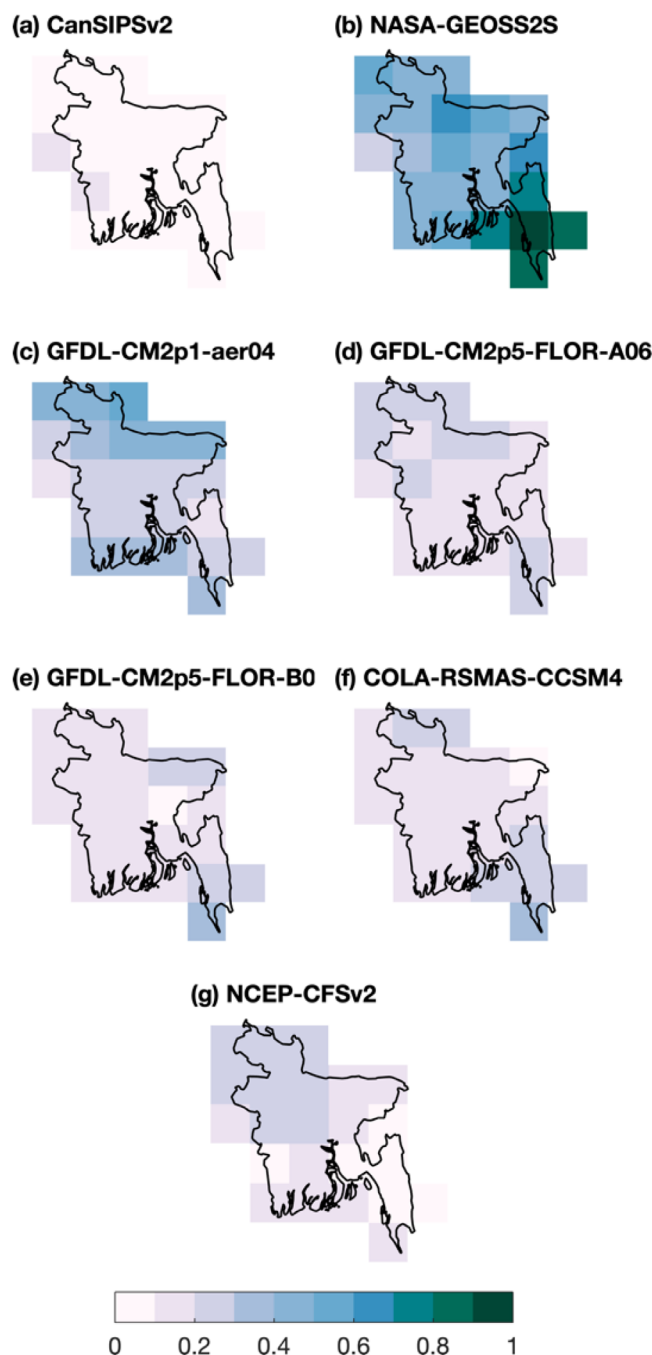
FIGURE 6 Signal-to-noise ratio of (a) CanSIPSv2, (b) NASA-GEOSS2S, (c) GFDL-CM2p1-aer04, (d) GFDL-CM2p5-FLOR-A06, (e) GFDL-CM2p5-FLOR-B01, (f) COLA-RSMAS-CCSM4, and (g) NCEP-CFSv2 over the period of 1982–2018. [Colour figure can be viewed at wileyonlinelibrary.com]

methods to partially or wholly remove systematic biases before computing a multi-model ensemble-based forecast. As described in Section 3.1, CCA-based calibration is useful in this regard as it projects the GCM rainfall onto the observed spatio-temporal patterns.

