

High-resolution multisite daily rainfall projections in India with statistical downscaling for climate change impacts assessment

Kaustubh Salvi,¹ S. Kannan,¹ and Subimal Ghosh¹

Received 2 September 2012; revised 3 January 2013; accepted 15 February 2013; published 13 May 2013.

[1] Climate change impacts assessment involves downscaling of coarse-resolution climate variables simulated by general circulation models (GCMs) using dynamic (physics-based) or statistical (data-driven) approaches. Here we use a statistical downscaling technique for projections of all-India monsoon rainfall at a resolution of 0.5° in latitude/longitude. The present statistical downscaling model utilizes classification and regression tree, and kernel regression and develops a statistical relationship between large-scale climate variables from reanalysis data and fine-resolution observed rainfall, and then applies the relationship to coarse-resolution GCM outputs. A GCM developed by the Canadian Centre for Climate Modeling and Analysis is employed for this study with its five ensemble runs for capturing intramodel uncertainty. The model appears to effectively capture individual station means, the spatial patterns of the standard deviations, and the cross correlation between station rainfalls. Computationally expensive dynamic downscaling models have been applied for India. However, our study is the first to attempt statistical downscaling for the entire country at a resolution of 0.5° . The downscaling model seems to capture the orographic effect on rainfall in mountainous areas of the Western Ghats and northeast India. The model also reveals spatially nonuniform changes in rainfall, with a possible increase for the western coastline and northeastern India (rainfall surplus areas) and a decrease in northern India, western India (rainfall deficit areas), and on the southeastern coastline, highlighting the need for a detailed hydrologic study that includes future projections regarding water availability which may be useful for water resource policy decisions.

Citation: Salvi, K., S. Kannan, and S. Ghosh (2013), High-resolution multisite daily rainfall projections in India with statistical downscaling for climate change impacts assessment, *J. Geophys. Res. Atmos.*, 118, 3557–3578, doi:10.1002/jgrd.50280.

1. Introduction

[2] South Asian summer monsoon is the source of 75% of total annual rainfall in major parts of southern Asia. More than 22% of the world's population resides in southern Asia and depends totally on the monsoon as their primary source of water [Dhar and Nandargi, 2003]. All-India monsoon rainfall (AIMR), the major portion of the south Asian monsoon, is a major factor controlling water resources, agriculture, and ecosystems throughout India. Therefore, since it is one of the key inputs for climate change adaptation and policy making, accurate projections of AIMR are of paramount importance. General circulation models (GCMs) are mathematical models that take into consideration the physics involved in various atmospheric, oceanic, and land processes in the form of a set of linear and nonlinear partial differential equations, and project climatic variables globally at a very coarse resolution. However, GCMs do not project

all climatic variables with the same level of accuracy. In general, GCMs show good skill in projecting large-scale circulation pattern but fail miserably when it comes to projections of rainfall [Ghosh and Mujumdar, 2006].

[3] The resolutions at which GCMs operate (generally more than 1.8°) directly hamper the accuracy of rainfall projections since subgrid features such as topography, cloud physics, and land surface processes cannot be incorporated into models that actually influence rainfall. Furthermore, the grid-level projection of rainfall at coarse resolution makes the situation worse for impact analyses that demand fine-resolution projections for regional decision making. Therefore, GCM simulations on a coarse resolution cannot be directly used in hydrologic impacts assessment. Downscaling techniques take coarse-resolution data as input and generate high-resolution data by applying various mathematical models. Such downscaling techniques are categorized based on the following headings: (1) dynamic and (2) statistical downscaling. Dynamic downscaling leads to the development of finer-scale physics-based models known as regional climate models (RCMs) that take input from GCMs' simulations as initial and boundary conditions, incorporate the subgrid features, and produce very high resolution results. Using various RCMs, dynamic downscaling has been attempted successfully for rainfall projections. Mesoscale atmospheric simulations for the projection of rainfall in the

¹Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai, India.

Corresponding author: S. Ghosh, Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India.
(subimal@civil.iitb.ac.in)

©2013. American Geophysical Union. All Rights Reserved.
2169-897X/13/10.1002/jgrd.50280

western United States [*Kim et al.*, 2000], the application of lateral boundary conditions (LBCs) in synchronous nested simulations for projecting climate variables [*Druyan et al.*, 2002], supplying the outputs generated by atmospheric GCMs to the regional spectral model developed by *Juang and Kanamitsu* [1994] with subsequent modifications in deep convection schemes for rainfall projections [*Misra et al.*, 2003], and applying HadRM3H (RCM) for simulating extreme rainfall over the United Kingdom [*Fowler et al.*, 2005] are some of the instances where dynamic downscaling is used for rainfall projections in different parts of the world. Using different RCMs such as the “PRECIS” [*Rupa Kumar et al.*, 2006; *Krishna Kumar et al.*, 2011], the “RegCM3” [*Moetasim et al.*, 2009], and the global atmosphere–ocean model ECHAM5 as the driving model and the “COSMO-CLM” as the regional model [*Dobler and Ahrens*, 2011], dynamic downscaling techniques have also been applied for the projection of AIMR. The various dynamic downscaling techniques used for fine-resolution climate modeling are summarized in Table 1.

[4] Regional climate models have the advantage of very fine resolution but are computationally expensive. On the other hand, statistical downscaling relies on data-driven approaches that involve deriving empirical relationships that transform the large-scale features of GCM-simulated climate variables (predictors) into regional-scale variables (predictand) such as rainfall. Statistical downscaling methods are computationally inexpensive and are useful if sufficient historical data are available for generating probability distribution functions and establishing statistical relationships. Statistical downscaling methodologies can broadly be classified

as weather generators, weather typing, and transfer functions. Markov chain models [*Bardossy and Plate*, 1991; *Hughes et al.*, 1993], spell length models [*Lall and Sharma*, 1996], nonparametric nonhomogeneous hidden Markov models [*Mehrotra and Sharma*, 2005; *Mehrotra and Sharma*, 2006; *Mehrotra and Sharma*, 2010], conditional random fields [*Raje and Mujumdar*, 2009], transfer function methodologies using an artificial neural network [*Hewitson and Crane*, 1992], the fuzzy clustering technique [*Ghosh and Mujumdar*, 2006], the support vector machine (SVM) approach [*Anandhi et al.*, 2008; *Ghosh*, 2010], the Bayesian joint probability (BJP) modeling approach [*Wang et al.*, 2009], and the stochastic space random cascade (SSRC) approach for downscaling rainfall from GCMs [*Groppelli et al.*, 2011] are some of the statistical downscaling techniques deployed for climate variable projections. The various approaches are summarized in Table 2.

[5] *Kannan and Ghosh* [2011] addressed the problem of multisite downscaling using representations of the pattern of multisite rainfall using the rainfall state in a river basin. A model based on a K-means clustering technique coupled with a supervised data classification technique, classification and regression tree (CART), is employed for the generation of rainfall states using large-scale atmospheric variables within a river basin. Nonparametric kernel regression is used for rainfall projections. *Kundu and Siddani* [2011] explored the dependence of the probability of zero rain on the averaging space and time scales in large multiyear data sets based on radar and rain gauge observations. It is also found that a stretched exponential formula fits the observed

Table 1. Dynamic Downscaling Methods

Sr. No.	Citation	Model Application Details	Result/Comments
1	<i>Kim et al.</i> , 2000	Model used: “Mesoscale Atmospheric Simulation” (MAS) Study region: Western United States (rainfall projection) and northern California coastal basin (streamflow projection)	High correlation between observed and simulated streamflow (0.88)
2	<i>Druyan et al.</i> , 2002	GCM used: NASA Goddard Institute for Space Studies (GISS) RCM: GISS/Columbia University Center for Climate Systems Research GCM simulations are interpolated to create LBCs for synchronous nested simulations of RCM. Study region: South America	Climatic variables are projected for March-April-May (MAM) 1985 and 1997. The downscaled predictions and the downscaled analyses both capture the meridional displacement of the intertropical convergence (ITC) rainfall maximum over northern Brazil between the two seasons.
3	<i>Misra et al.</i> , 2003	GCM used: Center for Ocean-Land-Atmosphere Studies (V 2.2) RCM: study regional spectral model developed by <i>Juang and Kanamitsu</i> [1994] with some modifications in deep convection schemes and land surface processes Study region: south Atlantic convergence zone Amazon River Basin and subtropical South America Resolution: 80 km	Observed and projected rainfall show fair match
4	<i>Fowler et al.</i> , 2005	RCM: HadRM3H Study region: UK	Study suggests that the study regional climate models have skills in predicting how rainfall extreme might change under enhanced greenhouse conditions.
5	<i>Rupa Kumar et al.</i> , 2006 and <i>Krishna Kumar et al.</i> , 2011	RCM: PRECIS Study region: India	9–16% increase in summer monsoon rainfall in future as compared to baseline period 1961–1990
6	<i>Moetasim et al.</i> , 2009	RCM: RegCM3 Study region: Asia Resolution: 0.25°	Overall suppression of summer monsoon rainfall
7	<i>Dobler and Ahrens</i> , 2011	GCM used: atmosphere-ocean model ECHAM5 RCM: COSMO-CLM Resolution: 0.44° Study region: India	70% decrease in the monsoon rainfall over century

Table 2. Statistical Downscaling Techniques

Sr. No.	Citation	Model Application Details	Result/Comments
1	Bardossy and Plate, 1991	Atmospheric circulation is divided into patterns and time series of each pattern is modeled using semi-Markov fields. Using the model, different circulation patterns and corresponding rainfall are simulated.	The statistics of observed and simulated sequences match well. The model is applicable for the simulation of nonstationary atmospheric conditions.
2	Hughes et al., 1993	GCM used: General Fluid Dynamics Laboratory Study region: Columbia River Basin The weather states based on present and previous day's sea level pressure are stochastically related to gridded precipitation and temperature.	Weather states are generated using present and previous day's sea level pressure and stochastically linked to gauge level rainfall and temperature.
3	Wilks, 1999	GCM used: HadCM2 Study region: United States Rain gauge stations: six groups of five U.S. stations are considered for downscaling. A stochastic weather generator model is developed and used for downscaling.	The model is validated for six diverse study regions in the United States by using natural climatic variability as a proxy for climate change. The output of the stochastic model is in a convenient form for stochastic simulation ("weather generation") in order to provide input for various impacts models.
4	Mehrotra and Sharma, 2005, 2006	Study region: Sydney, eastern Australia (30 rain gauge stations) A K-nearest neighbor-based nonparametric nonhomogeneous hidden Markov model is developed and applied for spatial downscaling of multistation daily rainfall occurrences using atmospheric circulations variables	The introduction of weather state results in significant reduction in dimension and also improves temporal dependence. The study shows that nonparametric continuous weather state formulation with optimized influence weights performs better than parametric discrete weather state alternative in terms of various statistics, e.g., spatial correlations, wet day probabilities, and maximum wet and dry spells in a season.
5	Raje and Mujumdar, 2009	GCM used: MIROC3.2 Study region: Mahanadi river basin, Orissa, India Conditional random field based downscaling technique is developed. Broyden-Fletcher-Goldfarb-Shanno optimization is used for maximum likelihood parameter estimation, and uncertainty is addressed through modified Viterbi algorithm.	CRFs can be used for high-dimensional feature vectors as they do not make assumptions on independence of observations. The results show increase in the probability of wet and high rainfall days.
6	Hewitson and Crane, 1992	Study region: Chiapas, Mexico Empirical relationship between sea level and 500 mb circulation field and local precipitation in Chiapas is established using Artificial neural network.	Neural network seems to capture the onset of the precipitation season and the phase of individual precipitation events which are controlled to a large extent by atmospheric circulation.
7	Ghosh and Mujumdar, 2006	GCM used: CCCma (CGCM2) Study region: Orissa The linear regression model is developed between GCM-simulated predictors and rainfall data. Dimensionality is reduced using principal component analysis, and fuzzy clustering technique is used to classify the components.	The model is computationally simple and still models rainfall with high value of goodness of fit.
8	Anandhi et al., 2008	GCM used: CCCma (CGCM3) The monthly precipitation is downscaled using support vector machine.	The results show that the precipitation is projected to increase in future for almost all the IPCC scenarios considered.
9	Wang et al., 2009	Study region: Queensland, Australia Bayesian joint probability (BJP) modeling approach is developed for seasonal forecasting of streamflows at multiple sites and extended for zero-flow occurrences.	The observed and projected streamflows are consistent.
10	Ghosh, 2010	Study region: Assam and Meghalaya SVM is used to establish relationship between predictors and predictand. Probabilistic Global Search Algorithm, a probabilistic search technique, is used to compute the optimum parameters of SVM. Uncertainty is modeled with modified reliability ensemble averaging method.	The results show significant improvement when compared with the results obtained with earlier developed SVM-based downscaling technique.
11	Groppelli et al., 2011	Stochastic space random cascade (SSRC) approach is developed to downscale rainfall from a general circulation models (GCMs), developed for the assessment of water resources under climate change scenarios for the Oglia river (1440 km^2), in the Italian Alps.	The SSRC approach reproduces well the spatial clustering, intermittency, self-similarity, and spatial correlation structure of precipitation fields and requires low computational power.

scale dependence of the zero-rain probability. Wang and Robertson [2011] extended the applicability of the Bayesian joint probability (BJP) modeling approach to streams with zero-flow occurrences with an aim to produce forecasts for these streams in the form of probabilities of zero seasonal flows and probability distributions of above zero seasonal flows.

