Assessing Severe Drought and Wet Events over India in a Future Climate Using a Nested Bias-Correction Approach

Richa Ojha[1]; D. Nagesh Kumar, M.ASCE[2]; A. Sharma[3]; and R. Mehrotra[4]

Abstract: General circulation models (GCMs) are routinely used to simulate future climatic conditions. However, rainfall outputs from GCMs are highly uncertain in preserving temporal correlations, frequencies, and intensity distributions, which limits their direct application for downscaling and hydrological modeling studies. To address these limitations, raw outputs of GCMs or regional climate models are often bias corrected using past observations. In this paper, a methodology is presented for using a nested bias-correction approach to predict the frequencies and occurrences of severe droughts and wet conditions across India for a 48-year period (2050–2099) centered at 2075. Specifically, monthly time series of rainfall from 17 GCMs are used to draw conclusions for extreme events. An increasing trend in the frequencies of droughts and wet events is observed. The northern part of India and coastal regions show maximum increase in the frequency of wet events. Drought events are expected to increase in the west central, peninsular, and central northeast regions of India. **DOI: 10.1061/(ASCE)HE.1943-5584.0000585.** © *2013 American Society of Civil Engineers.*

CE Database subject headings: Droughts; India; Climates; Hydrologic models.

Author keywords: Nested bias correction; General circulation model; Drought; Wet; Standardized Precipitation Index; Climate; Precipitation; India.

Introduction

Impact-assessment studies related to climate change have underlined the sensitivity of water resources to climatic variations. Climatic changes are expected to cause increase in temperatures and changes in precipitation patterns and other climatic variables across the globe (Houghton et al. 1990). The tremendous importance of water in both society and nature underscores the necessity of understanding the impacts of changes in global climate over water resources at local regional scales.

General circulation models (GCMs) constitute important tools for assessing plausible impacts of climate change over a range of human and natural systems. Although GCMs perform relatively well at continental and large regional scales, their ability to simulate climate at finer spatial scales is still limited. Simulations at these finer scales are of considerable interest to hydrologists for assessing the possible impacts of climate change on water supply and related attributes. This has led to the development of a suite of statistical and dynamical downscaling methods that use coarse-scale GCM atmospheric simulations and yield estimates of variables at finer scales of interest (Anandhi et al. 2008; Sharma et al. 2011).

However, uncertainties associated with the formulation of GCMs lead to predictions that are biased when compared with observations. Specifically, GCM results show a large degree of uncertainty for key hydrologic variables such as precipitation, evapotranspiration, and runoff (Xu 1999). The models tend to incorrectly estimate rainfall occurrences and rainfall intensities at finer spatial and temporal scales. The interannual variability observed in precipitation patterns is affected by regional and global teleconnections, and the extent of this variability is nonuniform over different parts of the globe (Johnson and Sharma 2012). The GCMs tend to imperfectly model the effects of these teleconnections. Their inability to represent interannual variability in precipitation patterns results in poor simulations and biased estimates. Weather conditions in India are greatly influenced by the El Niño Southern Oscillation (ENSO) that cause widespread floods and droughts (Maity et al. 2007; Maity and Nagesh Kumar 2008; Dhanya and Nagesh Kumar 2009). The frequency, intensity, and impacts of such extreme events vary dramatically with geographical location over the Indian subcontinent. Addressing the interannual variability in precipitation patterns becomes an important concern over India. Thus, even though of prime importance for hydrologic studies, reliable prediction of future hydrologic variables from GCMs is very challenging.

Nested bias correction (NBC) (Johnson and Sharma 2011, 2012) is a recently developed, relatively simple approach that compensates for some of the shortcomings of GCM-predicted rainfall values, corrects for systematic bias at multiple timescales, and can be used to predict rainfall over different regions. In the present paper, the NBC approach is adopted to produce more reliable GCM predictions of droughts and wet events over India.

The main objectives of this paper are twofold: first, to highlight the utility of the NBC approach to correct the biases in the GCMs' rainfall outputs, and second, to predict the frequency of severe drought and wet conditions for a future climate scenario over India using the Standardized Precipitation Index (SPI).

The remainder of this paper is organized as follows: in the "Background" section, details are provided for some of the earlier work

[1]Dept. of Civil Engineering, Indian Institute of Science, Bangalore 560012, India.

[2]Dept. of Civil Engineering, Indian Institute of Science, Bangalore 560012, India; and Centre for Earth Sciences, Indian Institute of Science, Bangalore, India (corresponding author). E-mail: nagesh@civil.iisc.ernet.in

[3]School of Civil and Environmental Engineering, Univ. of New South Wales, Sydney, NSW 2052, Australia.

[4]School of Civil and Environmental Engineering, Univ. of New South Wales, Sydney, NSW 2052, Australia.

related to analysis and prediction of droughts and floods, the need for bias correction in GCM outputs, and the methodology adopted. In the "Methodology" section, an application of the methodology, followed by the discussion of study results in the "Results and Discussion" section. Finally, the study conclusions are presented.

Background

In this section, some of the work related to bias-correction techniques for GCM simulations and use of the SPI to characterize drought and wet conditions are reviewed. This is followed by a description of the methodology adopted in this paper.

Correction for Bias in GCM outputs

Significant systematic biases between modeled GCM outputs and observations require statistical correction of GCM-derived variables. Statistical bias-correction techniques allow the use of GCM outputs directly in hydrologic studies, while accepting that there are limitations in GCMs. When combined with spatial disaggregation, bias-correction techniques can provide model inputs at a range of scales suitable for hydrologic studies (Johnson and Sharma 2012; Mehrotra and Sharma 2010).

In a previous study, statistical bias correction was applied for correcting climate-model output to produce internally consistent fields that have the same statistical intensity distribution as observations (Piani et al. 2010). The study concluded that spatial distributions of time-based statistics of daily precipitation from climate models are significantly and consistently improved by a solely intensity-based statistical bias-correction method. In another study, several bias-correction methods were estimated for ensemble streamflow volume forecasts (Hashino et al. 2007). Specifically, an event-based bias-correction method, a regression method, and a quantile-mapping method were used for correcting the bias in historical weather data. It was observed that bias-correction methods achieved better skill by reducing the unconditional bias and increasing the potential skill of probabilistic forecasts. Three simple statistical downscaling approaches were compared in a study for a future climate scenario (Wood et al. 2004). The bias in monthly model data was corrected using quantile-mapping method. The authors of that study observed that bias correction yielded consistently plausible streamflow simulations, whether or not dynamical downscaling was used. Dynamic downscaling of model results without bias correction generated results that showed greater hydrologic sensitivity to climate change.

The bias-correction approaches mentioned so far mainly focus on either monthly or daily statistics. However, these approaches allowed for bias correction to be implemented at one timescale only. Further, even the bias-corrected results frequently failed to capture the interannual variability that is prevalent in observations. A previously published paper proposed the NBC methodology that preserves the interannual variability by applying bias correction simultaneously at two timescales (monthly and annually) (Johnson and Sharma 2012). Subsequently, a comparative study of two scaling and four bias-correction approaches including NBC was performed, and concluded that NBC corrects for the interannual variability to a greater extent than the other methods (Johnson and Sharma 2011). This interannual variability in hydrologic variables across India plays a strong role in identifying extreme events at long timescales (greater than 1 year). The NBC on atmospheric variables was applied as a postprocessor of GCM-simulated atmospheric variables before applying a stochastic downscaling to simulate rainfall in a future climate over the Malaprabha catchment in India (Mehrotra et al. 2011). Therefore, NBC was identified as a

suitable bias-correction approach when using GCM precipitation outputs for SPI calculations in this study. The need for SPI assessment is discussed in the next subsection.

Standard Precipitation Index

Even though several indexes are available for drought characterization (Mishra and Singh 2010), the SPI, developed by McKee et al. (1993, 1995), was adopted in this study because of its flexibility, simplicity, and its ability to classify both wet and dry conditions. The SPI classification is sample dependent and involves transformation of the precipitation time series into a standardized normal distribution (z distribution).

Numerous studies have adopted SPI for drought classification. For example, Bussay et al. (1999) and Szalai and Szinell (2000) assessed the utility of the SPI for prediction of droughts in Hungary. They concluded that the SPI can be used for describing different types of drought events. Hughes and Saunders (2002) performed drought-related studies for the European region. They compared the characteristics of SPI computed over different timescales of 3–24 months with the Palmer Drought Severity Index (PDSI) and found a close resemblance between the values corresponding to 12-month SPI and PDSI. Ntale and Gan (2003) analyzed the properties of three drought indexes for East Africa and found that SPI was more appropriate for drought monitoring. Seiler et al. (2002) used SPI as a tool to monitor wetter than normal conditions and assessed the potential of SPI for monitoring flood risks in Argentina. The SPI was successfully used to explain the development of conditions leading to some of the major flood events in the region. Seven meteorological indexes were compared by Morid et al. (2006)—SPI being one amongst them—and it was concluded that SPI can be effectively used to analyze both dry and wet events. Linear and nonlinear trends of droughts and wetness were observed in Europe using SPI by Bordi et al. (2009). The advantages and disadvantages of using SPI were discussed by Hayes et al. (1999). Guttman (1999) performed a detailed study of SPI and found that the value of SPI strongly depends on the probability-distribution model used to define the time series of precipitation.