4.2 | Skill of calibrated GCMs predictions

The Taylor diagram of calibrated GCMs of Figure 4b shows that after calibration the root mean square differences range from 200 to 230 mm, representing an improvement over the uncalibrated GCMs. Moreover, it is also noticed the correlation also improved after calibration. For instance, the highly negative correlations between observations and models such as NASA-GEOSS2S and GFDL-CM2p1-aer04 become positively correlated after calibration. To examine the performance of the CCA-based calibration method at grid point scale, RMSE, Spearman's correlation coefficients and 2AFC scores are computed before and after calibration for each NMME model. For uncalibrated models, we interpolated GCMs to the ENACTS-BMD's resolution for a fair comparison as CCA produced the same resolution products as ENACTS-BMD. As similar results are found for all
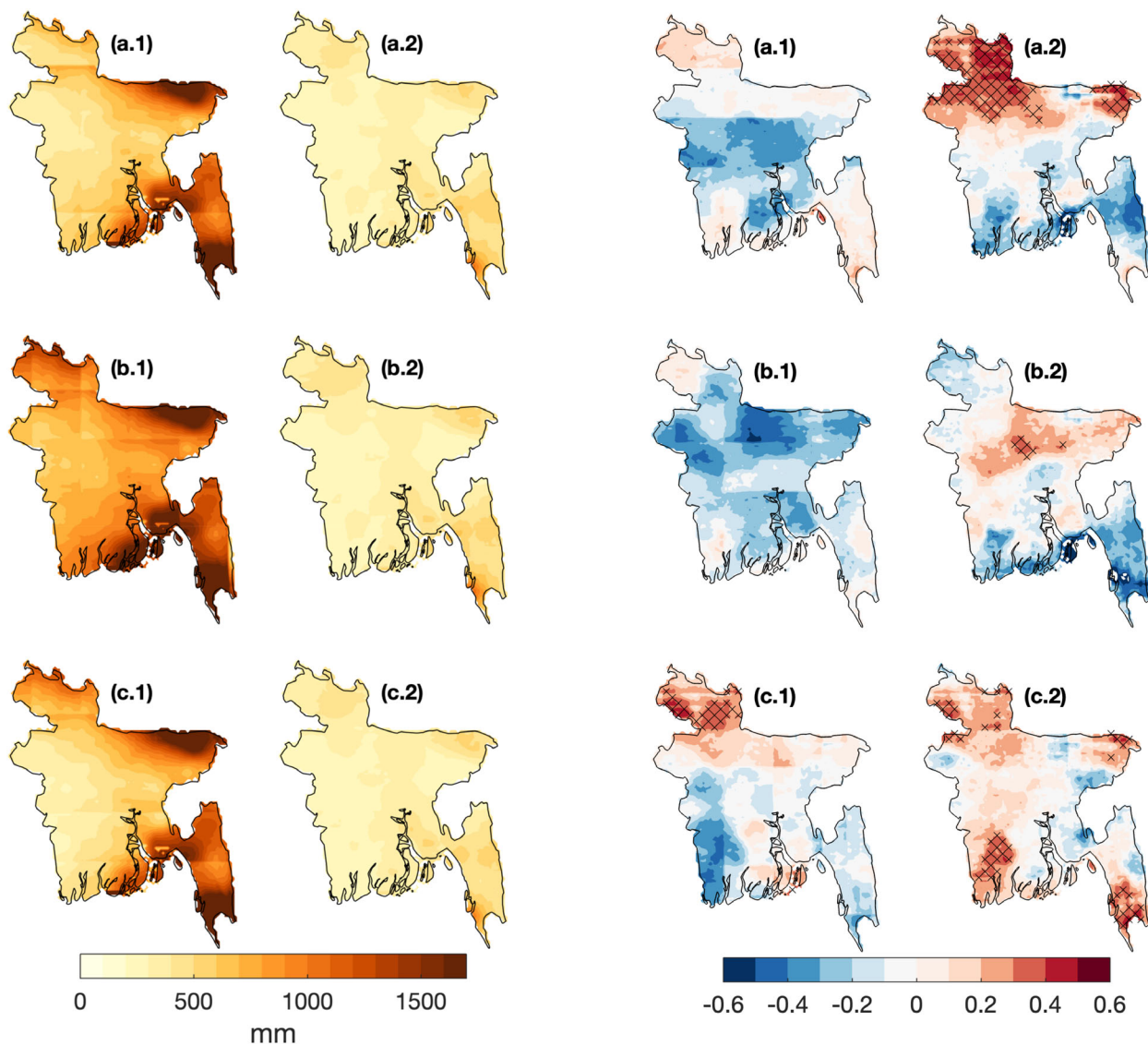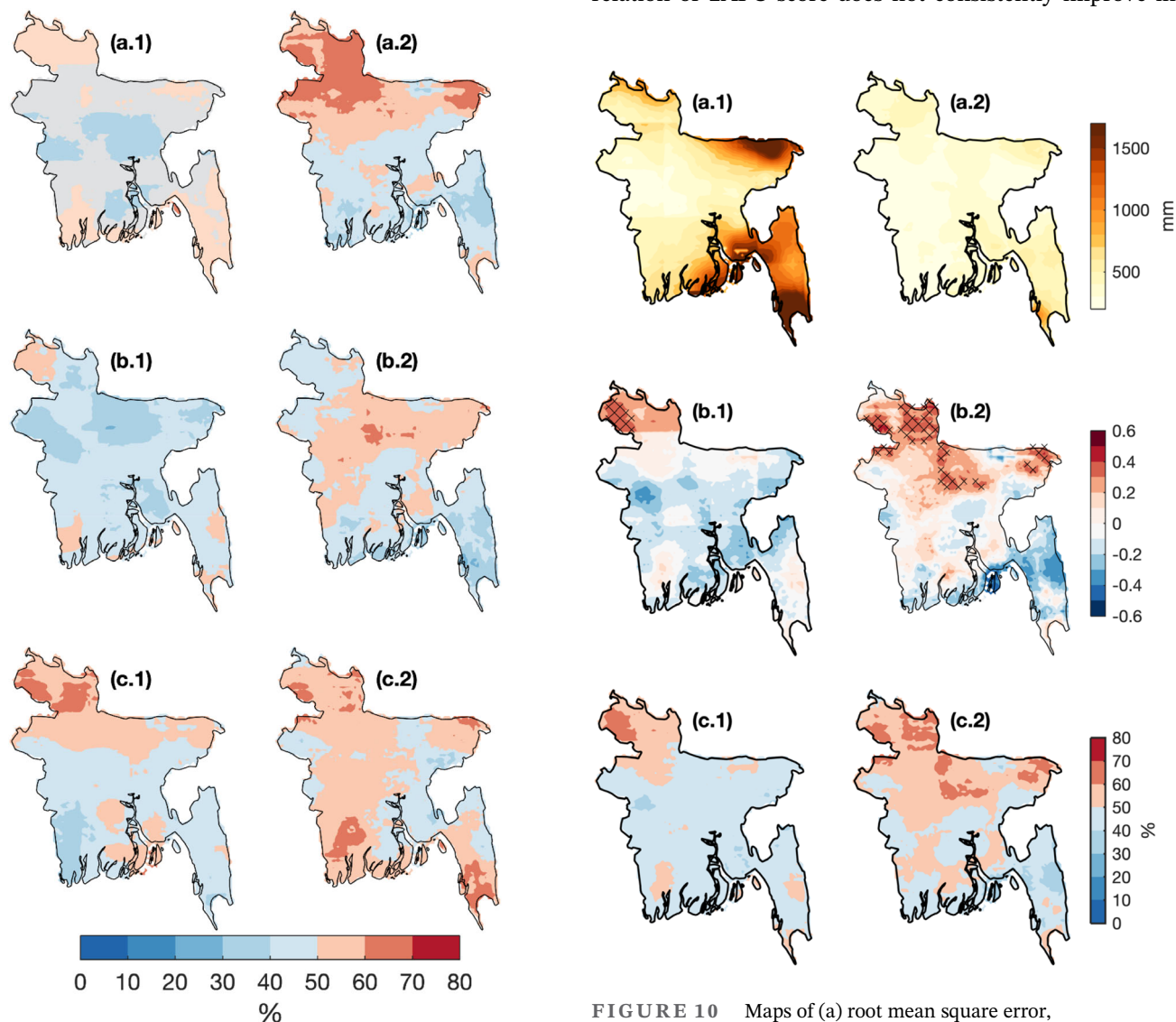