[6] Precipitation is a highly heterogeneous spatial phenomenon that normally occurs as a result of the complex interaction between various climatic variables, the ocean, and the landmass. Therefore, the spatial modeling of precipitation is a major research challenge. A common problem encountered with multisite statistical downscaling is

modeling and capturing the intersite rainfall cross correlation. The problem becomes more challenging when downscaling is applied to a daily scale. In this work, the problem is managed using the rainfall state-based approach as introduced by *Kannan and Ghosh* [2013].

[7] The daily rainfall state is the single representative rainfall status for a given day expressed qualitatively for a region to which multiple sites of interest belong. The state is generated using a K-means clustering approach that helps categorize a day by considering rainfall at all of the rain gauge stations simultaneously. A statistical relationship is established between the predictors (the climate variables from the reanalysis data) that have undergone “principal component analysis (PCA)” and the rainfall state that is generated using K-means clustering. Assuming that the relationship holds for the future, future rainfall states are generated using the preestablished relationship and the predictors consist of either climate variables from the reanalysis data (the validation of the model using the reanalysis data) or the climate variable simulated using the GCM (for the 20th century and for future projections). For the current study, we selected the GCM developed by the Canadian Centre for Climate Modeling and Analysis. More details regarding this GCM, as well as the reasons behind selecting it for the analysis, are discussed in section 2. Conditional on the derived state, the kernel regression is applied for modeling multisite rainfall. The derived rainfall state is a proxy for multisite rainfall patterns and helps to model the cross correlation between rainfall from different stations. The present study consists of rainfall projections on a 0.5° resolution for the entire Indian landmass, using the methodology developed by *Kannan and Ghosh* [2011]. *Kannan and Ghosh* [2013] used smaller regions that consisted of eight stations. The objective of the current study is to apply a statistical downscaling technique in order to obtain future rainfall projections (complete for the 21st century) at a high resolution that will be useful for impact analyses for the entire Indian landmass, policies regarding the efficient usage and distribution of water resources taking into consideration alterations in future rainfall, and validating whether the physics behind the climatic processes related to rainfall is modeled effectively using the data-driven approach. Given the objectives, it is imperative to determine trend in average annual rainfall for India, as well as the spatial distribution, in order to comment on water availability for the future. The work described here is the first attempt at statistical downscaling for the entire area of India on a daily scale for a spatial resolution of 0.5° latitude \times 0.5° longitude. As a result of size and meteorological nonhomogeneity, the entire Indian landmass cannot be chosen as a single unit to determine the daily state generation. Therefore, for the current study, the seven meteorologically homogeneous zones identified by the India Meteorological Department (IMD) are utilized [*Parthasarathy et al.*, 1996]. For each zone, the selection of the spatial extent for which the predictors are to be considered as influencing factors, clustering, and the rainfall projections are performed separately. The results indicate that the model is capable of capturing spatially heterogeneous rainfall over the entire country, as well as for the inter-station cross correlation in a region.

[8] The current paper is organized as follows. The selection of the GCM and the data-related details are discussed in section 2. The methodology and the models used in the current paper are provided in section 3. A discussion of the various results obtained in the study is presented in section 4. Finally, a summary, followed by concluding remarks, is presented in section 5.

2. Data

[9] GCMs are based on the primitive equations of motion, retain sufficient resolution for representing atmospheric structure at synoptic and planetary scales, and include explicit representations of the main physical processes that determine atmospheric circulation on seasonal and longer time scales [*Boer et al.*, 1984]. The selection of the GCM plays an important role in the success of the downscaling. Coarse-resolution climate variables, which are used as predictors in the current study, are simulated using a third-generation coupled GCM (CGCM3.1) developed by the Canadian Centre for Climate Modeling and Analysis. Hereafter, this specific GCM is referred to as CCCma (CGCM3.1). The use of a truncated expansion in spherical harmonics to represent model variables [*Boer et al.*, 1984] and the use of semi-implicit time stepping [*Robert et al.*, 1972] with a weak time filter [*Asselin*, 1972] are some of the features of the model. The GCM is rich in data and provides us with projections for all of the important climatic variables for the special request emission scenarios (SRESSs) prescribed by the Intergovernmental Panel on Climate Change (IPCC). The rationale behind the selection of the CCCma (CGCM3.1) is as follows:

[10] 1. For the CMIP3 simulations, CGCM3.1 outputs are available for the entire 21st century. The main objective of the present work is to project rainfall for the same time period.

[11] 2. For the CGCM3.1, outputs are available for the five ensembles, which also allow us to assess intramodel uncertainty.

[12] The CCCma (CGCM3.1) can also simulate the interannual monsoon for India and can capture the variability [*Kripalani et al.*, 2007]. However, for assessing the performance of the model developed by *Kannan and Ghosh* [2013], predictors simulated by other GCM are employed with respect to the CCCma (CGCM3.1). AIMR is projected using coarse-resolution climate variables simulated by the GCM developed by “Geophysical Fluid Dynamics Laboratory” (GFDL) for the time slices 1971–2000, 2046–2065, and 2081–2100. The CCCma (CGCM3.1) provides data on two resolutions, viz., T47 (3.75° latitude/longitude) and T63 (2.8° latitude/longitude). For this work, we selected data with a resolution of T47, taking into consideration the data availability which is greater for this resolution.

2.1. Predictors

[13] Predictors are climatic variables that are well simulated by GCMs and that are used in statistical downscaling for predicting local-scale hydrological variable of interest such as rainfall (predictand). Ideally, the geophysical processes associated with rainfall are of a fine resolution. Therefore, coarse gridded GCMs fail to simulate them well. Predictors

that directly affect rainfall processes are used in statistical downscaling as input variables. The choice of predictor variables is of the utmost importance when it comes to the accuracy of projected data. The selection of predictors should be dependent on the following criteria: (1) the data for the particular predictor should be available for the desired period, (2) the selected GCM should be capable of simulating the variable well, and (3) the predictor should show a good correlation with the predictand [Wilby *et al.*, 1999]. Hewitson and Crane [1996] demonstrated how the downscaled projection of future change in mean rainfall and extreme events may significantly alter depending on the selection of predictor. Downscaled results can also depend on whether or not absolute or relative humidity is used as a predictor [Charles *et al.*, 1999]. The implication here is that while a predictor may or may not appear to be the most significant when developing the downscaling function for present climates, changes in that predictor for a future climate may be critical for determining climate change. Some estimation procedures, for example, the stepwise regression, are not capable of recognizing the changes and exclude variables that may be vital for climate change. Sharma [2000] adopted partial mutual information (PMI) criterion in order to identify significant season-wise atmospheric predictors. Johnson and Sharma [2009] developed a new approach based on a variable convergence score. Using this approach, variables are ranked based on the coefficient of variation for the ensemble. Dobler and Ahrens [2011] employed u-wind and v-wind in the form of wind shear at higher altitudes as predictors. Wind shear is basically the difference between the wind velocities at the 850 and 200 hPa pressure levels. For the current study, the climatic variables described by Kannan and Ghosh [2011] are used as predictors, viz., temperature, pressure, specific humidity, u-wind, and v-wind at the surface. The goal of our work is to project rainfall at a very high resolution for the entire 21st century. Pressure level variables are not considered in this study because they are not available for the entire period of 100 years of model simulation. However, rainfall projections using wind shears as additional predictors are performed for a single run in order to check whether or not the statistical downscaling technique provided any additional advantages in terms of computational power or accuracy over projections obtained using surface-level variables.

2.2. Reanalysis Data

[14] The reanalysis project is an outgrowth of the Climate Data Assimilation System project undertaken by the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR). A frozen state-of-the-art analysis/forecast system is used to perform the data assimilation using past data [Kalnay *et al.*, 1996]. The components of the assimilated data sets are the following: (1) global rawinsonde data, (2) a comprehensive ocean-atmosphere data set that comprises a collection of surface marine data, (3) aircraft data, (4) surface land synoptic data, (5) satellite sounder data, (6) special sensing microwave/imager surface wind speeds, and (7) satellite cloud drift winds. The spectral statistical interpolation, a 3-D variational analysis scheme, is used as an analysis module to assimilate the various data sets. The results are combined in order to obtain NCEP/NCAR reanalysis data for various climate variables. Climatic variables are categorized into three types [Kalnay *et al.*,

1996]. Category "A" variables are those that are strongly influenced by observations (e.g., zonal and meridional wind). Category "A" variables are highly accurate. Category "B" variables are influenced by the model and the observations, but are not as accurate as Category "A" level variables (e.g., specific humidity). For Category "C" variables, no observations directly affect the variable and they are completely generated by the model (e.g., precipitation flux). Additional information on the various types of NCEP/NCAR reanalysis variables can be obtained from Kalnay *et al.* [1996]. For our work, the NCEP/NCAR reanalysis-I daily data for surface air temperature, mean sea level pressure, specific humidity, zonal wind velocity, and meridional wind velocity for a region delimited by the latitudes 5°–40°N and longitudes 60°–120°E, surrounding the entire study area for a period of 30 years from 1971–2000, are utilized for the bias correction, the training, and the validation of the downscaling model.

2.3. Observed Rainfall

[15] Gridded daily data for India (latitude: 6.5–38.5, longitude: 66.5–100.5) at a 0.5° resolution are provided by IMD [Rajeevan and Bhate, 2008]. Gridded data are derived from rainfall data collected from more than 6000 rain gauge stations over India. The "Shepard method" [Shepard, 1968] is employed for interpolating station data into regular grids of 0.5° resolution. Even though rain gauge stations have been established all over India, the density of the rain gauge stations considered for the interpolation is not the same throughout India. In the Himalayan area of India, fewer rain gauge stations are available. Therefore, there is less confidence in the accuracy of the gridded data generated using station-level data in this zone. In total, there are 1245 grids at a 0.5° resolution that actually constitute the landmass for India, and rainfall is only projected for these 1245 nodes. In the discussions that follow, gridded rainfall data of 0.5° resolution are referred to as observed rainfall.

3. Methodology

[16] Statistical downscaling is the methodology by which coarse-resolution predictors are linked to the fine-resolution predictand using a statistical relationship. For the current study, we adopted the methodology developed by Kannan and Ghosh [2013]. Figure 1 provides a flowchart depicting the stepwise mathematical operations performed on the data for the rainfall projections. The GCM-simulated predictors and the observed rainfall, as a predictand, undergo different mathematical operations before actually becoming statistically linked. The predictors undergo a bias correction operation where the systematic error is removed using a quantile-based remapping technique [Li *et al.*, 2010]. The bias-corrected predictors go through a principal component analysis (PCA) that involves the application of orthogonal transformation on a set of correlated predictor variables, producing principal components. The resulting principal components are dimensionally reduced and uncorrelated to one another. Principal components carry almost the same variability as that of the original data. Hence, the PCA helps to reduce both dimensionality and multicollinearity. A reduction in the dimensions also results in a reduction in the computational effort.