SPI has been popular in India as well. Ghosh and Mujumdar (2007) examined the severity of future droughts for a meteorological subdivision of India using SPI as a drought indicator and the results from different GCM simulations. Bhuiyan et al. (2006) monitored drought dynamics in the Aravalli region of India using several indexes and showed that SPI is efficient in quantifying the precipitation deficit. Because of its relative simplicity and effectiveness in assessing both abnormally dry and wet conditions, SPI is used in the present study as an indicator to analyze severe drought and wet conditions across India.

Methodology

In this study, SPI calculation initially requires prior selection of GCM, data processing of GCM precipitation-rate output, and identifying the probability distribution of the observed rainfall data. Data processing includes GCM data interpolation to a common grid, grid-cell averaging, and bias correction. The steps involved in data processing are explained in the subsequent sections.

Data Processing

Interpolation to a Common Grid
Different GCMs use grid cells of different sizes and nonuniform placement. For the grid-based analysis and bias-correction method

described here, it is required that the GCM outputs and monthly gridded rainfall data at 1° grid spacing from the India Meteorological Department (IMD) (Rajeevan et al. 2006) be interpolated on a common grid. As the IMD rainfall data have been compared with all 17 GCMs, a grid size of $1° × 1°$ has been adopted as the reference. A total of 357 grid points were identified across India for which interpolated rainfall data, strongly constrained by observations from a dense network, were available. The GCM outputs were interpolated at these grid points using weighted mean of the four nearest values, with weights being assigned based on inverse-square distance relationship (Johnson and Sharma 2009).

Grid-Cell Averaging

The interpolated GCM variables show large fluctuations in space that confound relationships with predictor variables (Wilby and Wigley 2000). To remove spurious spatial fluctuations, spatial smoothing was adopted wherein each grid-cell value is replaced by the average of the nine nearest grid-cell values.

Bias Correction

The goal of the present study was to look at longer term extreme events as strong seasonality in rainfall does not readily permit smaller timescale drought computations. Even though the drought indicator used for the present study, SPI, actually assumes independence in monthly estimates, the data are aggregated for larger time windows. While constructing the cumulative values over larger time windows, it was felt necessary to account for variability in monthly and annual values. Heuristically, because monthly data were used, the first two moments at the monthly scale are obvious choices. As drought timescales examined here were greater than 1 year, annual mean and variance were also obvious choices.

Further, the precipitation patterns over India are very much affected by regional and global teleconnections; however, the influence of these teleconnections is not properly represented in the GCM outputs. By bias correcting for annual lag-1 correlation, an improvement in interannual variability was demonstrated previously (Johnson and Sharma 2012). This correction is important for ensuring that drought and flood periods are modeled properly. Therefore, bias correction for lag-1 correlation was also implemented here. When compared with other existing techniques for bias correction, NBC is the only available method that allows correcting for bias in mean, standard deviation, and lag-1 correlation of GCM precipitation data at monthly and annual scales in a simple manner. Thus, for the present analysis, NBC was adopted as a bias-correction technique. However, monthly bias-correction (MBC) approach is one of the simplest techniques available for bias correction. Therefore, both MBC and NBC were used to correct for the biases in the grid cell–averaged GCM outputs. The MBC corrects the mean and standard deviation of the GCM-generated rainfall data to match the observations. The underlying assumption of MBC is that the distribution of GCM-generated rainfall data is sufficiently similar to that of the observations such that the GCM rainfall only needs to be shifted and scaled to match the observations.

In the first step for MBC, raw monthly time series of a GCM-generated rainfall data ($y_{i,k}$) are standardized to create $y'_{i,k}$ for each month (i) in year (k) by removing the model monthly mean ($\mu_{\text{mod},i}$) and standard deviation ($\sigma_{\text{mod},i}$) for that month (i) as shown in Eq. (1):

$$y'_{i,k} = \frac{y_{i,k} - \mu_{\text{mod},i}}{\sigma_{\text{mod},i}} \quad (1)$$

The second step is to impose the observed mean ($\mu_{\text{obs},i}$) and standard deviation ($\sigma_{\text{obs},i}$) to create a transformed time series $z_{i,k}$ at the monthly level as follows:

$$z_{i,k} = y'_{i,k}\sigma_{\text{obs},i} + \mu_{\text{obs},i} \quad (2)$$

For the validation period, monthly mean and standard deviation values of the calibration period are used, with Eqs. (1) and (2) remaining the same.

The NBC is based on the assumption of linear autoregressive model. The choice of statistics that deserve correction depends on the application at hand. To meet the goals of this study for characterizing extreme events at timescales greater than 1 year, NBC was designed to correct for biases in means, variances, and lag-1 autocorrelations, along with replicating the interannual variability that is often ignored by other bias-correction techniques. The steps involved for NBC are presented in the following section.

Similar to MBC, in the first step, raw monthly time series of a GCM-generated rainfall data ($y_{i,k}$) are standardized to create $y'_{i,k}$ for each month (i) in year (k) by removing the model monthly mean ($\mu_{\text{mod},i}$) and standard deviation ($\sigma_{\text{mod},i}$) for that month (i) as shown in Eq. (3):

$$y'_{i,k} = \frac{y_{i,k} - \mu_{\text{mod},i}}{\sigma_{\text{mod},i}} \quad (3)$$

In the second step, monthly lag-1 autocorrelations ($\rho_{\text{mod},i}$), which are present in the model results from the standardized time series, are removed and instead the observed monthly lag-1 autocorrelations ($\rho_{\text{obs},i}$) are applied to create $y''_{i,k}$ as shown in Eq. (4). Monthly lag-1 autocorrelations are defined as the correlation of the time series of the values from month (i) with the time series of month ($i-1$):

$$y''_{i,k} = \rho_{\text{obs},i}y''_{i-1,k} + \sqrt{1 - \rho_{\text{obs},i}{}^2}\left(\frac{y'_{i,k} - \rho_{\text{mod},i}y'_{i-1,k}}{\sqrt{1 - \rho_{\text{mod},i}{}^2}}\right) \quad (4)$$

The third step is to rescale with the observed means ($\mu_{\text{obs},i}$) and standard deviations ($\sigma_{\text{obs},i}$) to create the nested time series $y'''_{i,k}$ at the monthly level:

$$y'''_{i,k} = y''_{i,k} \times \sigma_{\text{obs},i} + \mu_{\text{obs},i} \quad (5)$$

In the fourth step, the monthly transformed values ($y'''_{i,k}$) are aggregated to the annual scale (z_k). The monthly process is repeated for the annual time step, with the difference that there is no further need of calculating the model parameters separately for each month. The annual time series (z_k) is modified by standardizing the mean (μ_{mod}) and standard deviation (σ_{mod}) of the annual time series, such that for year k

$$z'_k = \frac{z_k - \mu_{\text{mod}}}{\sigma_{\text{mod}}} \quad (6)$$

The fifth step is to remove any modeled annual lag-1 autocorrelations (ρ_{mod}) and apply the observed annual lag-1 autocorrelations (ρ_{obs}). Yearly lag-1 autocorrelations are calculated as the correlation between the variable in one year and the next as follows:

$$z''_k = \rho_{\text{obs}}z''_{k-1} + \sqrt{1 - \rho_{\text{obs}}{}^2}\left(\frac{z'_k - \rho_{\text{mod}}z'_{k-1}}{\sqrt{1 - \rho_{\text{mod}}{}^2}}\right) \quad (7)$$

The last step is to create the final annual time series by rescaling with the observed annual means (μ_{obs}) and standard deviations (σ_{obs}) as follows:

Table 1. Event Classification Based on SPI Value

SPI value	Category	Probability (%)
≥2.00	Extremely wet	2.3
1.50–1.99	Severely wet	4.4
1.00–1.49	Moderately wet	9.2
0–0.99	Mildly wet	34.1
0 to −0.99	Mild drought	34.1
−1.00 to −1.49	Moderate drought	9.2
−1.50 to −1.99	Severe drought	4.4
−2 or less	Extreme drought	2.3

$$z_j''' = z_j'' \times \sigma_{\text{obs}} + \mu_{\text{obs}} \tag{8}$$

The time series obtained in Eq. (8) are used to correct the monthly GCM time series ($y_{i,k}$). The final bias-corrected value is calculated using the following expression:

$$Y_{i,k} = \frac{z_k'''}{z_k} \times y_{i,k}''' \tag{9}$$

For validation and the future study period, the observed and modeled statistics for the calibration period are used to adjust the model results, assuming that the biases in the model for the calibration period remain the same in the validation and future period as well (Johnson and Sharma 2012). To be specific, the validation and future bias-correction steps with Eqs. (3), (4), (6), and (7) use the monthly and annual statistics from the GCM data for the climate-calibration period. Thus, monthly and annual means, standard deviations, and lag-1 correlations serve as the calibration parameters. Parameters in Eq. (9) are kept unchanged for the validation and future period.
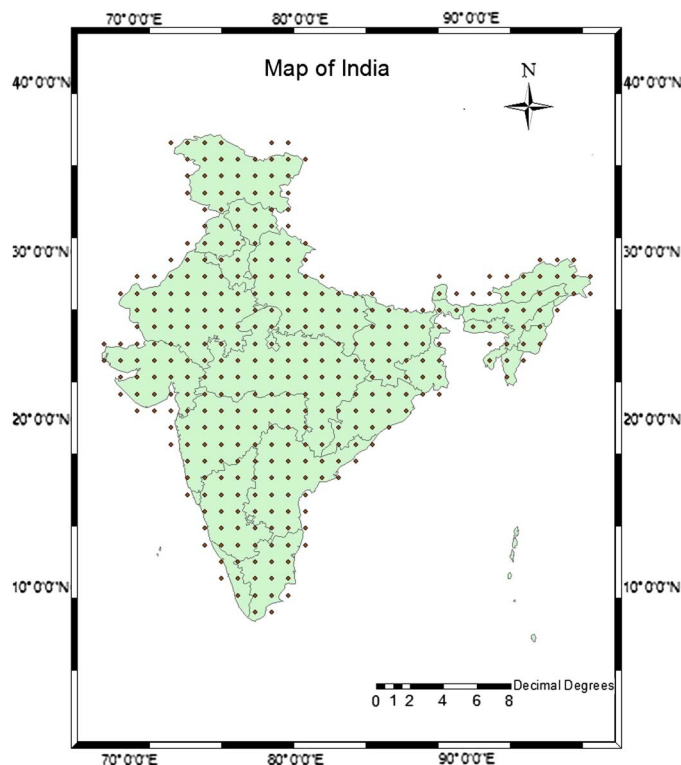
Distribution Fitted for Observed Rainfall Data

The SPI value is dependent on the probability distribution and the model used to characterize the time series of precipitation. In the present study, the assumption that a gamma distribution would provide the best representation to the observed rainfall data was tested by computing the monthly Kolmogorov–Smirnov (K–S) statistics for each grid cell for three timescales (12, 24, and 60 months) using gamma, lognormal, and normal distributions (Hughes and Saunders 2002). The test statistic is formulated as follows:

$$D_n = \max |F_n(x) - F(x)| \tag{10}$$

Fig. 1. Grid map of the study area

where $F_n(x)$ = empirical cumulative probability, estimated as $F_n(x_i) = i/n$ for the ith smallest data value; $F(x)$ = theoretical cumulative probability distribution estimated at x. The D_n is compared with tabulated values appropriate to the sample size and the assumed distribution, under the null hypothesis that the data are drawn from the theoretical distribution. The null hypothesis is rejected at a given level of significance, if D_n exceeds the critical value (Hughes and Saunders 2002). The results of the K–S test for this study confirm that gamma distribution approximates the best empirical distribution of the observed rainfall at a majority of grid points over India.

Computation of SPI

After processing GCM rainfall data and identifying a suitable probability-distribution function for the observed rainfall data,

Table 2. List of GCMs Considered in the Study

Serial number	Model used	Organization
1	BCCR_BCM2_0	Bjerknes Centre for Climate Research, Bergen, Norway
2	CSIRO 3.5	Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia
3	GFDL 2.0	Geophysical Fluid Dynamic Laboratory, Princeton, NJ, USA
4	GFDL 2.1	Geophysical Fluid Dynamic Laboratory, Princeton, NJ, USA
5	GISS	Goddard Institute for Space Studies, New York, NY, USA
6	INGV- ECHAM 4	Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy
7	INMCM 3	Institute of Numerical Mathematics, Moscow, Russia
8	IPSL_CM4	Institut Pierre Simon Laplace, Paris, France
9	MIROC3	Centre for Climate System Research, Tokyo, Japan
10	MPI-ECHAM5	Max Planck Institute for Meteorology, Hamburg, Germany
11	MRI-CGCM2	Meteorological Research Institute, Tsukuba, Japan
12	NCAR_CCSM3	Community Earth System Models, NCAR, Boulder, Colorado, USA
13	NCAR-PCMI	Parallel Climate Models, NCAR, Boulder, Colorado, USA
14	UKMO-hadcm3	UK Met Office, Exeter, UK
15	UKMO-hadgem1	UK Met Office, Exeter, UK
16	CSIRO 3.0	Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia
17	CNRM_CM3	Canadian Climate Centre, Gatineau, Canada

the SPI values at each grid point for the time series of different timescales are calculated using the procedure described previously (McKee et al. 1993). The steps involved in SPI calculation are as follows:

- First, fit a gamma distribution to the time series of precipitation values for each timescale of interest (12, 24, and 48 months, etc.) without overlapping the data segments. Then compute the parameters of the gamma distribution.
- The second step is to compute the value of cumulative distribution function (CDF) $[G(x)]$ corresponding to each value of precipitation (x).
- Finally, the value of the standard normal deviate corresponding to the value of CDF $[G(x)]$ is calculated, which is the SPI value for the precipitation (x).

On the basis of the value, dry and wet events are identified. Table 1 presents the categories of events corresponding to SPI values (McKee et al. 1993).

The parameters of the gamma distribution are estimated based on the observed precipitation data for three different timescales, namely, 12, 24, and 60 months. Shorter durations were not considered as they will lead to too many zero precipitation values in the data set for analysis. Using these parameters, the bias-corrected future precipitation values obtained from different GCMs are used to calculate 12-month SPI (SPI-12), 24-month SPI (SPI-24), and 60-month SPI (SPI-60) values.

Application

Data Selection

Precipitation outputs from different GCMs are highly uncertain and so are the results of the impact studies performed using these outputs. It has been found that the mean of an ensemble prediction has, on an average, a smaller error than the mean error of any of the individual forecasts (Tang et al. 2008). However, in most regional hydrologic assessments, due to time and resource constraints, the number of GCMs used for impact studies is limited. In the present
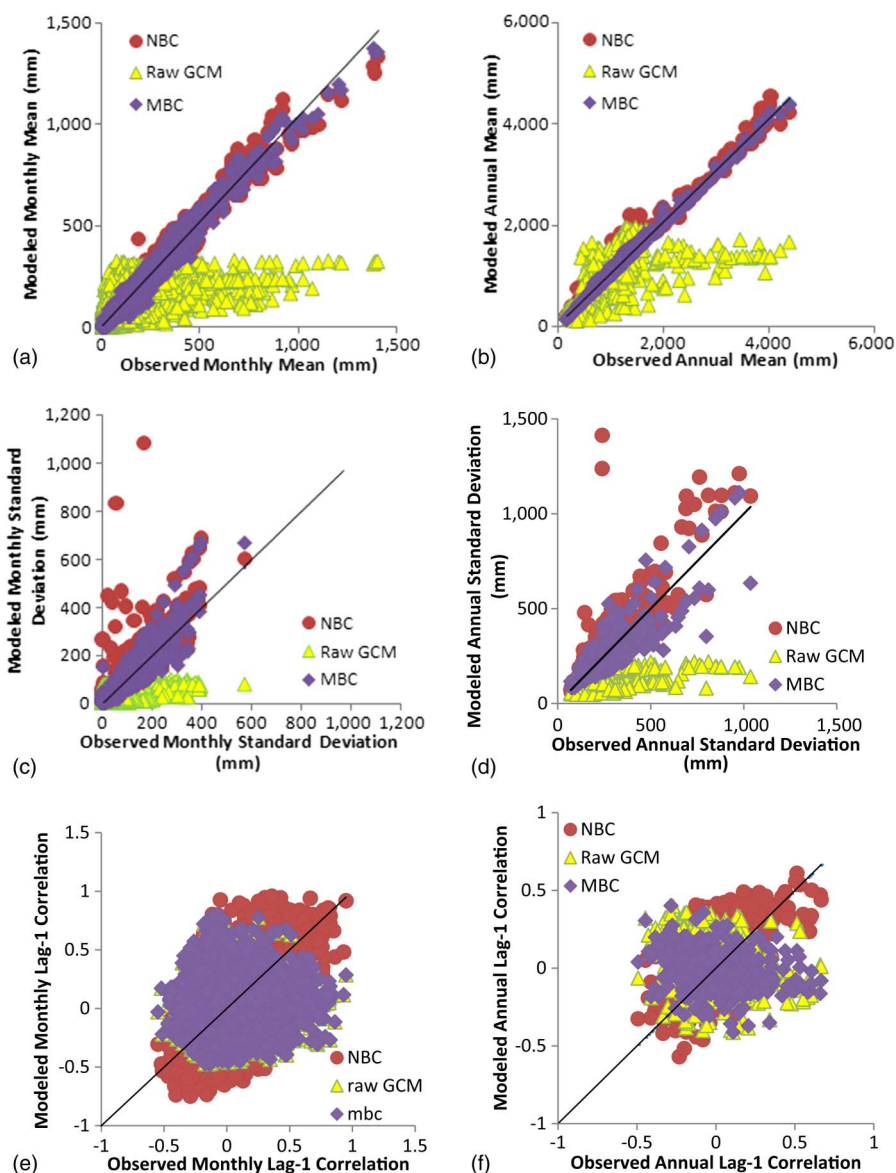
Fig. 2. Modeled rainfall data for BCCR_BCM2_0 versus the observed IMD rainfall data statistics of monthly and annual precipitation rate for raw GCM outputs and the MBC and NBC model outputs across India for the period between 1976 and 1999: (a) monthly mean; (b) annual mean; (c) monthly standard deviation; (d) annual standard deviation; (e) monthly lag-1 correlation; (f) annual lag-1 correlation