FIGURE 7 Maps of root mean square error for uncalibrated (left column) and calibrated (right column) total June–September rainfall (1982–2018) for (a) GFDL-CM2p5-aer04, (b) NASA-GEOSS2S, and (c) NCEP-CFSv2. Colour bar in mm/season. [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 8 Maps of Spearman correlation for uncalibrated (left column) and calibrated (right column) total June–September rainfall (1982–2018) for (a) GFDL-CM2p5-aer04, (b) NASA-GEOSS2S, and (c) NCEP-CFSv2. Dashed areas denote statistically significant correlations. [Colour figure can be viewed at wileyonlinelibrary.com]

NMME models, we selected GFDL-CM2p5-aer04, NASA-GEOSS2S and NCEP-CFSv2 models for illustrative purposes. The north-eastern and south-eastern portions of Bangladesh exhibit the highest RMSE, which correspond to the rainiest areas of the country (Figure 7). Notably, the calibration reduces the RMSE, with values below 200 mm over most of the country, except for rainier regions where RMSE is around 300 mm for most models. Calibrated models show higher skill in terms of correlation for most of the country area (Figure 8). The correlation coefficients of GFDL-CM2p5-aer04 and NASA-GEOSS2S before calibration are mostly negative, but in general, improved after calibration, except in south-eastern Bangladesh. Over southern and eastern parts of the country, NCEP-CFSv2 correlations turn from negative to positive. Also, positive correlations in the north are similar before and after calibration. For all models, the 2AFC score also improved: areas where 2AFC was less than 50% for uncalibrated model outputs became higher than 50% after calibration (Figure 9). Moreover, the spatial pattern of improvement is similar for 2AFC scores and Spearman's correlation coefficients.