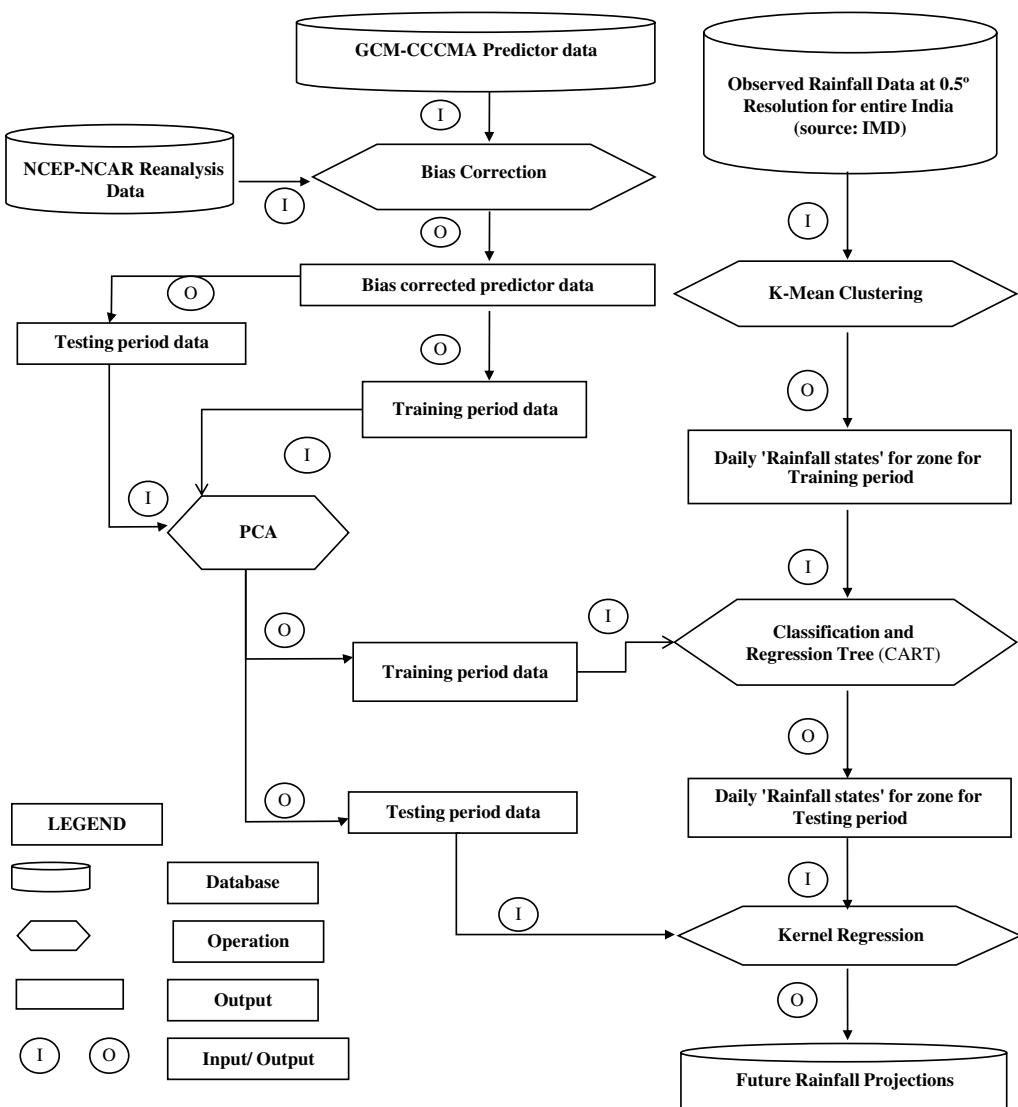


Figure 1. Flowchart for multisite statistical downscaling model enlisting various mathematical operations that are performed on predictors (GCM-simulated climate variables) and predictand (rainfall) which take part in statistical downscaling as inputs. The current statistical downscaling model [Kannan and Ghosh, 2013] is a combination of daily weather state generator and transfer function method.

[17] A K-means clustering technique is applied in order to individually derive the daily rainfall states for the seven Indian zones. The step helped to reserve cross correlation among rainfall for multiple grids in one zone. The daily rainfall states and the bias-corrected predictors, which all undergo principal component analysis, are key inputs to the kernel regression model for establishing the statistical relationship for the training period. Assuming that the relationship holds for the future, future states can be generated with the help of preestablished relationships and predictors for the future period. By applying a nonparametric kernel regression, rainfall is projected at each node. Rainfall projections are obtained based on the following three rationales:

[18] 1. For the model validations, the projections are obtained using NCEP/NCAR reanalysis climate variables as predictors. The period 1971–1985 is the time slice considered as the training period for which the statistical relationship is established. The period from 1986–2000 is the time slice

considered as the validation period for which projections are obtained and compared with the observed rainfall data for the same time period.

[19] 2. The 20th century projections are obtained from the 20C3M scenario of the CCCma (CGCM3.1) for the time slice 1971–2000.

[20] 3. Future projections are obtained from the same GCM for the time slice from 2001–2100 for three different scenarios: SRESA2, SRESA1B, and SRESB1. Details regarding the projections for each scenario are described in section 4.

[21] Each of the mathematical operations discussed above is illustrated in the flowchart in Figure 1 and is summarized in the subsequent text.

3.1. Bias Correction

[22] Due to incomplete knowledge regarding geophysical processes, assumptions are made in the development of

the GCM in terms of the parameterizations and the empirical formula. As a result of these assumptions, the GCM may not simulate climate variables accurately and there is a systematic difference, known as bias, between the observed and simulated climate variables for almost all of the GCMs. For projecting future hydrologic and climatic scenario correctly, it is important to remove the bias from the GCM output. Standardization [Wilby *et al.*, 2004] is used to reduce systematic biases in the mean and the variances in the GCM predictors relative to the observations or NCEP/NCAR data. The procedure typically involves a subtraction of the mean and division by the standard deviation of the predictor variable over a predefined baseline period in both NCEP/NCAR and GCM outputs. The methodology recently developed by *Li et al.* [2010] is a quantile-based mapping method that is based on the cumulative distribution functions (CDFs) of the reference data, the training data, and the future data. Reference data can either be station-level observed data or reanalysis data in the gridded format. *Johnson and Sharma* [2012] proposed a nesting model for bias correction that imparts correct distributional and persistence attributes and results in the sequence that is representative of reference data. For the current study, the methodology proposed by *Li et al.* [2010], as discussed in the following section, is used for the bias correction. The NCEP/NCAR reanalysis data are used as reference data.

3.1.1. Bilinear Interpolation

[23] The bias correction methodology developed by *Li et al.* [2010] demands that the reference data (observed data/reanalysis data) time series and the GCM-simulated variable time series be at the same location in terms of latitude and longitude so that their cumulative distribution functions (CDFs) can be compared. The GCM-simulated data used for this study are at a $3.75^\circ \times 3.75^\circ$ (latitude longitude) resolution. The NCEP/NCAR reanalysis data is at a $2.5^\circ \times 2.5^\circ$ (latitude \times longitude) resolution. Therefore, using the bilinear interpolation technique, first the GCM-simulated data are scaled to the NCEP/NCAR resolution, then the bias correction is performed.

3.1.2. Bias Correction for the Observed Period

[24] The availability of reanalysis data (station level or gridded) is an important asset that can directly be used for the bias correction of GCM-simulated data over the same period. Under ideal conditions (zero bias), GCM-simulated data will be exactly the same as that of the reanalysis data. As a result of the presence of bias, the two time series displayed a systematic difference. Figure 2 provides a pictorial representation of the bias correction methodology (Figure 2a), the bias correction on a sample grid in the form of a time series (Figure 2b), and the grid-level comparison of the mean in the form of India plots (Figures 2c–2e). The following provides a stepwise description of the methodology for removing the bias:

[25] 1. Fit the probability distribution for both time series (the reanalysis and the GCM simulated) on the same grid. In this study, we used the gamma distribution since it proved to be the best fit.

[26] 2. Generate the cumulative distribution functions (CDFs) for both time series.

[27] 3. Replace the individual values from the GCM-simulated time series using the data from the reanalysis data time series that has an equal CDF.

[28] The CDF matching technique adopted for the bias correction ensures that the statistical properties (the mean and the standard deviation) of the reanalysis data time series and the GCM simulate data time series match well.

3.1.3. Bias Correction for Future Data

[29] Future data lack the advantage of the availability of reanalysis data for the same period. Therefore, the bias correction of the future data is slightly different from that of the observed period. To remove bias in future data, the following steps should be followed:

[30] 1. Fit the probability distribution for the three time series (the reanalysis, the GCM-simulated observed period data, and the GCM-simulated future data) to the same node.

[31] 2. Generate the cumulative distribution functions (CDFs) for all of the time series.

[32] 3. Calculate the shift between the GCM-simulated future data time series and the GCM-simulated observed data time series by subtracting the GCM-simulated observed period data from the GCM-simulated future data time series at the same CDF level. The shift in the predictor variable is due to climate change. The step is essential because we cannot directly correct the future time series using past observations.

[33] 4. Replace projected data from the GCM-simulated time series for the future time period and the data from the observed data time series with the same probability, and add the shift (calculated in the previous step) in order to obtain a bias-corrected GCM-simulated future time series.

[34] The quantile-based remapping methodology provides fairly accurate results as far as the statistical properties of the reanalysis and the GCM-simulated time series are concerned. For a detailed description of the methodology, *Li et al.* [2010] can be referred. Figures 2c–2e provide the results for the bias correction performed over India in the form of the mean humidity at the grid level. The mean for the observed data (Figure 2c) and the mean for the bias-corrected GCM projected data (Figure 2e) revealed a good match.

3.2. Principal Component Analysis

[35] Bias-corrected data cannot be directly used for the development of statistical relationship since it suffers from two major impediments, viz., multidimensionality and multicollinearity. The multidimensionality problem, normally referred to as the curse of dimensionality, arises while dealing with high-dimensional space. The problem leads to an increase in the degree of sparseness in data that may impact outputs where a statistical significance is required. Multidimensionality also causes a drastic increase in the computational time required for the statistical analysis (e.g., multivariate regression). On the other hand, if the dimensions are reduced without considering the internal data pattern and its variability, the accuracy of the model output will get hampered. The identification of the pattern of multidimensional data is a tough task, especially when the comfort of visual representations is not available. A “trade off” always exists between multidimensions, computational power, and accuracy. Multicollinearity indicates the presence of a high correlation among two or more predictor variables in cases of multiple regression analyses and may lead to erratic changes in the coefficient estimation for small changes in data. Therefore, it is necessary to

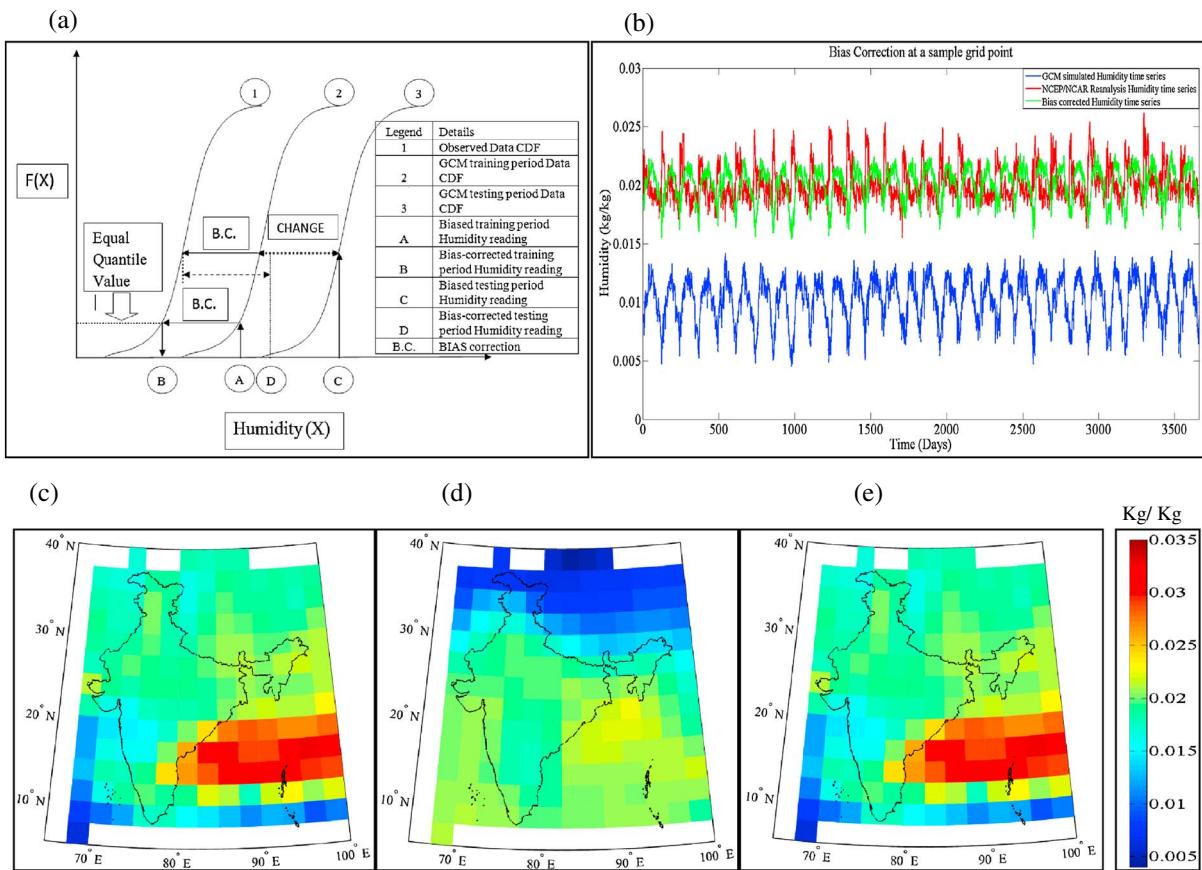


Figure 2. Bias correction methodology [Li *et al.*, 2010] applied to GCM-simulated predictors and sample results. The bias correction method is a CDF matching technique (a) in which the GCM-simulated values are replaced by observed values having equal probability (same quantile). Corrections for future are obtained with the changes projected by GCMs. Application of the method is demonstrated with humidity time series for (blue) observed, (red) uncorrected, and (green) corrected GCM simulations (b) at an arbitrarily selected location. Uncorrected GCM simulations (mean humidity for 1971–2000, monsoon) fail (d) to capture the spatial distribution as observed (c); however, the corrected simulations capture spatial variation (e).

identify these problems and to tackle them before the predictors actually get involved in the statistical downscaling. PCA is a powerful mathematical tool that is mainly used for identifying patterns in multidimensional data and reducing the number of dimensions without reducing the variability of the original data to the best possible extent. For the current study, PCA is used to obtain principal components from the large-scale climate variables acting as predictors in the statistical relationship.