study, monthly precipitation outputs of 17 GCMs (Table 2), a subset of the 23 available GCMs [used as part of the Intergovernmental Panel on Climate Change assessment report (Meehl et al. 2007) for which time series of precipitation are readily available] was used. The sequence number associated with each model identifier in Table 2 is further used to refer to them in the results to follow.

The precipitation outputs for 20C3M (representing current climate) and Special Report on Emissions Scenarios (SRES) A2 (representing future climate) were obtained from the World Climate Research Programme's Coupled Model Intercomparison Project Phase 3 multimodel data set.

Daily gridded rainfall data at 1° grid spacing from the IMD (Rajeevan et al. 2006) were used as surrogate for observed data. This data set is based on 1,803 stations over India with a minimum of 90% data availability during the period 1951–2003. Using the data from these stations, the grid-point analysis of rainfall was carried out by the IMD using Shepard's directional interpolation method over the Indian subcontinent and finally daily rainfall values were obtained for different grid points. Overall, 357 grid points were identified for which daily rainfall data were available. The daily data set was aggregated to monthly data and used for the present study.

Study Region

The methodology has been applied across India (latitude from 6.5° to 37.5° N and longitude from 66.5° to 101.5° E) covering 357 grid points at a resolution of 1° × 1°. Fig. 1 shows the study region along with the grid points for which interpolated rainfall data are available. The shape, size, location, latitudinal extent, and contrasting relief features of India bring great climatic diversity across different regions of India. Aberrations in the total volume and pattern of rains from the Southwest monsoon are observed in different parts of the country. The ENSO also plays an important role in the rainfall patterns across India.

Performance Assessment

Different GCM outputs are available for the 20C3M scenario (current climate) over a wide period. The period between 1951 and 1999 was identified as a common period for which precipitation output were available from all the 17 GCMs and IMD-gridded rainfall data. Calibration and validation were performed for both MBC and NBC GCM data. The data set for the current period was divided into two parts, namely the calibration period (1951–1975) and the validation period (1976–1999). The GCMs were ranked
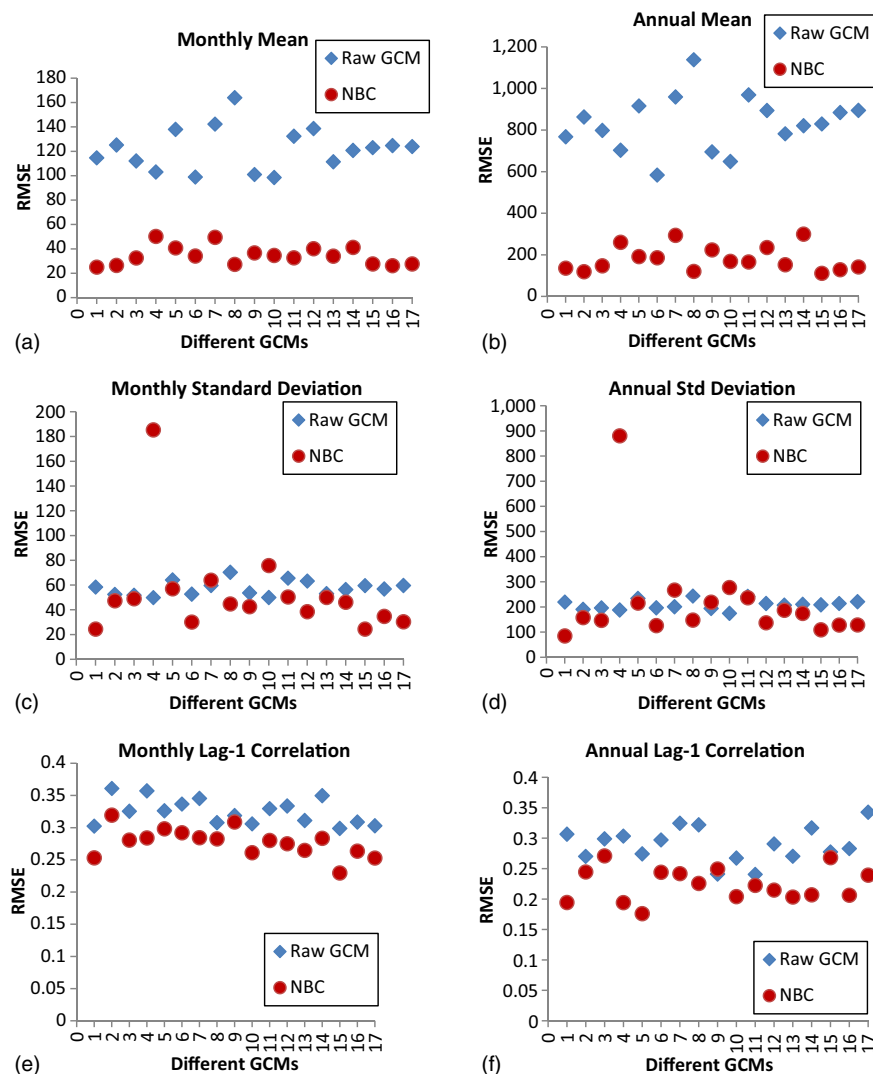
Fig. 3. RMSE for NBC and raw GCM data compared with the observed IMD rainfall data for the validation period (1976–1999) from 17 GCMs as listed in Table 2: (a) RMSE for monthly mean; (b) RMSE for annual mean; (c) RMSE for monthly standard deviation; (d) RMSE for annual standard deviation; (e) RMSE for monthly lag-1 correlation; (f) RMSE for annual lag-1 correlation

Table 3. Ranking of GCMs Based on RMSE

|	Monthly Statistics			Annual Statistics				
GCM	Mean	Standard deviation	Lag-1 correlation	Mean	Standard deviation	Lag-1 correlation	Average ranking	Overall ranking
BCCR_BCM2_0	1	2	3	6	1	3	2.67	1
CSIRO3.5	3	10	17	2	9	14	9.16	8
GFDL 2.0	7	11	9	3	7	17	9	7
GFDL2.1	17	17	12	14	17	2	13.16	14
GISS	15	14	15	12	12	1	11.5	12
INGV ECHAM4	10	3	14	11	3	13	9	7
INMCM 3.0	16	15	13	16	15	12	14.5	15
IPSL_CM4	4	8	10	4	8	10	7.3	5
MIROC3	12	7	16	13	13	15	12.67	13
MPI_ECHAM5	11	16	4	10	16	5	10.33	10
MRI_CGCM2	8	13	8	9	14	9	10.16	9
NCAR-PCMI	13	6	7	15	6	8	9.16	8
NCAR_CCSM3	5	1	1	1	2	16	4.33	2
UKMO-hadcm3	9	12	6	8	11	4	8.33	6
UKMO-hadgem1	14	9	11	17	10	7	11.33	11
CSIRO 3.0	2	5	5	5	4	6	4.5	3
CNRM-CM3	6	4	2	7	5	11	5.83	4

based on their performance during the validation period, which was assessed using the following six statistics, i.e., monthly and annual values of mean, standard deviation, and lag-1 correlation. Five best models were identified based on improvement in root-mean-square error (RMSE) and further used for SPI calculation. The K–S test was used to identify the suitable probability-distribution function for the observed IMD rainfall data, in this case gamma distribution. For the present study, 24 years of data were available for the

validation period. It has been previously suggested that a minimum of 30–40 years of time series data should be available for reliable prediction of SPI (McKee et al. 1993). Hence, the parameters required for estimation of SPI, that is, the parameters of gamma distribution, are estimated based on 49 years of observed precipitation data (1951–1999).