In general, the CCA-based calibration improves the forecast skill of uncalibrated models. Moreover, when RMSE is used as the verification metric, CCA calibration appears to improve the forecasting skill strongly, but correlation or 2AFC score does not consistently improve in

FIGURE 9 Maps of generalised discrimination score (2AFC) for uncalibrated (left column) and calibrated (right column) total June–September rainfall (1982–2018) for GFDL-CM2p5-aer04 (a), NASA-GEOSS2S (b), and NCEP-CFSv2 (c). [Colour figure can be viewed at wileyonlinelibrary.com]
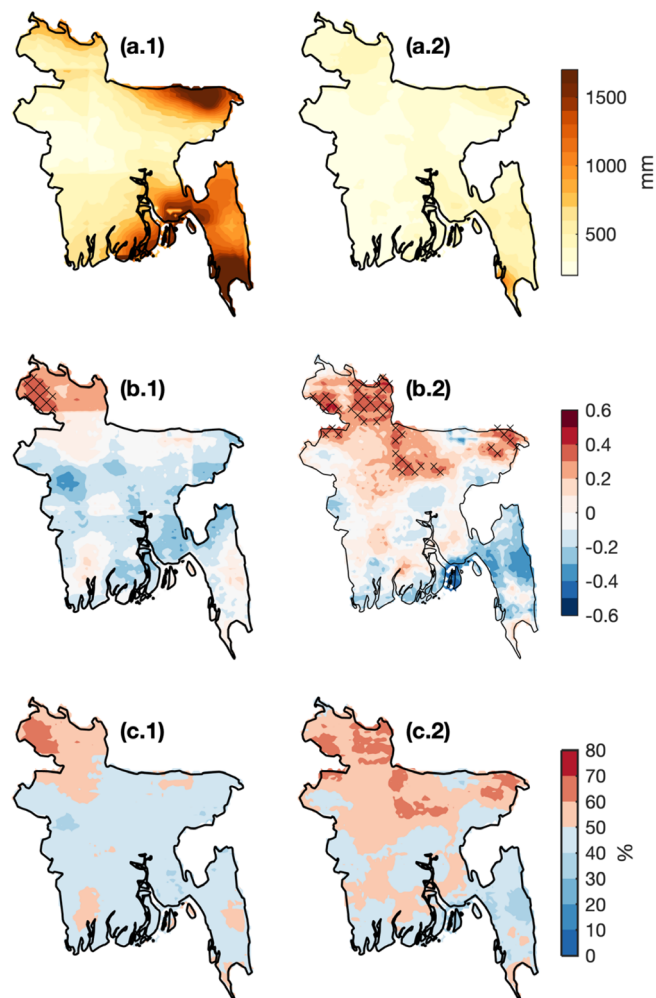
FIGURE 10 Maps of (a) root mean square error, (b) Spearman's correlation coefficient, and (c) Generalised discrimination score (2AFC) of total June–September rainfall (1982–2018) for uncalibrated (left column) and calibrated (right column) multi-model ensemble predictions. Dashed areas in (b) and (c) denote statistically significant correlations. [Colour figure can be viewed at wileyonlinelibrary.com]

every case, especially where the models show poor skill in the uncalibrated version, such as the case of NASA-GEOSS2S, which can be explained by the limited sample data to train the CCA.

4.3 | Skill of calibrated multi-model ensemble

To assess the performance of CMME, its skill is compared with uncalibrated MME, namely, UMME (averaging uncalibrated individual model) and presented in Figure 10. The skill of the UMME can be used as a benchmark. In general, CMME outperformed UMME in all skill scores. The RMSE is much lower in CMME, especially in north and south-eastern Bangladesh. Considering Spearman's correlation coefficient, UMME shows positive values only over a small area in the northern and drier areas of Bangladesh, whereas CMME shows widespread positive values except over a small area in the more mountainous southeastern part of the country where the correlations are close to zero or slightly negative. In addition, CMME Spearman's correlation coefficients are higher compared to most calibrated individual models. In terms of the 2AFC score, Figure 9c shows that values higher than 50% are dominant in CMME, except for the same region over the southeast. These results suggest an overall improvement of skill in BSMR prediction when CMME is used; however, high within-country differences are also observed, which can be associated with the complex local-scale precipitation mechanisms and the high spatial variability in climatological rainfall in Bangladesh.