[36] The meteorological homogeneous zones identified by the IMD are shown in Figure 3a [Parthasarathy *et al.*, 1996]. Each zone is treated as an individual entity, and for each zone, a corresponding region is fixed that is regular in shape (rectangle or square) and large enough to completely encompass the zone (Figure 3b). Rainfall in a particular zone is assumed to be influenced more by the predictors in the selected region surrounding the zone than the predictors that are outside the region. The regions are fixed after examining the correlation contour plots of the coefficient of correlation between the area-averaged observed rainfall time series in

the given zone and the individual predictor time series available at a 2.5° resolution after the bias correction, as seen in Figures 3c–3g. The validity of the assumption is depicted in terms of the correlation plot.

[37] Stepwise details of the PCA are explained using the central zone as an example. Figure 4b indicates the central zone as well as the region of predictors for the central zone. In the central zone, there are 64 nodes in the region of the predictors (Figure 3b). At each node, there are five predictors, implying that for rainfall projections in the central zone, we have in all $64 \times 5 = 320$ variables available as predictors. Considering all of the 320 predictors for projecting rainfall in the central zone will lead to the problems of multicollinearity and multidimensionality. Therefore, PCA is applied in order to avoid multidimensionality and multicollinearity problems.

[38] Figure 4 provides the region encompassing each zone, assumed to be the region of predictors for the corresponding zone. Each meteorologically homogeneous zone is highlighted with a gray shade and the region of predictors is shown with a black rectangle (refer to

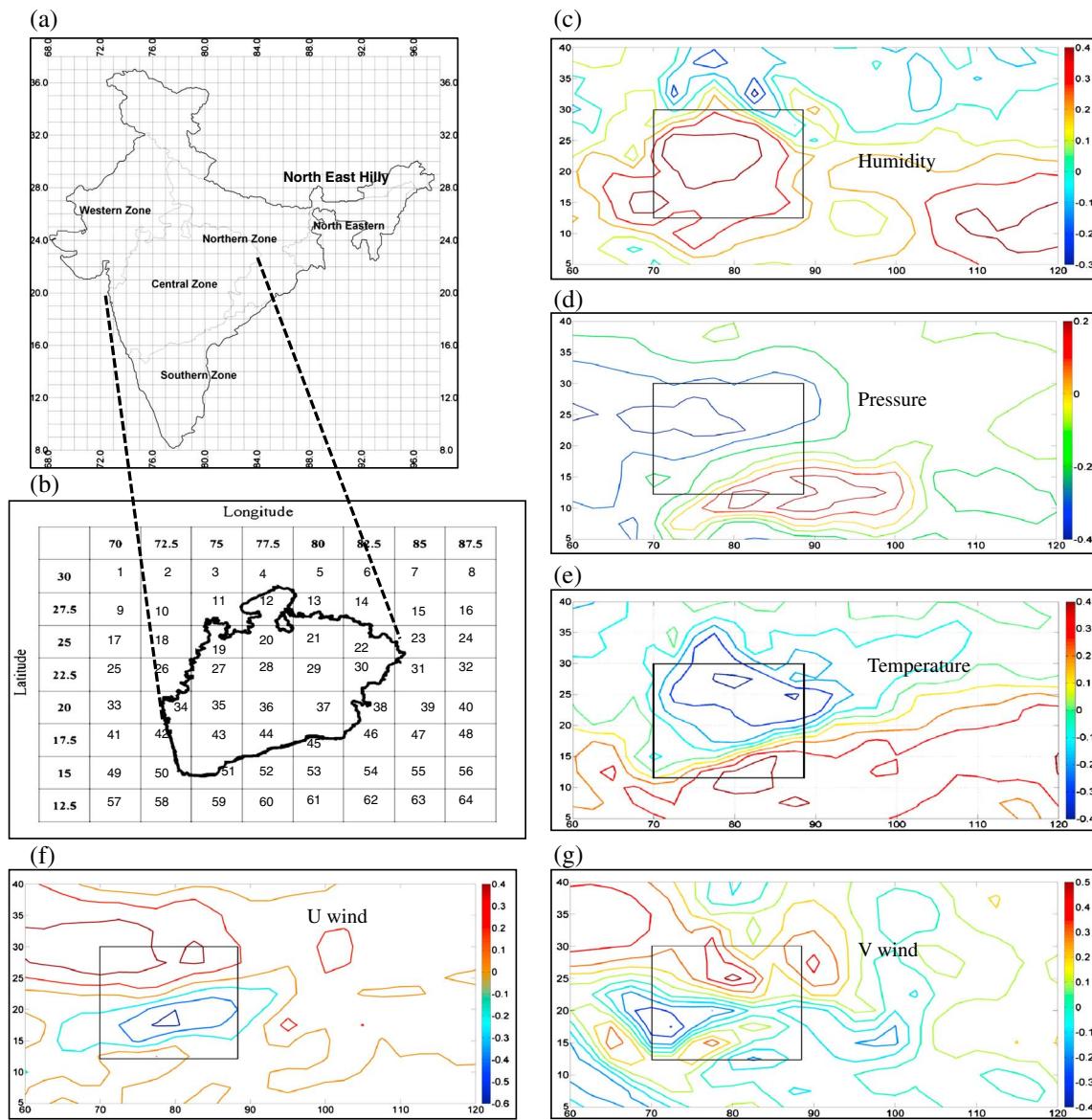


Figure 3. Spatial extent of predictor selection for meteorologically homogeneous subdivisions in India (a) and is illustrated for central zone (b). The correlation contours between zonal average rainfall time series and (c) humidity, (d) pressure (e), temperature (f), u-wind, and (g) v-wind are used, and a common region is selected from all plots (shown by black rectangle) in such a way that it encompasses central zone completely and contains the region having high correlation contours for all predictors.

Figures 4a–4g). The exact locations of the region of predictors in terms of latitude and longitude boundaries are shown in Figure 4h in the form of a table. From this point forward, these regions are referred to as the region of predictors.

[39] The obtained principal components are uncorrelated and carry most of the variability present in the original data. No single criterion exists for selecting the number of principal components [Kannan and Ghosh, 2013]. For this study, we selected the number of principal components based on the percentage of variability that they collectively represented. Our threshold is 98% (i.e., the total percentage of the variability carried by the selected principal components should be more than or equal to 98% of the original data). The same approach is adopted in order to obtain the principal components for the remaining zones in India. The operation

performed on the predictand data (rainfall), the establishment of a statistical relationship, and actual rainfall projections are discussed in subsequent sections.

3.3. The Daily Rainfall State Estimation

[40] Capturing the cross correlation among the rainfall at nearby rain gauge stations or nearby nodes (in the case of gridded data) is one of the research problems in multisite statistical downscaling. Cross correlation is the correlation between the rainfall time series at two or more rain gauge stations. Rainfall is a spatial phenomenon. Thus, it is highly unlikely that two rain gauge stations that are in the proximity of one another receive rainfall with magnitudes that are significantly different. However, when we attempted single-site statistical downscaling (i.e., projecting rainfall at

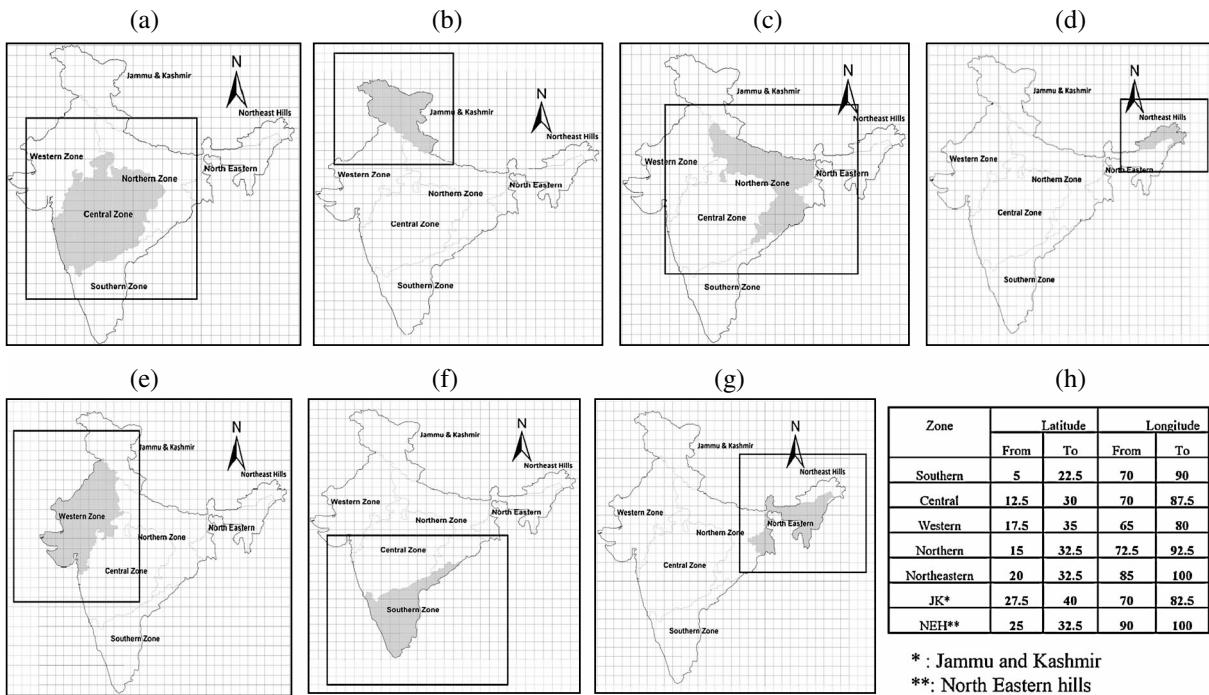


Figure 4. The region of predictors (shown by black rectangles) for each meteorologically homogeneous zone (shown by gray shade) is illustrated: (a) central, (b) Jammu and Kashmir, (c) north, (d) northeast hills, (e) western, (f) south, and (g) northeast. The extent of region of predictors in terms of latitude and longitudes is detailed in Figure 4h.

individual stations), we actually neglected the influence of rainfall occurring at nearby points on the location where downscaling is performed. As a result, the model failed to capture the spatial pattern of rainfall that could be visualized using cross-correlation plots. Mehrotra and Sharma [2005] introduced the concept of the weather state in order to tackle the multisite problem. The weather state captures the spatial distribution of rainfall on a particular day over a selected region. Kannan and Ghosh [2011] adopted an unsupervised data classification technique, K-means clustering, in order to group together the rainfall data at all nodes in one zone and to assign a state in order to represent the data. The same approach is used in this study to negotiate the multisite problem. K-means clustering is a method of cluster analysis that aims to partition n observations into K clusters for which each observation belongs to the cluster with the nearest mean. In the current study, K-means clustering is used to deal with the spatial variation of rainfall in a particular zone. In order to preserve the same imprint of spatial variation in further analysis and to maintain the cross correlation among the grid points within a particular zone, the K-means clustering technique is adopted. The technique reads the observed rainfall values for all nodes in a zone on any day, clusters them, and provides one representative value that is referred to as the state for that day. The step is important, provides the representative rainfall category for a particular day, and is linked to the predictors for establishing the statistical relationship. The total number of clusters that formed from the data should be optimized to represent the true classification that is inherently present in the data. Clustering is used to classify multisite rainfall for each zone into different states (in the form of clusters). Clusters can be considered as groups into which we divide the

data. Each cluster has a corresponding state to which it is associated. Hence, the state can be viewed as an identity number assigned to each cluster. The observed rainfall data for each zone are clustered using the K-means clustering technique. The total number of clusters presented in the data is determined to be three after comparing the Dunn's index and the Silhouette index obtained for different combinations of numbers of clusters (the number of clusters varied from 2 to 11). The highest values for these indices, indicating the optimum number of clusters in the data, are obtained when the data are classified into two clusters. However, two clusters imply a dry wet day classification. Therefore, in order to capture the spatial patterns of nonzero multisite rainfall [Kannan and Ghosh, 2013], the optimum number of clusters is identified as three for which the second highest values of the cluster validity indices are obtained. Since we used a three-cluster formation, based on rainfall in all of the nodes, any zone is assigned one of the three state values (1, 2, and 3). After fixing the states, the next step is to model the state using large-scale atmospheric data. For this determination, we used the classification and regression tree (CART).