Results and Discussion

Implications of NBC

A comparative study was performed for the validation period (1976–1999) between NBC and MBC. Fig. 2 presents these results for the BCCR_BCM2_0 GCM rainfall. Each individual point in the graphs represents the respective monthly and annual statistic for
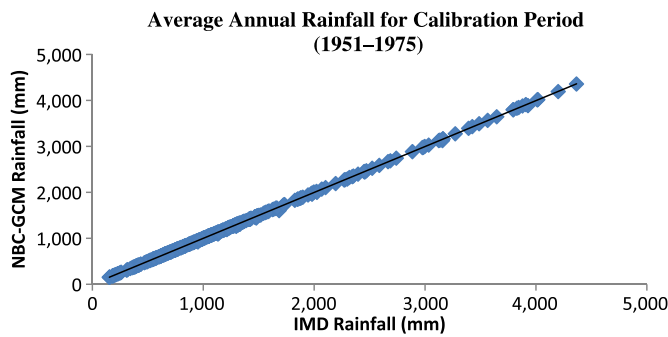
Fig. 4. Mean of average annual rainfall for NBC five GCMs rainfall data versus average annual rainfall from the observed IMD rainfall data for the calibration period (1951–1975)
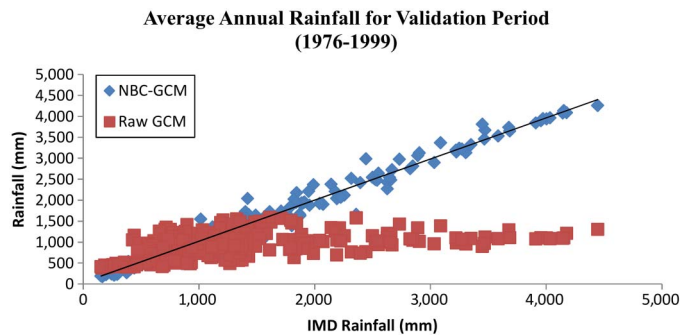
Fig. 5. Mean of average annual rainfall for NBC and raw five GCMs rainfall data versus average annual rainfall from the observed IMD rainfall data for the validation period (1976–1999)
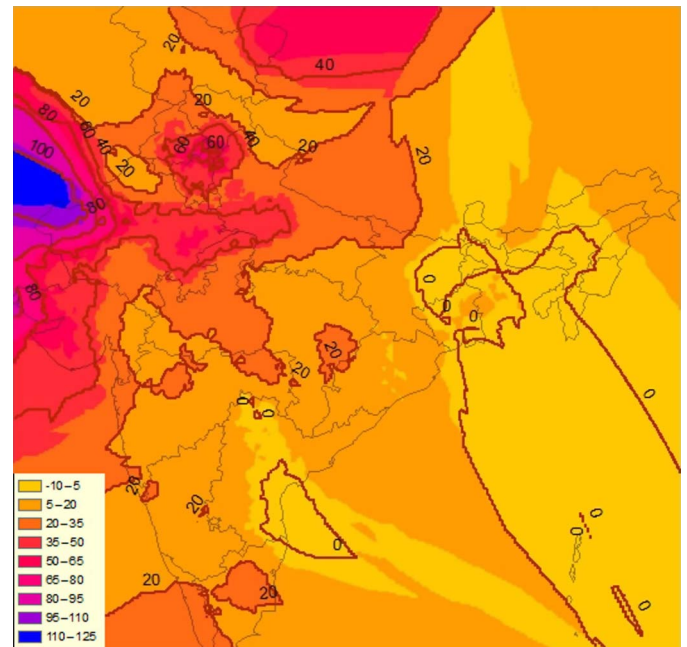
Fig. 6. Percentage change in average annual rainfall across India for mid-2075

each grid cell (i.e., each location in which the bias correction has been applied). Both NBC and MBC perform well while correcting for the annual and monthly mean and standard deviation. However, in the case of monthly and annual lag-1 correlation, NBC shows more improvement. The RMSE values obtained in the case of NBC for monthly and annual lag-1 correlation are 0.254 and 0.19, which are smaller compared with the RMSE values of 0.302 and 0.305 obtained in the case of MBC.

Fig. 3 presents the RMSE values for monthly and annual mean, standard deviation, and lag-1 correlation as observed for the validation period (1976–1999) for the raw and NBC rainfall from all the 17 GCMs. The RMSEs show significant improvement in the monthly and annual means for the NBC data from the 17 GCMs when compared with raw GCM rainfall data. Improvement in RMSE is demonstrated in case of other statistics as well. However, monthly lag-1 correlation does not show as good an improvement as the other statistics. Because of the ability of NBC to correct for bias in the data at multiple timescales and also improve lag-1 autocorrelation, it was adopted for the present study. The GCMs were

ranked based on the RMSE for all the six statistics shown in Table 3. Average ranking was obtained by averaging the six ranks. Finally, based on the value of average ranking, the overall ranking was obtained. The model with the least average ranking was ranked 1 and similarly other models were also ranked. It is not possible to determine a priori which statistic is most relevant for analyzing extreme events. For robustness, rankings over multiple statistics were adopted here. In spite of bias correction, the future projections from different GCMs revealed substantial inter-GCM variability. The initial pool of 17 GCMs was further screened to obtain more reliable predictions of SPI. The top five performing models were selected and used for further analysis of severe droughts and wet events in future to demonstrate the methodology. This number would vary based on the resources available to conduct such an analysis.

The performances of the best five selected GCMs in predicting average annual rainfall for the calibration and validation periods were analyzed. The mean of the average annual rainfall predicted from all the five GCMs using NBC and raw GCM data was compared with the average annual rainfall from the observed
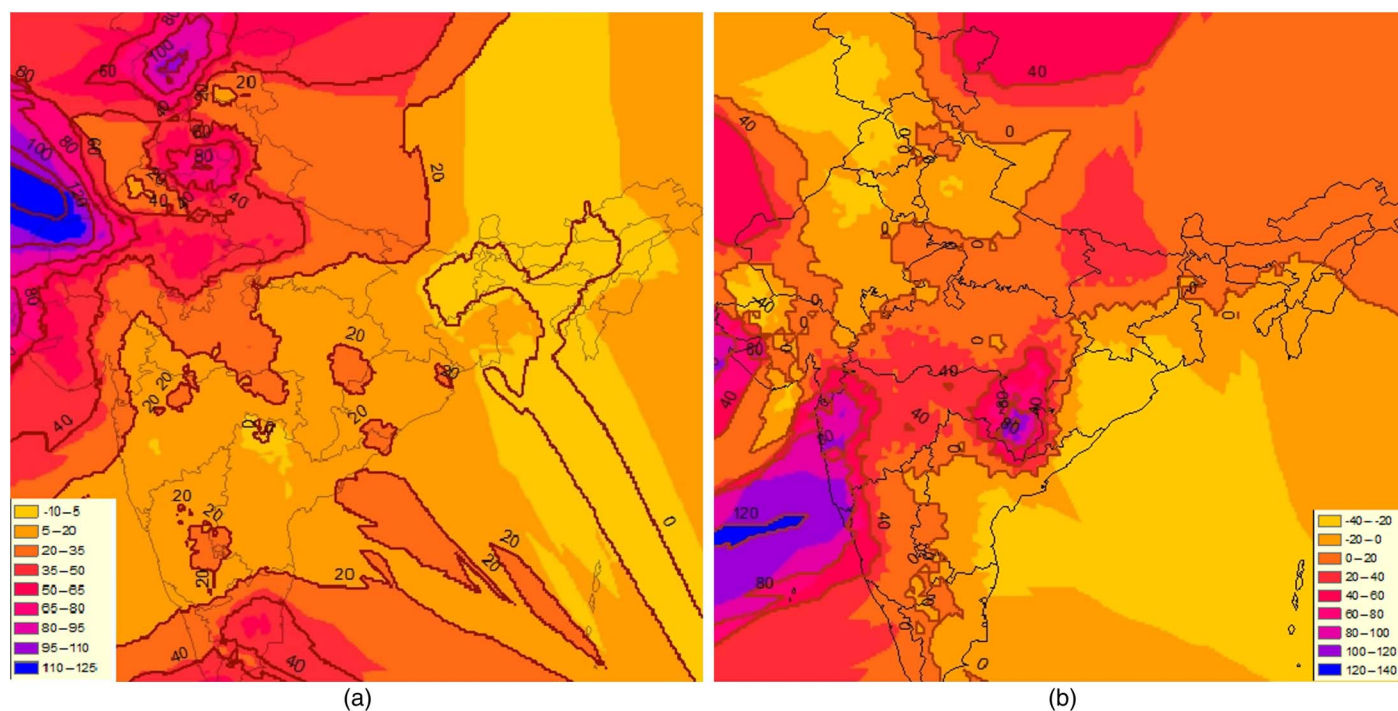