5 | CONCLUDING REMARKS

This study aimed to develop an improved seasonal forecast system based on calibrated multi-model ensemble for the prediction of BSMR. For this purpose, we developed a hybrid dynamical–statistical technique using state-of-the-art GCMs from the NMME project. The individual GCM's seasonal predictions have been calibrated using a CCA approach to correct large systematic biases. These calibrated individual model predictions were then combined with equal weighting to obtain the final CMME forecast. Although similar multi-model prediction approaches have been used extensively, to the best of our knowledge, this is the first time that it has been used to produce seasonal forecasts of the BSMR. Since October 2019, each month this CMME-based forecast is prepared in real-time by the BMD for the next season. Therefore, from an operational perspective, the potential benefits of

such a forecasting system need to be illustrated and documented in terms of the gain in quality of forecasts in real-time. Although this study only focuses on the skill of this forecast system for the summer monsoon season as the primary period of precipitation in Bangladesh, additional research should also document the predictability of pre- and post-monsoon precipitation, as well the applicability of our predictions for practical climate services in Bangladesh.

In conclusion, we found that although GCMs provide a solid non-linear approach to alternative statistical modelling to predict the BSMR, the calibration of models is necessary to generate operational forecasts given the strong model biases over Bangladesh. The biased performance of GCMs may be partly related to the model's coarse spatial resolution, their over-sensitivity to SST-rainfall teleconnections and lower signal-to-noise ratio which explains the predictability limit. Our results strongly indicate that CCA-based calibration can generate significant improvements that reduce the magnitude of systematic errors (RMSE) compared to individual uncalibrated models. Calibration also appears to improve Spearman's correlation coefficients and 2AFC scores over most of Bangladesh, exempting a few locations in the north- and south-east of the country. In conclusion, our analysis demonstrates that the skill of CMME is much better than the UMME and in comparison, to individual calibrated models, especially in the northern part of the country. However, due to limited sample data to train the CCA (32 years; using leave-5-out cross-validation in 37 years of hindcast data), further room for skill improvement which would be the subject of future research and will require a large sample to achieve increased robustness.

AUTHOR CONTRIBUTIONS
Nachiketa Acharya: Conceptualization; data curation; formal analysis; investigation; methodology; validation; writing – original draft; writing – review and editing. **Carlo Montes:** Visualization; writing – review and editing. **S. M. Q. Hassan:** Writing – review and editing. **Razia Sultana:** Writing – review and editing. **Md. Bazlur Rashid:** Writing – review and editing. **Md. Abdul Mannan:** Writing – review and editing. **Timothy J. Krupnik:** Funding acquisition; writing – review and editing.

ORCID

Nachiketa Acharya https://orcid.org/0000-0003-3010-2158

Md. Bazlur Rashid https://orcid.org/0000-0003-1789-6379

REFERENCES

Acharya, N., Chattopadhayay, S., Mohanty, U.C., Dash, S.K. & Sahoo, L.N. (2013) On the bias correction of general circulation model output for Indian summer monsoon. *Meteorological Applications*, 20(3), 349–356.

Acharya, N., Dinku, T. & Hall, K.J.C. (2020) A next generation (NextGen) approach to improve the seasonal prediction system in East Africa. Earth and Space Science Open Archive. https://doi.org/10.1002/essoar.10504989.1

Acharya, N., Ehsan, M.A., Admasu, A., Teshome, A. & Hall, K.J.C. (2021) On the next generation (NextGen) seasonal prediction system to enhance climate services over Ethiopia. *Climate Services*, 24, 100272. Available from: https://doi.org/10.1016/j.cliser.2021.100272

Acharya, N., Faniriantsoa, R., Rashid, B., Sultana, R., Montes, C., Dinku, T. et al. (2020) Developing high-resolution gridded rainfall and temperature data for Bangladesh: the ENACTS-BMD dataset. *Dew-Drop*, 7(1), 155–164.

Acharya, N., Kar, S.C., Kulkarni, M.A., Mohanty, U.C. & Sahoo, L.N. (2011) Multi-model ensemble schemes for predicting northeast monsoon rainfall over peninsular India. *Journal of Earth System Science*, 120(5), 795–805.

Acharya, N., Kar, S.C., Mohanty, U.C., Kulkarni, M.A. & Dash, S.K. (2011) Performance of GCMs for seasonal prediction over India—a case study for 2009 monsoon. *Theoretical and Applied Climatology*, 105(3–4), 505–520.

Acharya, N., Srivastava, N.A., Panigrahi, B.K. & Mohanty, U.C. (2014) Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. *Climate Dynamics*, 43(5), 1303–1310.