3.4. The Statistical Relationship Using Classification and Regression Trees

[41] The bias-corrected predictors that undergo PCA and the rainfall states obtained using the K-means clustering serve as inputs for establishing the statistical relationship. Since rainfall state is a categorical variable, we cannot use a normal multivariate regression analysis. Instead, we used a CART analysis, a decision tree learning technique. For this study, the CART analysis is used to obtain the relationship

between the principal components and the rainfall states for the observed data periods, and to generate future rainfall states using the developed relationship and the principal components for the future period (as simulated by the GCM). In general, the purpose of the analysis using tree-building algorithms is to determine a set of logical conditions (split) that permit accurate predictions or classifications for cases. A binary recursive partitioning algorithm is employed for building trees. The algorithm is an iterative procedure that splits the data into child nodes and applies the conditions at those nodes. The procedure repeats itself until some rule is met or when further splitting is not possible. The CART analysis is applied based on an assumption that the relationship developed for the observed data period holds for the future, as well. Once the future states are simulated, kernel regression is used conditional on the derived states in order to project future rainfall.

3.5. Kernel Regression

[42] Nonparametric statistical downscaling techniques such as the k-nearest neighborhood are widely used for multisite rainfall projections [Lall and Sharma, 1996; Yates et al., 2003; Mehrotra et al., 2004; Mehrotra and Sharma, 2006]. Any nonparametric regression estimator is generally a smoothing filter that projects the predictand for a desired set of predictors by applying weights to the other predictand in the neighboring region of the one desired. In general, a weight function is deployed in order to fulfill this task, which assigns heavy weights to the nearby data and very low weights to data that are far away. Multivariate kernel regression is a nonparametric technique in statistics that is used for estimating the conditional expectation of a random variable. The objective is to capture the nonlinear relationship between a pair of random variables X and Y [Wand and Jones, 1995]. Therefore, for the present case, by considering the nonlinear relationship between daily rainfall and the predictors, this technique is adopted. Mathematically, the general form for the conditional expectation of the kernel regression can be written as follows:

$$E(Y|X) = m(X) = \frac{\int y f(y|x)}{f(x)} \quad (1)$$

where Y is the set of predictands at the node level, X is the set of principal components, $f(y|x)$ is the conditional probability density function (pdf) of Y given $X=x$, and $f_X(x)$ is the marginal pdf of X . Nadaraya [1964] provided an estimator by replacing the multivariate pdf with kernel density estimates.

$$m_h(x) = \widehat{\sum_{i=1}^n K_h(x-X_i) Y_i} \quad (2)$$

where $m_h(x)$ is the expected value Y for a condition of $X_i=x$, and K_h is the kernel with bandwidth h .

[43] Efromovich [1999] indicated that the shape of the kernel estimate is heavily dependent on the bandwidth selection. For this study, the asymptotic mean integrated square error for the multivariate kernel density estimator is used for the bandwidth estimation [Scott, 1992; Wand and Jones, 1995]. Kannan and Ghosh [2013] provided additional details regarding bandwidth selection. In this study, three different kernels are used based on the projected state of rainfall as computed in the CART analysis. For example,

if the state for day 1 (the future period) is 2, then all the rainfall values for those days in observed data period, having state as 2, will be used in the kernel to generate the day 1 rainfall. The kernel used here is an exponential function that assigned probabilistic weights based on the difference between the predictor vector for the testing period (day 1) and the training period predictor vectors. The advantage of using an exponential function is that if the difference is small (the day 1 predictors for the testing period well resembled a particular day predictor vector during the training period), a very high weight is assigned. As the distance increases, a low weight is assigned to that day in the training period. Details of the methodology may be found in Kannan and Ghosh [2013].

4. Results and Discussion

[44] Using the statistical downscaling technique, rainfall projections for the SRES are obtained at a 0.5° resolution for the entire area of India. The obtained results are validated using various statistical checks. In this section, we provide a detailed discussion regarding the obtained results and their interpretation.

[45] The capabilities of the model in terms of accuracy and spatial variations are validated by projecting rainfall using NCEP/NCAR reanalysis climate variables as predictors and by comparing the results to the observed data. The training period is selected as 1971–1985 and the testing period is selected as 1986–2000. Figure 5 provides plots of the comparison of statistical properties of observed and projected rainfall at the grid level. The spatial distribution of mean rainfall shown by the observed data is well replicated by projected mean rainfall (Figures 5a and 5b). Although the maximum absolute difference between observed mean rainfall and projected mean rainfall is approximately 10 mm, for most of the parts of India, the difference is around 3 mm (Figure 5c). As compared to other sections of India, the difference is greater where actual mean rainfall is greater. The difference of approximately 3 mm in the projection of the mean for a majority of the nodes indicates that the projected mean rainfall matched well with the observed mean rainfall. The difference between the observed standard deviation and the projected standard deviation is delimited between ± 20 mm (Figure 5f). A comparison between the plots in Figures 5d and 5e indicates that the spatial pattern (the spatial variability) of the standard deviation shown by observed rainfall data is captured by the projected rainfall data. However, the standard deviation for observed rainfall is greater in magnitude as compared to that of projected rainfall, as evident from Figure 5f that shows the dominance of a blue color indicating a higher standard deviation for observed rainfall. Although the maximum difference is delimited between ± 20 mm for the majority of the nodes (more than 75% of the grids points), the difference between the standard deviations for the observed and projected rainfall differed by 7.5 mm/d.

[46] The 20C3M is a scenario that deals with 20th century projections. Therefore, for this scenario, rainfall is projected using GCM-simulated predictors and is compared with observed data in order to validate the capability of the GCM in terms of the projections of the climate variables for all five available simulations. Since CCCma (CGCM3.1) has five runs available for each scenario, ensemble mean

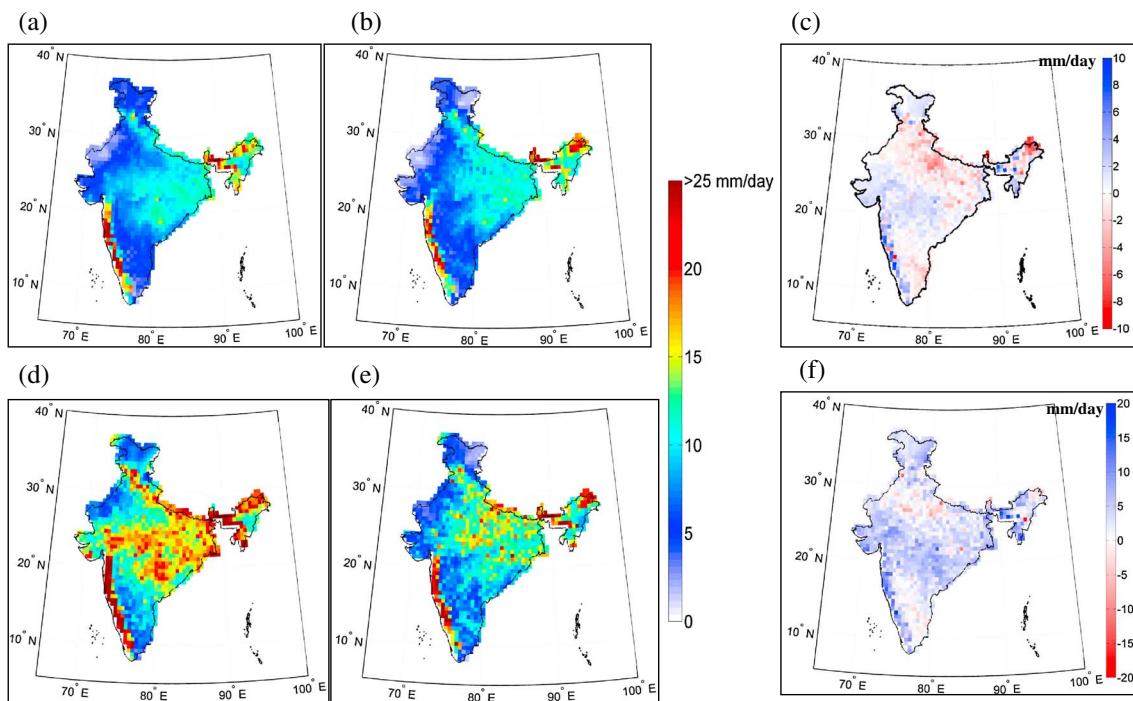


Figure 5. Comparison of statistical properties of observed rainfall and projected rainfall (using NCEP/NCAR reanalysis data predictors in downscaling) for entire India at 0.5° resolution (1986–2000). The gridded mean of (a) observed rainfall data and (b) projected rainfall data for the entire India at 0.5° shows good match in magnitude as well as spatial variability resulting in low error values (c) with ranges around –3 to 3 mm for the majority of the country. For few grids where mean rainfall is more as compared to other grids in India, error magnitude is also more. The standard deviation of observed rainfall data (d) shows higher values at grids when compared with the standard deviation of the projected rainfall (e) resulting errors (f) in the ranges around –10 to 10 mm for most of the grid in India. The higher magnitude of error and the dominance of blue color for error plot indicate that the model is underestimating rainfall variability.

results are obtained in order to deal with intramodel uncertainty. As mentioned previously, five surface-level GCM-simulated variables are used as predictors in the current study. To validate the effect of the inclusion of high altitude variables on the accuracy of the projection, one simulation is performed using wind shear as an extra predictor [Dobler and Ahrens, 2011]. Figure 6 provides the grid-wise details of the mean and standard deviation comparison between observed and projected rainfall (the 20C3M ensemble). Figures 6a and 6b indicate that the spatial pattern of mean observed rainfall is well captured by the projected rainfall obtained using a multisite downscaling technique. Along with the spatial pattern, the model has captured the mean with a high degree of accuracy, as evident from Figure 6d that provides the error plot obtained by subtracting the projected rainfall mean from the observed rainfall mean. A major portion of the Indian landmass has a range of error between –2.5 and 2.5 mm/d. Some of the nodes in the western coastal region and in the northeast displayed an absolute difference of 5 mm/d for the mean. However, the mean rainfall in those regions is also very high compared to mean rainfall in the rest of the country. The addition of wind shear as a predictor did not significantly improve the accuracy of the mean projections, as evident from Figures 6b and 6c. Similar to the rainfall projections obtained using NCEP/NCAR predictors, the spatial pattern (the spatial variability) of the standard deviation is also

captured by the model (refer to Figures 6e and 6f). The difference in the standard deviation for the observed and the projected rainfall depicted in Figure 6h provides overall positive differences, indicating that the model projects the rainfall values that are close to mean rainfall resulting in a lower magnitude for the standard deviation as compared to the observed rainfall. Again, the inclusion of high altitude variables did not significantly improve the results for the standard deviations (Figures 6f and 6g). Here, as well as for the kernel regression, GCM-simulated predictors are also not good for capturing the fluctuations since GCMS are not good for simulating variability.