Fig. 7. Percentage change in average rainfall across India for mid-2075: (a) monsoon period; (b) nonmonsoon period

Table 4. Percentage of Grid Points That Agree for a Particular Distribution

Month	Timescale: 12 months			Timescale: 24 months			Timescale: 60 months		
	Normal	Lognormal	Gamma	Normal	Lognormal	Gamma	Normal	Lognormal	Gamma
January	95	95	96	95	96	96	92	92	92
February	95	94	95	94	95	94	93	93	93
March	95	95	95	92	93	93	93	93	93
April	95	95	95	94	94	94	93	93	93
May	94	95	96	94	95	94	92	92	91
June	93	94	94	92	93	93	92	92	92
July	93	94	94	96	96	96	89	89	89
August	95	95	96	95	96	96	92	93	93
September	96	96	96	98	99	98	91	92	92
October	97	97	97	95	95	95	91	92	92
November	95	95	95	96	97	97	92	93	93
December	95	95	96	96	97	97	92	93	93

IMD rainfall data. Figs. 4 and 5 show the plots for the calibration (1951–1975) and validation (1976–1999) periods, respectively. In case of NBC GCM data, R^2 of 0.99 was obtained for the calibration period and 0.97 for the validation period. It can be observed from Fig. 5 that the raw GCM rainfall data exhibit large bias for the validation period 1975–1999, as the uncorrected GCM-generated values deviate to a great extent from the best fit line, and an R^2 value of only 0.2321 could be obtained, thereby highlighting the need for bias correction in GCM rainfall outputs.

Most GCMs show large uncertainties in the projected regional climate changes for near-future results up to 2030, mostly due to differences between the results of the climate models rather than the different emission scenarios (Hennessy et al. 2008). In the present study, future period from 2051 to 2099 was chosen because GCM outputs for this period are likely more strongly affected by emission scenarios rather than model uncertainties. The average annual rainfall was estimated using the best five performing models for future period (2051–2099) and was reported here, for convenience, as the average annual rainfall for 2075. The percentage change in average annual rainfall at all grid points when compared with average annual rainfall from the observed IMD data is shown in Fig. 6.

**Table 5.** RMSE Values of Raw GCM-, MBC-, and NBC GCM–Prediction Error for SPI Values

Statistic	Raw GCM	MBC	NBC
5th Percentile: 12-month SPI	2.11	0.40	0.36
5th Percentile: 24-month SPI	2.95	0.52	0.41
5th Percentile: 60-month SPI	4.88	0.66	0.55
95th Percentile: 12-month SPI	2.49	0.29	0.20
95th Percentile: 24-month SPI	3.46	0.42	0.25
95th Percentile: 60-month SPI	5.10	0.61	0.45

An increase in annual rainfall is observed across many parts of India. A greater increase is observed for many regions of northwest and central northwest parts of India. The northeastern part of India does not show appreciable change in rainfall. Further, to investigate the changes in monsoon patterns, percentage changes in average rainfall for the monsoon (June–October) and nonmonsoon periods (November–May) for the year 2075 were calculated and are presented in Fig. 7.