Ahasan, M.N., Chowdhary, M.A.M. & Quadir, D.A. (2010) Variability and trends of summer monsoon rainfall over Bangladesh. *Journal of Hydrology and Meteorology*, 7, 1–7.

Ahmed, M.K., Alam, M.S., Yousuf, A.H.M. & Islam, M.M. (2017) A long-term trend in precipitation of different spatial regions of Bangladesh and its teleconnections with El Niño/southern oscillation and Indian Ocean dipole. *Theoretical and Applied Climatology*, 129, 473–486. Available from: https://doi.org/10.1007/s00704-016-1765-2

Attada, R., Ehsan, M.A. & Pillai, P.A. (2022) Evaluation of potential pre-dictability of Indian summer monsoon rainfall in ECMWF'sFifth-generation seasonal forecast system (SEAS5). *Pure and Applied Geophysics*, 179, 4639–4655. Available from: https://doi.org/10.1007/s00024-022-03184-9

Banik, S., Anwer, M., Khan, A.F.M.K., Rouf, R.A. & Chanchary, F.H. (2009) Forecasting Bangladeshi monsoon rainfall using neural network and genetic algorithm approaches. *International Technology Management Review*, 2(1), 1100–1107.

Bari, S.H., Rahman, M.T., Hussain, M.M. & Ray, S. (2015) Forecasting monthly precipitation in Sylhet city using ARIMA model. *Civil and Environmental Research*, 7, 69–77.

Barnston, A.G. & Tippett, M.K. (2017) Do statistical pattern corrections improve seasonal climate predictions in the North American multimodel ensemble models? *Journal of Climate*, 30(20), 8335–8355. Available from: https://doi.org/10.1175/JCLI-D-17-0054.1

Borovikov, A., Cullather, R., Kovach, R., Marshak, J., Vernieres, G., Vikhliaev, Y. et al. (2017) GEOS-5 seasonal forecast system. *Climate Dynamics*, 1–27, 7335–7361. Available from: https://doi.org/10.1007/s00382-017-3835-2

Cash, B.A., Barimalala, R., Kinter, J.L., Altshuler, E.L., Fennessy, M.J., Manganello, J.V. et al. (2017) Sampling variability and the changing ENSO–monsoon relationship. *Climate Dynamics*, 48, 4071–4079. Available from: https://doi.org/10.1007/s00382-016-3320-3

Choudhury, B.A., Saha, S.K., Konwar, M., Sujith, K. & Deshamukhya, A. (2019) Rapid drying of Northeast India in the last three decades: climate change or natural variability? *Journal of Geophysical Research*, 124, 227–237.

Dahale, S.D. & Singh, S.V. (1993) Modelling of Indian monsoon rainfall time series by univariate box-Jenkins type of models. *Advances in Atmospheric Sciences*, 10, 211–220.

Delworth, T.L., Broccoli, A.J., Rosati, A., Stouffer, R.J., Balaji, V., Beesley, J.A. et al. (2006) GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate*, 19, 643–674. Available from: https://doi.org/10.1175/JCLI3629.1

Doblas-Reyes, F.J., Hagedorn, R. & Palmer, T.N. (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting—II. Calibration and combination. *Tellus A: Dynamic Meteorology and Oceanography*, 57, 234–252. Available from: https://doi.org/10.1007/s10584-10018-12267-10582

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S. et al. (2015) The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes. *Scientific Data*, 2, 150066.

Hansen, J.W., Challinor, A., Ines, A., Wheeler, T. & Moron, V. (2006) Translating climate forecasts into agricultural terms: advances and challenges. *Climate Research*, 33, 27–41.

Hossain, M., Kalam Azad, M., Karmakar, S., Islam Mondal, M., Das, M., Rahman, M.M. et al. (2019) Assessment of better prediction of seasonal rainfall by climate predictability tool using Global Sea surface temperature in Bangladesh. *Asian Journal of*

Advanced Research and Reports, 4(4), 1–13. Available from: https://doi.org/10.9734/ajarr/2019/v4i430116

Huq, S. (2001) Climate change and Bangladesh. *Science*, 294, 1617.

IRI. (2020) *NextGen 2-pager*. New York: IRI. https://iri.columbia.edu/wp-content/uploads/2020/06/Fact-Sheet_Next-Gen_small.pdf

Kang, I.S., Lee, J. & Park, C.K. (2004) Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. *Journal of Climate*, 17, 834–844.

Kang, I.-S., & Shukla, J. (2005) Dynamical seasonal prediction and predictability of monsoon. In: Wang, B. (Ed.), *The Asian Monsoon*. Chichester, UK: Praxis Pub. Ltd, pp. 585–612.