[47] Cross correlation between rain gauge stations is an important property depicting the spatial nature of rainfall. The K-means clustering technique is adopted in order to cope with the multisite downscaling problem. The technique takes care of the cross correlation of projected rainfall for different grid points within the same zone. If a particular zone has n grid points, by using two stations at a time, the total number of cross correlation coefficients for the given zone will be ${}^n C_2$ (for any two selected from n points). The results are validated for the ${}^n C_2$ correlation values within each zone. Figure 7 provides a plot that displays the cross correlation between the observed rainfall versus the projected rainfall for the same time period for the grid points of all of the zones. The plots indicate that the zone-wise cross correlation among the observed rainfall data in various

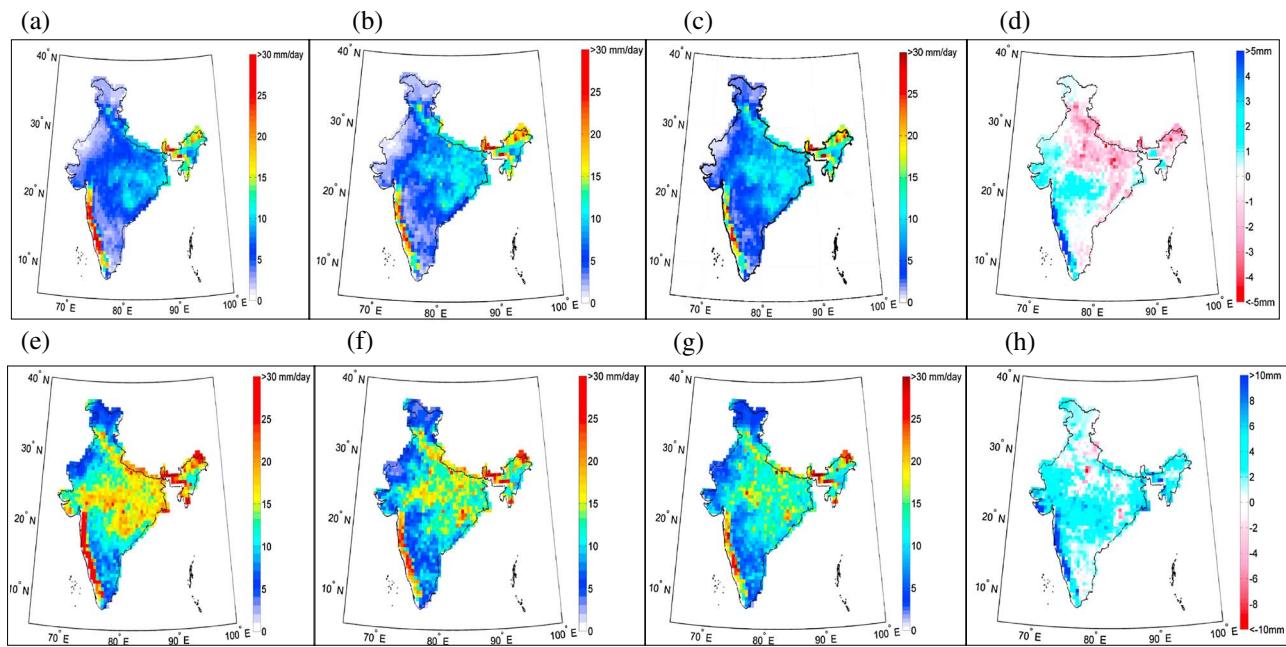


Figure 6. Comparison of statistical properties (mean and standard deviation) of observed and projected downscaled GCM-simulated rainfall (20C3M) for the entire India at 0.5° resolution (1971–2000). The GCM simulations are presented as ensemble average from five runs. The mean of observed rainfall data (a) and of the projected rainfall data obtained using five surface-level predictors (b) for the entire India at 0.5° show good match in terms of magnitude and spatial variability, leading to errors in the range of 2 to ± 2 mm for most of the Indian landmass (d). Inclusion of high altitude predictors does not improve the simulation performance (c) as compared to that in Figure 6b. Similarly, the standard deviation of rainfall for observed (e), downscaled with surface predictors (f), downscaled with surface and high altitude predictors (g), and error plots (h) is presented. The model is underestimating the variability.

nodes is nicely captured by the projected rainfall data for various nodes within the same zone.

[48] Rainfall projections using the NCEP/NCAR predictors and the 20C3M scenario predictors mainly served the purpose of validating the model of *Kannan and Ghosh* [2013], as well as the capability of the GCM. High-resolution rainfall projection results for the 21st century are discussed in subsequent sections. Figure 8 provides complete projection results for the future scenarios A2, A1B, and B1 in the form of a change in mean rainfall with respect to mean rainfall projected for the 20C3M scenario. As far as the intensity of carbon dioxide (CO_2) is concerned, A2 is the most critical scenario (850 ppm of CO_2) and provides maximum deviations in mean rainfall with respect to the 20C3M mean as compared to the deviations for scenarios A1B and B1. The CO_2 concentration decreases from 850 to 550 ppm from A2 to B1. A similar decrease in the deviation from the mean is seen as we move from scenario A2 to A1B and B1. Figures 8a–8i provide the ensemble difference in mean rainfall between A2, A1B, and B1 (future scenarios) and the 20C3M. As far as differences in mean rainfall (as compared to the mean rainfall projected for the 20C3M scenario) are concerned, a spatial nonuniformity is clearly seen in the plots irrespective of the scenarios. Mean rainfall values indicate an increase in some regions and a decrease in others. The western coast, the central portion of India, and the northeast region show an increase in daily mean rainfall for the future, for northern India, and for the southern portion of the leeward side of the Western Ghats. The southeastern coast of India displays a

decrease in daily mean rainfall. The deviation (an increase or a decrease) seems to be more intensified in magnitude with time, as well as with the stringency of the scenarios based on the carbon dioxide content. The magnitude of the increase in mean rainfall is more intense for the western coast and northeast India, the zones of maximum rainfall in India. For central India, there is also an increase in mean rainfall. However, the magnitude of the increase is less as compared to the western coast and northeast India, possibly because central India itself receives less rainfall. Hence, changes are present but not prominent. The role played by orography is also illustrated in the plots. The windward sides of the mountains show an increase in mean rainfall for the future. The leeward sides display a decrease for future rainfall. All of the plots display one common but important aspect of projection—the consistency of change. Regions showing positive change in mean rainfall maintain the same trend irrespective of the time slice and scenario. The same is true for regions showing negative changes. The behavior should be analyzed in further detail. However, it is not a part of the scope of the current study.

[49] In order to check the performance of the model developed by *Kannan and Ghosh* [2013], rainfall projections are obtained using the same predictors, but simulated by different GCM. Here we used the GCM developed by the GFDL. Performance validations are restricted to a single simulation for the scenarios 20C3M and SRESA2. At the surface, specific humidity data are not available for this GCM.

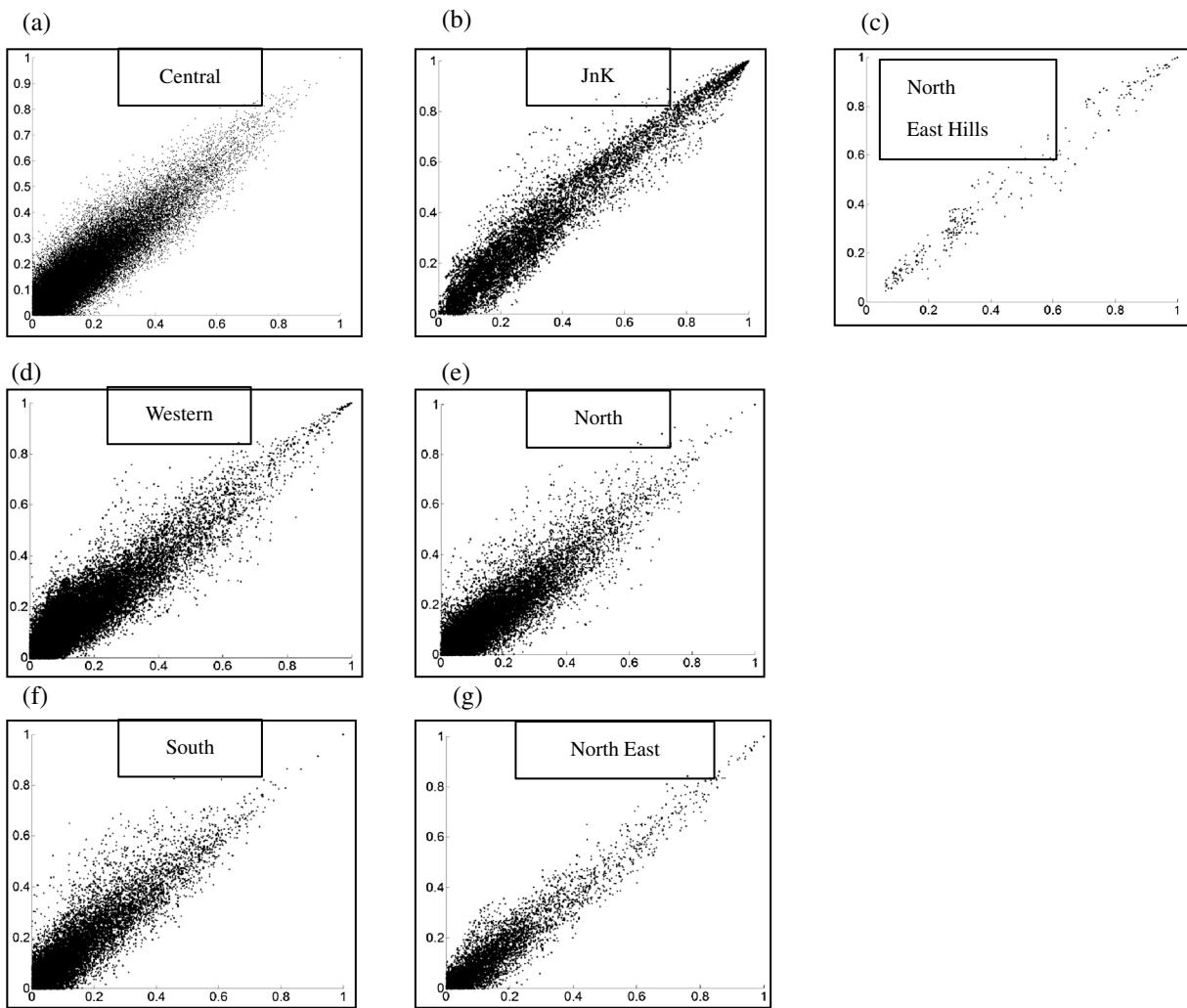


Figure 7. Scatter plots for zone wise cross correlations of multisite rainfall between observed and projected. The zones are as follows: (a) central, (b) Jammu and Kashmir, (c) northeast hilly region, (d) western, (e) north (f), south, and (g) northeast.

Therefore, we chose a specific humidity at 1000 Hp as one of the predictors. Figures 9a–9f provide a comparison of the projections in the form of a bias for the 20C3M (1971–2000) scenario and in the form of changes in mean rainfall (an increase or a decrease with respect to the 20C3M) for the SRESA2 scenario (2046–2065 and 2081–2100). A comparison of the bias between the rainfall projected using the CCCma (CGCM3.1) and the GFDL (Figures 9a and 9b) did not indicate a large difference for the magnitude of the bias, as well as the pattern, with the exception of the Gujarat Region where the GFDL model failed to capture the mean. Therefore, the results are heavily biased. For future projections, the spatial pattern indicated by the CCCma (CGCM3.1) and the GFDL matched fairly well, with the exception of the Gujarat and within the Jammu and Kashmir zones. For other regions (e.g., the west coast, central India, and the east coast), the spatial pattern remained the same. As far as the magnitude of change for mean rainfall as considered for future time slices, the GFDL provided a heavy difference between mean rainfall projected for the SRESA2 scenario and the 20C3M as compared to the CCCma (CGCM3.1), which is evident from the darker color shades (Figures 9c–9f). The exercise is

performed for model validation purpose using different GCM, not to compare the results obtained using two different GCMs in order to comment on their accuracy.