The percentage change in rainfall for monsoon period is quite similar to that obtained for average annual rainfall. However, for the
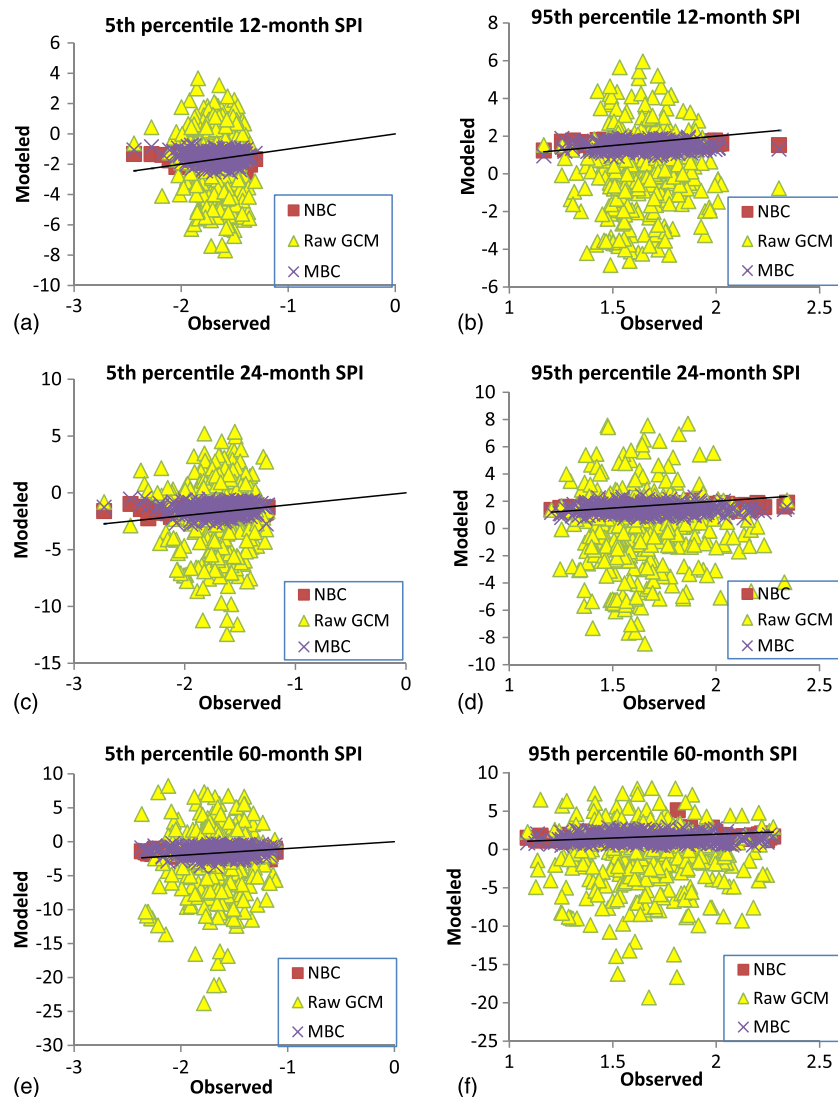
Fig. 8. SPI value for NBC, MBC, and raw GCM data for BCCR_BCM2_0 compared with the observed IMD rainfall data for the current period (1951–1999): (a) 5th percentile, 12-month SPI; (b) 95th percentile, 12-month SPI; (c) 5th percentile, 24-month SPI; (d) 95th percentile, 24-month SPI; (e) 5th percentile, 60-month SPI; (f) 95th percentile, 60-month SPI
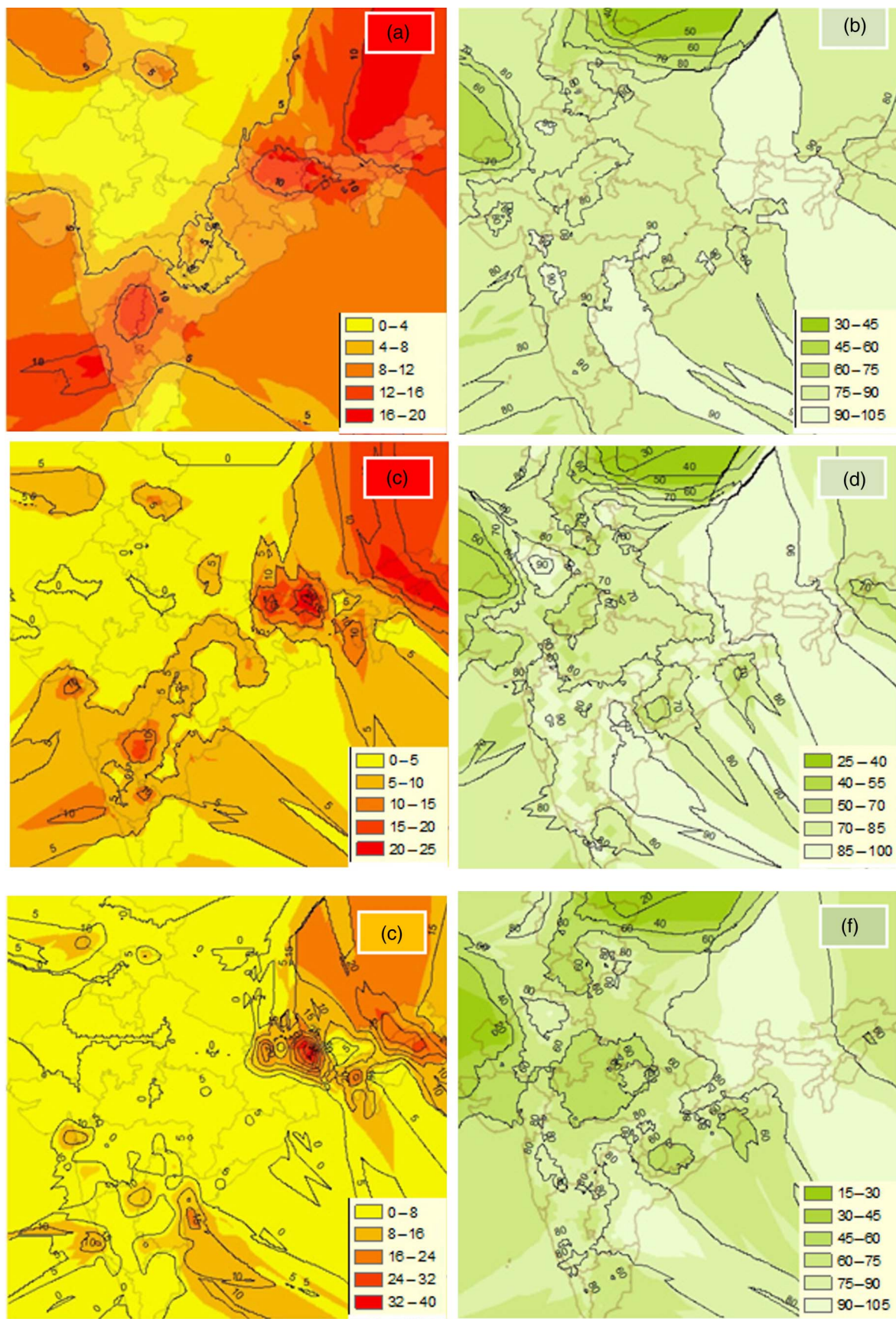
Fig. 9. Frequency plots for severe drought and wet events for NBC data around mid-2075: (a) frequency of severe drought event, SPI-12; (b) frequency of severe wet event, SPI-12; (c) frequency of severe drought event, SPI-24; (d) frequency of severe wet event, SPI-12; (e) frequency of severe drought event, SPI-60; (f) frequency of severe wet event, SPI-60
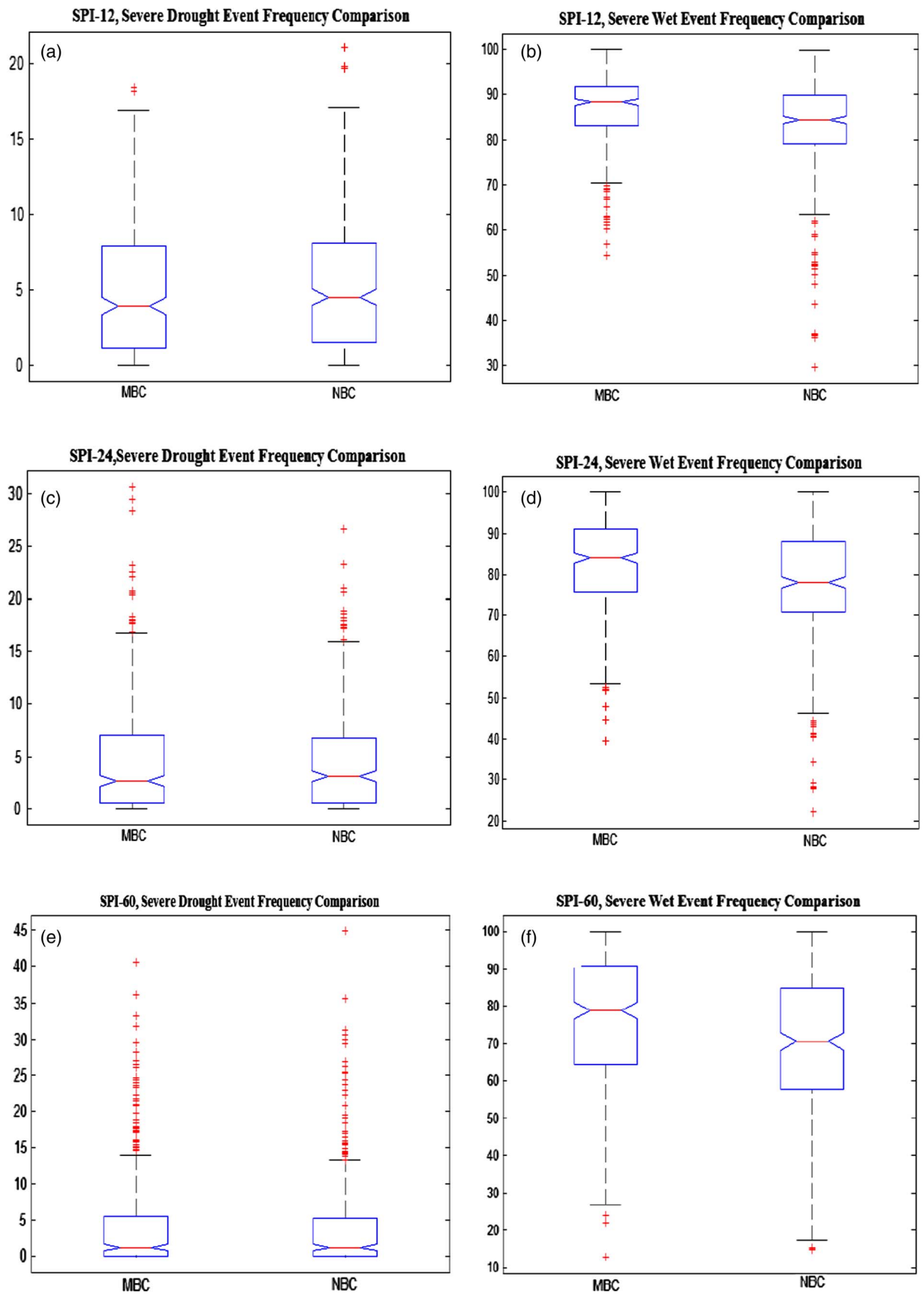
Fig. 10. Comparison of frequency values for severe drought and wet events for NBC and MBC data around mid-2075: (a) SPI-12, severe drought event; (b) SPI-12, severe wet event; (c) SPI-24, severe drought event; (d) SPI-24, severe wet event; (e) SPI-60, severe drought event; (f) SPI-60, severe wet event

nonmonsoon period the extent of increase in rainfall amount is high for some of the coastal regions of India and certain regions of west central part of India.

Probability Distribution for the Observed IMD Rainfall Data

The K–S test was performed to identify the suitable probability-distribution function for the observed IMD rainfall data for time-scales of 12, 24, and 60 months at all the 357 grid points. Rejection of the null hypothesis at 5% significance level was considered. Lognormal and gamma distribution fitted the time series data of different timescales for most grid points. Gamma distribution slightly outperformed for a timescale of 12 months, whereas for data of 24- and 60-month timescales both distributions fit almost equally well. Because gamma distribution was originally suggested for calculation of SPI previously (McKee et al. 1993), the same was adopted for the present study. Table 4 shows the percentage fit out of 357 grid points for all the three distributions for the three timescales.

Modeling of Current and Future Severe Droughts and Wet Events

Drought frequency was modeled for the current period (1951–1999) across India for raw and bias-corrected GCM data. Fig. 8 shows the 5th and 95th percentile plots for SPI values for time-scales of 12, 24, and 60 months for all the 357 grid points. The 5th and 95th percentile SPI values have been identified as severe droughts and wet events, respectively. The NBC and MBC provide better estimates of severe droughts and wet events when compared with the SPI values from raw GCM data series. The SPI values for raw GCM data deviate to a great extent from the best fit line in Fig. 8.

The RMSE values for raw, MBC, and NBC GCM data series are shown in Table 5. The RMSE values have significantly reduced for both the bias-corrected series. Both NBC data and MBC data show significant improvement in RMSE values compared with the raw GCM data series. However, RMSE values for NBC are better than those for MBC. The NBC data were further used for assessing the severe droughts and wet events in future. The top performing models were checked separately for their performance for a future period for predicting the 5th and 95th percentile SPI values. The NCAR_CCSM3 was not able to predict SPI values very successfully when compared with the other four models. Thus, even after bias correction and selection of a best subset of models, further processing was necessary to screen poorly performing models. It is suspected that the inability of NCAR_CCSM3 to correctly model annual lag-1 autocorrelation might be one of the reasons for this poorer performance. The remaining four models were further used for analysis.

The 5th and 95th percentile SPI values obtained from the time series of observed IMD rainfall for the three timescales were defined as the threshold values at each grid cell for severe drought and wet events (Burke and Brown 2008). The median frequency of severe drought and wet events was estimated based on future projections of the four GCMs. Fig. 9 shows the median frequency of severe droughts and wet events centered around 2075 for 12, 24, and 60 months for the NBC data.