Kar, S.C., Acharya, N., Mohanty, U.C. & Kulkarni, M.A. (2012) Skill of monthly rainfall forecasts over India using multi-model ensemble schemes. *International Journal of Climatology*, 32, 1271–1286.

Kelley, C., Acharya, N., Montes, C., Krupnik, T., Abdul, M. & Hassan, S.Q. (2020) Exploring the predictability of subseasonal rainfall statistics during the South Asian summer monsoon in Bangladesh using NMME seasonal rainfall forecasts. *Theoretical and Applied Climatology*, 141, 495–508.

Kirtman, B.P., Min, D., Infanti, J.M., Kinter, J.L., III, Paolino, D.A., Zhang, Q. et al. (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95, 585–601. Available from: https://doi.org/10.1175/BAMS-D-12-00050.1

Krishnamurti, T.N., Mishra, A.K., Chakraborty, A. & Rajeevan, M. (2009) Improving global model precipitation forecasts over India using downscaling and the FSU superensemble. Part I: 1–5-day forecasts. *Monthly Weather Review*, 137, 2713–2735.

Lin, H., Merryfield, W.J., Muncaster, R., Smith, G.C., Markovic, M., Dupont, F. et al., (2020) The Canadian Seasonal to Interannual Prediction System Version 2 (CanSIPSv2). *Weather and Forecasting*, 35(4), 1317–1343.

Mahmud, I., Sheikh, H.B. & Rahman, M. (2017) Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method. *Environmental Engineering Research*, 22(2), 162–168.

Mannan, M.A., Mannan Chowdhury, M.A. & Karmakar, S. (2015) Long-range and short-range prediction of Rainfall and Rainy days over northwestern part of Bangladesh during Monsoon Season. Conference: P-ID 31 *International Conference on Physics Sustainable Development & Technology* (*ICPSDT-2015*) (August 19–20, 2015) Department of Physics, CUET, p. 47.

Mason, S.J. & Weigel, A.P. (2009) A generic forecast verification framework for administrative purposes. *Monthly Weather Review*, 137, 331–349.

Mohsin, M., Akhter, Y. & Begum, M. (2012) Modeling rainfall in Dhaka division of Bangladesh using time series analysis. *Journal of Mathematical Modelling and Application*, 1, 67–73.

Montes, C., Acharya, N., Hassan, S.M.Q. & Krupnik, T.J. (2021) Intense precipitation events during the monsoon season in Bangladesh as captured by satellite-based products. *Journal of Hydrometeorology*, 22(6), 1405–1419.

Montes, C., Acharya, N. & Hassan, S.Q.M. (2021) Assessing the performance of satellite-based products for monitoring extreme rainfall events in Bangladesh. *Dew-Drop*, 7(1), 70–81.

Montes, C., Acharya, N., Hossain, P.R., Babu, T.S.A., Krupnik, T.J. & Hassan, S.Q.M. (2022) Developing a framework for an early warning system of seasonal temperature and rainfall tailored to aquaculture in Bangladesh. *Climate Services*, 26, 100292.

Montes, C., Acharya, N., Stiller-Reeve, M.A., Kelley, C. & Hassan, S.Q.M. (2021) Interannual variability of monsoon onset and withdrawal in Bangladesh. *Atmospheric Science Letters*, 22(11), e1069. Available from: https://doi.org/10.1002/ASL.1069

Nahar, A., Luckstead, J., Wailes, E.J. & Alam, M.J. (2018) An assessment of the potential impact of climate change on rice farmers and markets in Bangladesh. *Climate Change*, 150, 289–304.

Nair, A., Acharya, N., Singh, A., Mohanty, U.C. & Panda, T.C. (2013) On the predictability of northeast monsoon rainfall over south peninsular India in general circulation models. *Pure and Applied Geophysics*, 170(11), 1945–1967. Available from: https://doi.org/10.1007/s00024-012-0633-y

Nashwan, M.S., Shahid, S. & Wang, X. (2019) Uncertainty in estimated trends using gridded rainfall data: a case study of Bangladesh. *Water*, 11, 349.

Pillai, P.A., Rao, S.A., Ramu, D.A., Pradhan, M. & George, G. (2018) Seasonal prediction skill of Indian summer monsoon rainfall in NMME models and monsoon mission CFSv2. *International Journal of Climatology*, 38, e847–e861.

Rahman, M., Rafiuddin, M. & Alam, M. (2013a) Seasonal forecasting of Bangladesh summer monsoon rainfall using simple multiple regression model. *Journal of Earth System Science*, 122(2), 551–558.