[50] Extreme rainfall events normally have a catastrophic impact on society. In this study, extreme events for various scenarios are examined in the form of return periods on each grid. To test the model capability on the projection of extremes, 30 year return period rainfall values are compared between the observed and projected rainfall using NCEP/NCAR, and between observed and projected rainfall for the 20C3M scenario. The comparison period for rainfall projected using the NCEP/NCAR predictors is considered from 1986 to 2000, and the period for rainfall projected using the 20C3M predictors is 1971–2000. For both cases, annual maximum values for rainfall are determined for the desired period, and the Gumbel distribution is fitted for annual maximum rainfall values in each grid. Thirty year return level rainfall values are obtained and compared. Figures 10a–10c provide comparison plots between the 30 year return levels obtained using observed rainfall, projected rainfall using NCEP/NCAR predictors, and rainfall using the GCM-simulated predictors for each grid. The plots indicate

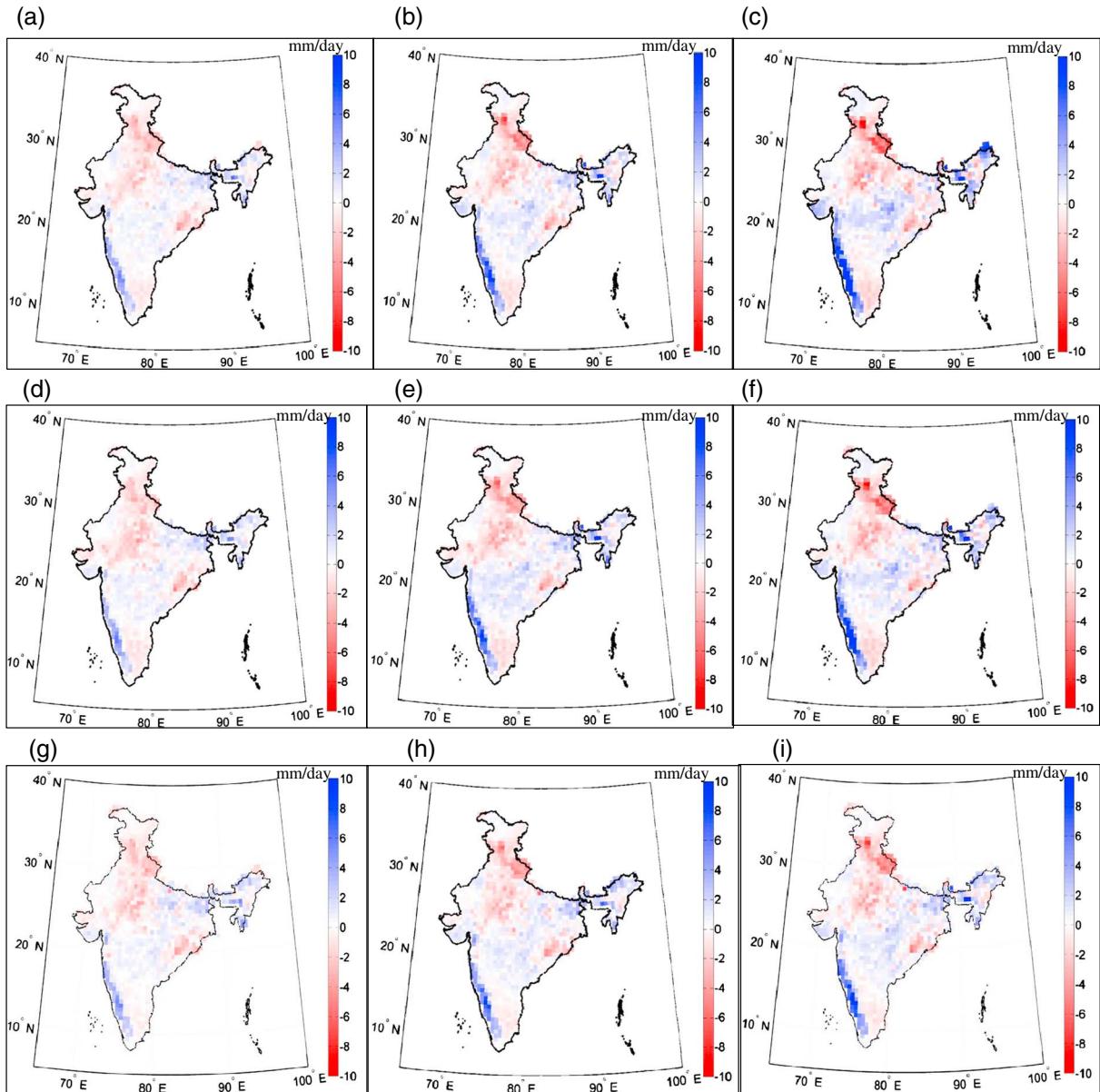


Figure 8. Rainfall projections for future for scenarios SRESA2, SRESA1B, and SRESB1. The plots present changes for A2 scenario in 2020s (2010–2039) (a), 2050s (2040–2069) (b), and 2080s (2070–2099) (c); for A1B scenario in 2020s (2010–2039) (d), 2050s (2040–2069) (e), and 2080s (2070–2099) (f); for B1 scenario in 2020s (2010–2039), (g) 2050s (2040–2069) (h), and 2080s (2070–2099) (i). Future projection plots reveal spatially nonuniform changes (increase at some grids and decrease at some grids). The west coast, central India, and northeast India show increase in the mean as compared to 20C3M, whereas east coast (southern part) and north India show decrease in the mean rainfall. This pattern is consistent irrespective of the scenario and time slice under consideration. A2 presents maximum changes being the worst-case scenario.

that extreme rainfall is not captured very well by the model. Extreme rainfall events are mainly captured using fluctuations in the rainfall series. Fluctuations are indicated by high standard deviations. Since the model did not capture the standard deviation with the same accuracy as that of the mean, extreme rainfall events are not modeled well. Figure 10d provides the 30 year return level of rainfall for the SRESA2 scenario for the near future (2010–2039). The results indicate spatial nonuniformity, with an overall decrease.

[51] The rainfall projection results discussed here are basically the ensemble of five runs for different scenarios: 20C3M, A2, A1B, and B1. Since the CCCma (CGCM3.1) provides five simulations for each scenario, rainfall that is projected separately for the five different simulations is utilized for quantifying intramodel uncertainty. The use of various physics-based mathematical models, different numerical schemes to obtain solutions, different initial and boundary conditions supplied to the same numeric scheme,

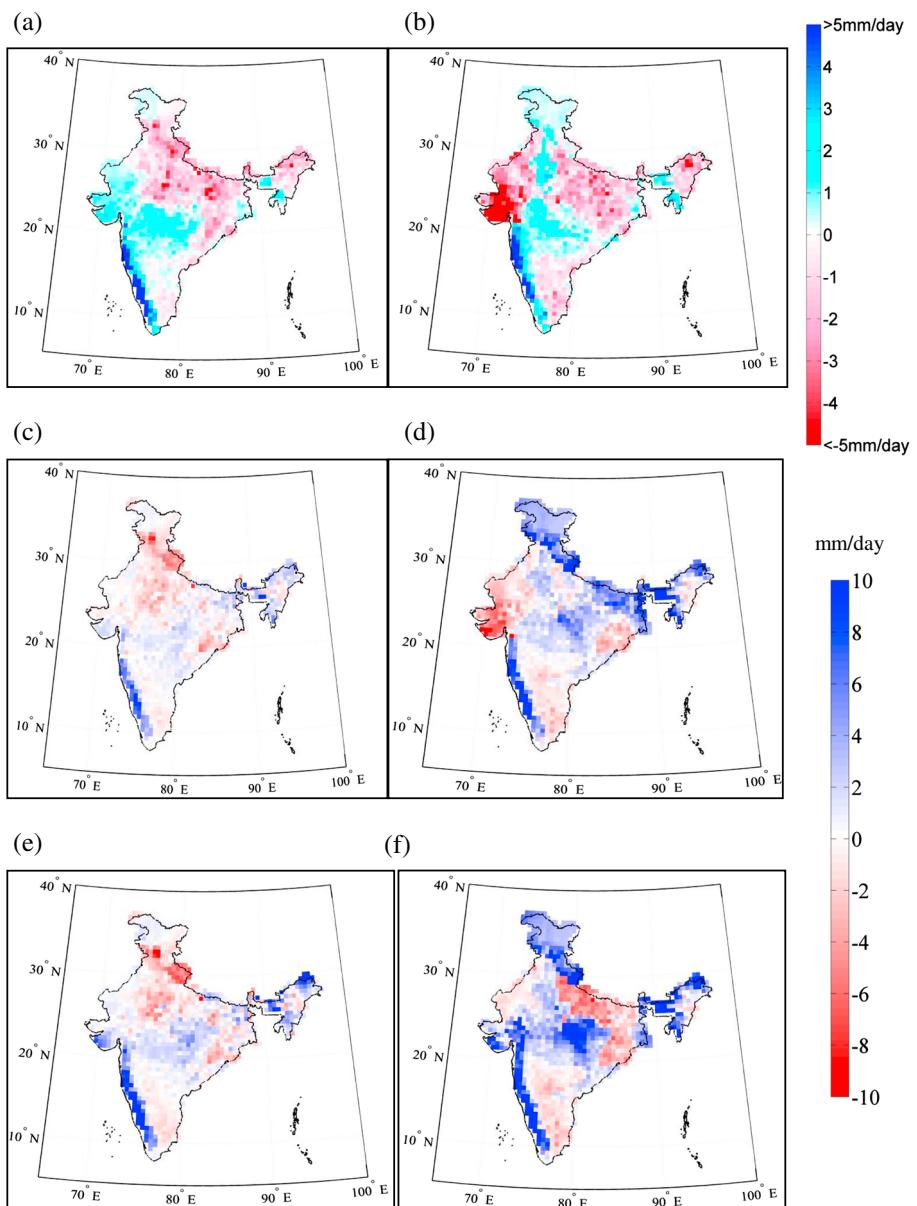


Figure 9. Model performance comparison between the projected rainfalls in terms of bias (for 20C3M) and changes (for future) for two GCMs CCCma (CGCM3.1) and GFDL are displayed. (a) Difference between observed mean rainfall and projected mean rainfall for CCCma (CGCM3.1) for the period 1971–2000. (b) Difference between observed mean rainfall and projected mean rainfall for GFDL for the period 1971–2000. (c) Difference between mean rainfall for SRESA2 scenario (2046–2065) and 20C3M for CCCma (CGCM3.1). (d) Difference between mean rainfall for SRESA2 scenario (2046–2065) and 20C3M GFDL. (e) Difference between mean rainfall for SRESA2 scenario (2081–2100) and 20C3M for CCCma (CGCM3.1). (f) Difference between mean rainfall for SRESA2 scenario (2081–2100) and 20C3M GFDL.

and downscaling techniques contributes to various types of uncertainties. For the current study, we performed an assessment of the uncertainty. An actual uncertainty analysis is beyond the scope of this work.

[52] In simulation number 1 using the 20C3M (the 20th century), rainfall is projected for the period from 1971 to 2000, and using the SRESA2 (the 21st century) for three time slices: the 2020s, the 2050s, and the 2080s. We subtract the 20C3M (30 year) mean from the 2020s, the 2050s, and

the 2080s (30 year) at each node in order to obtain the variation of mean rainfall for the future as compared to the 20th century. The procedure is repeated for all five simulations in order to obtain five differences for each time slice (2020s, 2050s, and 2080s) at each node. The minimum and maximum for these five values indicate an intramodel uncertainty, and the average of these five values indicates the ensemble mean difference between 21st century mean rainfall and 20th century mean rainfall for all three time slices at each node.

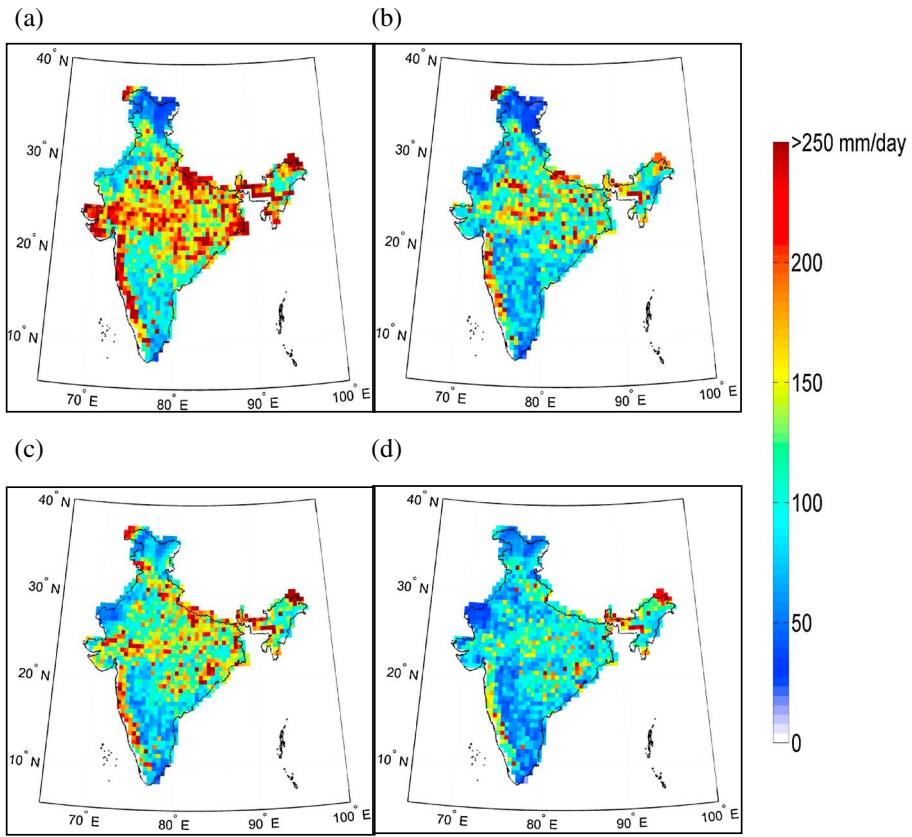


Figure 10. Thirty year return level rainfall (computed with Gumbel distribution) comparison between observed and projected rainfall at 0.5° resolution. The observed return levels are on higher side (a) as compared to projected form NCEP/ NCAR (b) and GCM (c). The underestimation of return levels is due to underestimation of rainfall variability and is a common limitation for any statistical downscaling model. The projections in near future for A2 (d) do not show spatially uniform changes, though most of the grids show decreasing pattern.