The frequency plots for drought events show increase in 1-year drought frequency for many parts of the country, while a great increase in drought events for some of the regions of central northeast, west central, and peninsular India is observed. Interestingly, the frequency of wet events increases across the northwest region of India. This region is considered to be drought prone for most of the year as it receives a very small amount of monsoon rainfall. An increase in the frequency of wet events was also observed for coastal regions of India as well as for most parts of northern India. Many parts of northern India show frequency of wet events at approximately 60–70%. A comparison of distribution of frequency values is shown in Fig. 10 for frequency values predicted using MBC and NBC for the year 2075.

The frequency of severe drought events predicted by both MBC and NBC is similar, with NBC value being slightly on the lower side. The frequency for severe drought events is approximately 6–7% meaning that in a particular year 6–7% of the whole country is likely to be suffering from drought. Frequency of severe wet events predicted by NBC is greater than that predicted by MBC. The frequency of severe wet events in one particular year is 12%. The frequency for severe wet events increased for 2- and 5-year time windows to 20 and 25%, respectively.

Conclusions

The NBC approach was used to correct for the biases in GCM-derived rainfall for SRES A2. The NBC performed better than the simple monthly means correction, which has often been used in climate change–impact assessments. To a great extent, NBC compensates for the known weaknesses in GCM outputs and corrects for mean, variance, and lag-1 correlation. Overall, a 10–30% increase in annual rainfall was projected for 2075. For the nonmonsoon period, some parts of coastal regions of India and west central part of India showed greater increase in rainfall. Severe droughts and wet events were predicted across India using SPI. The SPI was calculated using the bias-corrected rainfall data for three different timescales. The 5th and 95th percentile values were extracted to denote the severe drought and wet events across India. An increase in frequency of severe droughts and wet events was observed across most parts of India for the future period. The severe drought-event frequency will likely be high for central northeast, west central, and peninsular regions of India. The northwest region of India showed an increase in the frequency of severe wet events for future, which is rather contrary to the current trend observed for this region. The frequencies of severe drought events predicted by both MBC and NBC were similar, with the NBC value being slightly on the lower side. Frequency of severe wet events predicted by NBC was greater than that predicted by MBC. The frequency for severe wet events increased for 2- and 5-year durations, whereas the frequency for severe drought events remained almost same for these time windows.

Acknowledgments

References

Anandhi, A., Srinivas, V. V., Nanjundiah, R. S., and Nagesh Kumar, D. (2008). "Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine." *Int. J. Climatol.*, 28(3), 401–420.

Bhuiyan, C., Singh, R. P., and Kogan, F. N. (2006). "Monitoring drought dynamics in the Aravalli region (India) using different indices based on

ground and remote sensing data." *Int. J. Appl. Earth Obs. Geoinf.*, 8(4), 289–302.

Bordi, I., Fraedrich, K., and Sutera, A. (2009). "Observed drought and wetness trends in Europe: An update." *Hydrol. Earth Syst. Sci.*, 13(8), 1519–1530.

Burke, E. J., and Brown, S. J. (2008). "Evaluating uncertainties in the projection of future drought." *J. Hydrometeorol.*, 9(2), 292–299.

Bussay, A., Szinell, C., and Szentimery, T. (1999). *Investigation and measurements of droughts in Hungary*, Hungarian Meteorological Service, Budapest, Hungary.

Dhanya, C. T., and Nagesh Kumar, D. (2009). "Data mining for evolution of association rules for droughts and floods in India using climate inputs." *J. Geophys. Res.*, 114(D2), D02102.

Ghosh, S., and Mujumdar, P. P. (2007). "Nonparametric methods for modeling GCM and scenario uncertainty in drought assessment." *Water Resour. Res.*, 43(7), W07405.

Guttman, N. B. (1999). "Accepting the Standardized Precipitation Index: A calculation algorithm." *J. Am. Water Resour. Assoc.*, 35(2), 311–322.

Hashino, T., Bradley, A. A., and Schwartz, S. S. (2007). "Evaluation of bias-correction methods for ensemble streamflow forecasts." *Hydrol. Earth Syst. Sci.*, 11(2), 939–950.

Hayes, M. J., Svoboda, M. D., Wilhite, D. A., and Vanyarkho, O. V. (1999). "Monitoring the 1996 drought using the Standardized Precipitation Index." *Bull. Am. Meteorol. Soc.*, 80(3), 429–438.

Hennessy, K., et al. (2008). "An assessment of the impact of climate change on the nature and frequency of exceptional climatic events." *CSIRO and the Bureau of Meteorology Rep. to the Australian Government*, CSIRO and Bureau of Meteorology, Canberra, Australia, ⟨http://www.daff.gov.au/__data/assets/pdf_file/0007/721285/csiro-bom-report-future-droughts.pdf⟩ (Jul. 23, 2008).

Houghton, J. T., Jenkins, G. J., and Ephraums, J. J., eds. (1990). *Climate change: The IPCC scientific assessment*, Cambridge University Press, Cambridge, UK.

Hughes, B. L., and Saunders, M. A. (2002). "A drought climatology for Europe." *Int. J. Climatol.*, 22(13), 1571–1592.

Johnson, F. M., and Sharma, A. (2009). "Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments." *J. Clim.*, 22(16), 4373–4382.

Johnson, F. M., and Sharma, A. (2011). "Accounting for interannual variability: A comparison of options for water resources climate change impact assessments." *Water Resour. Res.*, 47(4), W04508.

Johnson, F. M., and Sharma, A. (2012). "A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations." *Water Resour. Res.*, 48(1), W01504.

Maity, R., and Nagesh Kumar, D. (2008). "Basin-scale streamflow forecasting using the information of large-scale atmospheric circulation phenomena." *Hydrol. Processes*, 22(5), 643–650.

Maity, R., Nagesh Kumar, D., and Nanjundiah, R. S. (2007). "Review of hydroclimatic teleconnection between hydrologic variables and large scale atmospheric circulation patterns with Indian perspective." *ISH J. Hydraul. Eng.*, 13(1), 77–92.

McKee, T. B., Doeskin, N. J., and Kieist, J. (1993). "The relationship of drought frequency and duration to time scales." *Proc., 8th Conf. on Applied Climatology*, American Meteorological Society, Boston, 179–184.

McKee, T. B., Doeskin, N. J., and Kieist, J. (1995). "Drought monitoring with multiple time scales." *Proc., 9th Conf. on Applied Climatology*, American Meteorological Society, Boston, 233–236.

Meehl, G. A., et al. (2007). "Global climate projections." *Climate change 2007: The physical science basis*, S. Solomon, et al., eds., Cambridge University Press, Cambridge, UK, 747–846.

Mehrotra, R., and Sharma, A. (2010). "Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment." *Water Resour. Res.*, 46(7), W07526.

Mehrotra, R., Sharma, A., Nagesh Kumar, D., and Reshmidevi, T. V. (2013). "Assessing future rainfall projections using multiple GCMs and a multi-site stochastic downscaling model." *J. Hydrol.*, 488, 84–100.

Mishra, A. K., and Singh, V. P. (2010). "A review of drought concepts." *J. Hydrol.*, 391(1–2), 202–216.

Morid, S., Smakhtin, V. E., and Moghaddasic, M. (2006). "Comparison of seven meteorological indices for drought monitoring in Iran." *Int. J. Climatol.*, 26(7), 971–985.

Ntale, H. K., and Gan, T. Y. (2003). "Drought indices and their application to East Africa." *Int. J. Climatol.*, 23(11), 1335–1357.

Piani, C., Haerter, J. O., and Coppola, E. (2010). "Statistical bias correction for daily precipitation in regional climate models over Europe." *Theor. Appl. Climatol.*, 99(1–2),187–192.

Rajeevan, M., Bhate, J., Kale, J. D., and Lal, B. (2006). "High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells." *Curr. Sci.*, 91(3), 296–306.

Seiler, R. A., Hayes, M., and Bressan, L. (2002). "Using the Standardized Precipitation Index for flood risk monitoring." *Int. J. Climatol.*, 22(11), 1365–1376.

Sharma, M., Coulibaly, P., and Dibike, Y. (2011). "Assessing the need for downscaling RCM data for hydrologic impact study." *J. Hydrol. Eng.*, 16(6), 534–539..

Szalai, S., and Szinell, C. (2000). "Comparison of two drought indices for drought monitoring in Hungary—a case study." *Drought and drought mitigation in Europe*, J. V. Vogt and F. Somma, eds., Kluwer, Dordrecht, Netherlands, 161–166.

Tang, Y., Lin, H., and Moore, A. M. (2008). "Measuring the potential predictability of ensemble climate predictions." *J. Geophys. Res.*, 113(D4), D04108.

Wilby, R. L., and Wigley, T. M. L. (2000). "Precipitation predictors for downscaling: Observed and general circulation model relationships." *Int. J. Climatol.*, 20(6), 641–661.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P. (2004). "Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs." *Clim. Change*, 62(1–3), 189–216.

Xu, C. Y. (1999). "Climate change and hydrologic models: A review of existing gaps and recent research developments." *Water Resour. Manage.*, 13(5), 369–382.