Rahman, M., Rafiuddin, M. & Alam, M. (2013b) Teleconnections between Bangladesh summer monsoon rainfall and sea surface temperature in the Indian Ocean. *The International Journal of Ocean and Climate Systems*, 4(4), 231–237.

Rahman, R. & Lateh, H. (2015) Climate change in Bangladesh: a spatio-temporal analysis and simulation of recent temperature and rainfall data using GIS and time series analysis model. *Theoretical and Applied Climatology*, 128, 27–41.

Rajeevan, M., Unnikrishnan, C.K. & Preethi, B. (2012) Evaluation of the ENSEMBLES multi-model seasonal forecasts of Indian summer monsoon variability. *Climate Dynamics*, 38, 2257–2274.

Robertson, A.W., Moron, V. & Swarinoto, Y. (2009) Seasonal predictability of daily rainfall statistics over Indramayu district, Indonesia. *International Journal Climatology*, 29, 1449–1462.

Saha, S.K., Hazra, A., Pokhrel, S., Chaudhari, H.S., Sujith, K., Rai, A. et al. (2019) Unraveling the mystery of Indian summer monsoon prediction: improved estimate of predictability limit. *Journal of Geophysical Research – Atmospheres*, 124, 1962–1974. Available from: https://doi.org/10.1029/2018JD030082

Singh, A., Acharya, N., Mohanty, U.C., Robertson, A.W. & Mishra, G. (2012) On the predictability of Indian summer monsoon rainfall in general circulation model at different lead time. *Dynamics of Atmospheres and Oceans*, 58, 108–127. Available from: https://doi.org/10.1016/j.dynatmoce.2012.09.004

Singh, B., Cash, B. & Kinter, J.L., III. (2019) Indian summer monsoon variability forecasts in the North American multimodel ensemble. *Climate Dynamics*, 53, 7321–7334. Available from: https://doi.org/10.1007/s00382-018-4203-6

Taylor, K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, 106(D7), 7183–7192.

Tippett, M., Barnston, A. & Robertson, A. (2007) Estimation of seasonal precipitation tercile-based categorical probabilities from ensembles. *Journal of Climate*, 20, 2210–2228.

Tippett, M., DelSole, T., Mason, S. & Barnston, A.G. (2008) Regression based methods for finding coupled patterns. *Journal of Climate*, 21, 4384–4398. Available from: https://doi.org/10.1175/2008JCLI2150.1

Vecchi, G.A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg, A.T. et al. (2014) On the seasonal forecasting of regional tropical cyclone activity. *Journal of Climate*, 27, 7994–8016.

Wang, H.-M., Chen, J., Xu, C.-Y., Chen, H., Guo, S., Xie, P. et al. (2019) Does the weighting of climate simulations result in a better quantification of hydrological impacts? *Hydrology and Earth System Sciences*, 23, 4033–4050. Available from: https://doi.org/10.5194/hess-23-4033-2019

Weigel, A.P., Knutti, R., Liniger, M.A. & Appenzeller, C. (2010) Risks of model weighting in multimodel climate projections. *Journal of Climate*, 23(15), 4175–4191.

Weigel, A.P., Liniger, M.A. & Appenzeller, C. (2008) Can multimodel combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society*, 134, 241–260.

Wilks, D. (2002) Smoothing forecast ensembles with fitted probability distributions. *Quarterly Journal of the Royal Meteorological Society*, 128, 2821–2836.

Wilks, D. (2017) Enforcing calibration in ensemble postprocessing. *Quarterly Journal of the Royal Meteorological Society*, 144(710), 76–84.

Wilks, D.S. (2020) *Statistical methods in atmospheric sciences*, 4th edition. Cambridge, MA: Elsevier. Available from: https://doi.org/10.1016/C2017-0-03921-6

World Meteorological Organization (WMO). (2018) Guidance on Verification of Operational Seasonal Climate Forecasts, WMO-No. 1220. https://library.wmo.int/doc_num.php?explnum_id=4886.

World Meteorological Organization (WMO). (2020) Guidance on operational practices for objective seasonal forecasting (2020). https://wmoomm.sharepoint.com/:b:/s/wmocpdb/EQeDmiRg-QZBvwAYIa0Z83EBGK3GcN5Ye73nNaVAvam3hg?e=h2lAL7.

World Meteorological Organization (WMO). (2017) International Workshop on Global Review of Regional Climate Outlook Forums. http://www.wmo.int/pages/prog/wcp/wcasp/meetings/documents/rcofs2017/Report_RCOF_Review_2017_final.pdf

Zhang, S., Harrison, M.J., Rosati, A. & Wittenberg, A., (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather. Review*, 135(10), 3541–3564.