The plots shown in Figure 11 reveal the quantified uncertainty for the SRESA2 scenario. Plots depicting the minimum difference for all three time slices indicate a decrease in mean rainfall as compared to the 20C3M with the exception of the west coast and northeast India. The plots are similar to the projections provided by Moetasim *et al.* [2009] that indicated the complete suppression of the South Asian Summer Monsoon during the 21st century with the exception of the mentioned regions. The plots depicting the maximum and mean difference display a similar spatial variation but higher intensities for the maximum difference. The variations are similar to those provided by Krishna Kumar *et al.* [2011] using dynamic downscaling.

[53] Figure 12 provides an uncertainty assessment in the form of time series and bar charts for rainfall considered over the entire Indian landmass in terms of the spatial mean. The observed rainfall time series indicates high fluctuations as compared to the 20C3M time series for the same period. Both time series show almost the same trend. However, the fluctuations are suppressed in the 20C3M time series on account of the kernel regression that represents the conditional expectation for the predictand for given predictor values. The fluctuations remained suppressed even for the future time series obtained from the projected annual mean for entire landmass of India during the 21st century. The

lack of a significant trend for mean projected rainfall, averaged over India for the 20C3M scenario, is an important finding that matches the results obtained by Goswami *et al.* [2006]. In terms of the CO₂ concentration at the end of the 21st century, the SRESA2 (A2) is the most stringent scenario, the SRESA1B (A1B) is a medium stringent scenario, and the SRESB1 (B1) is the least stringent scenario. Time series plots for these scenarios revealed the same in terms of the annual mean rainfall time series for all of India. As compared to the A1B and B1, especially in the 2080s, the mean time series for the A2 scenario displayed more of an increasing trend. Even though a small upward trend is determined, the trend is not actually significant, indicating that for all of India, the annual mean rainfall will not differ much in the future but that the spatial distribution of mean rainfall will change, as evident from the plots in Figure 12. The result is an important finding and indicates that, in the future, the total amount of rainfall will not vary significantly but that the spatial pattern will be modified in terms of the rainfall magnitude at grid level.

4.3. The Orography Captured by the Data and the Downscaling Technique

[54] Indian summer monsoon rainfall is basically the result of the interaction between the moisture carried by

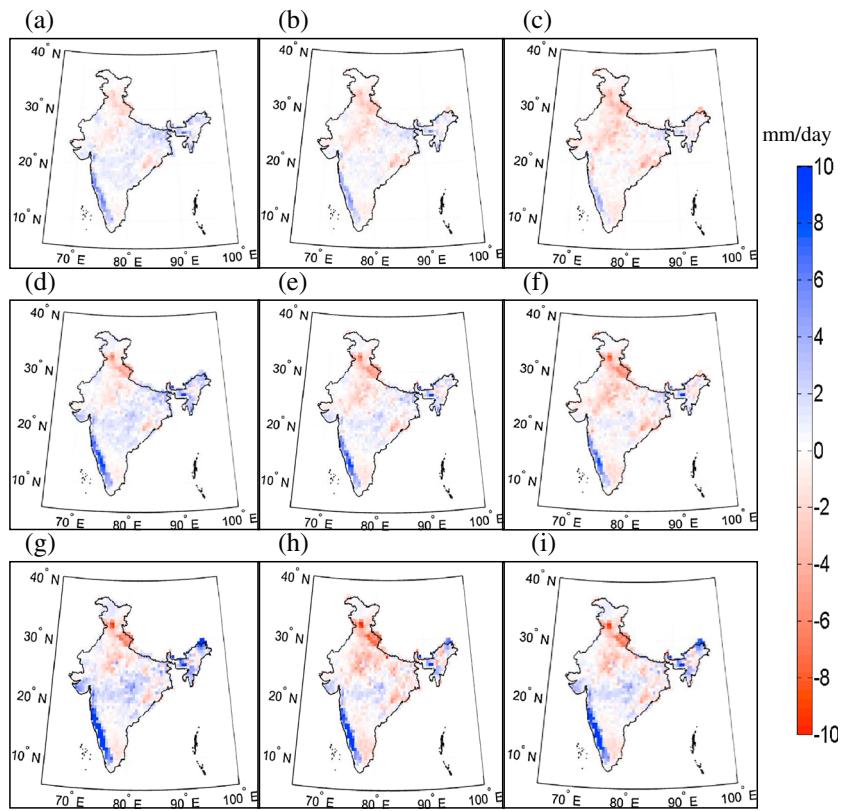


Figure 11. Lower bound and mean and upper bound of changes in rainfall for A2 scenarios, representing the uncertainty across the runs. The left panel shows upper bounds of changes for (a) 2020s, (d) 2050s, and (g) 2080s. Similarly, the (b, e, and h) mean and (c, f, and i) lower bounds are presented in the same order.

winds and the orography. The Indian landmass consists of hilly regions that, to a great extent, control rainfall. Figure 13a provides a physical map of India. The Western Ghats, Satpura, and Vindhya Ranges are important hilly regions playing a major role in the rainfall over west coast and central India. Figure 13b displays high amounts of rainfall for nodes that are on the windward side of these hills and low rainfall on the leeward side, indicating that the model understands the physics behind orographic rainfall and captures it through trends in observed data. The Asian monsoon has two branches: the Arabian Sea branch and the Bay of Bengal branch. The Arabian Sea branch monsoon enters India in Kerala, and wind moves in the southeast direction. The first encounter of monsoon winds in the Western Ghats range produces heavy rainfall in the west coast region. In the future, increases in mean rainfall will be greater on the western coast. The leeward side of the Western Ghats receives much less rainfall. Therefore, some parts of the Maharashtra, such as Vidarbha and Marathwada, and some parts of Karnataka that are located on the leeward side of the Western Ghats receive less rainfall. The central part of India on the windward side of the Satpura and Vindhya Ranges also receives more rainfall, mainly in Madhya Pradesh. The Aravali Range fails to perform the same job mainly due to alignment. The ranges are almost parallel to the direction of the monsoon wind and, therefore, offer the least resistance, leading to less rainfall near these ranges and in the Rajasthan. The region of India located in the northern part of the Aravali receives a benefit due to its

alignment and receives some rainfall. As the monsoon moves forward, the majority of the moisture present in the wind gets exhausted as a result of the rainfall at various locations. Therefore, the extreme north of India receives little rainfall. The northeast region of India displays very heavy rainfall, mainly as a result of the Bay of Bengal branch of AIMR. Although topography does not provide an input for the employed downscaling model directly, as evident from Figure 13b, all of the rainfall patterns and the nonuniformity displayed by the monsoon for the Indian landmass is well replicated in the projected rainfall. Figure 13c provides the changes in the mean projected rainfall for the SRESA2 scenario as compared to the 20C3M. The figure exemplifies an interesting relationship between the change in mean rainfall (positive or negative) and orography. Clearly seen is that regions with high mean rainfall exist on the windward side of the hilly regions. For example, the west coast is on the windward side of the Western Ghats (as shown by black triangles), and part of central India exists on the windward side of the Satpura Range (as shown by pink triangles) and in the northeast region on the windward side of the Himalayas. In the future, changes in mean rainfall are positive in these regions. In a similar manner, in regions with a comparatively lower mean rainfall (the east coast and northern India), the changes are negative. Global warming [Goswami *et al.*, 2006] will lead to higher evaporation rate, resulting in an increase in the moisture content of the monsoon wind, a possible reason for the increase in orographic precipitation on the windward side of the hilly regions. The finding led us

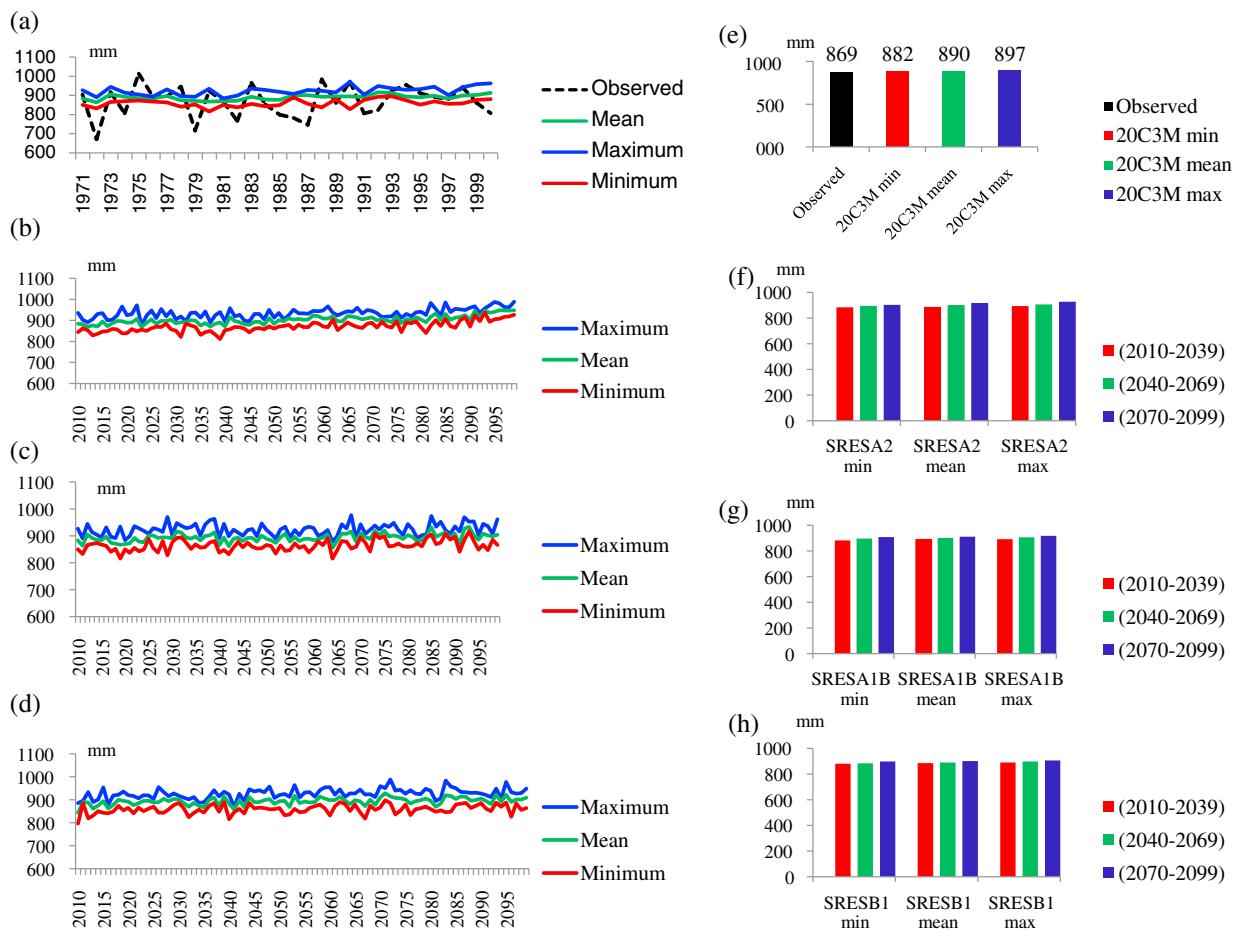


Figure 12. Uncertainty assessment in the form of annual mean time series of projected rainfall and bar chart comparison for the entire India. (a) Comparison of observed annual mean rainfall time series with projected annual mean rainfall time series (minimum, mean, and maximum) for the period 1971–2000. (b) Maximum, minimum, and mean annual time series for projected rainfall for the period 2010–2099 for SRESA2 scenario. (c) Maximum, minimum, and mean annual time series for projected rainfall for the period 2010–2099 for SRESA1B scenario. (d) Maximum, minimum, and mean annual time series for projected rainfall for the period 2010–2099 for SRESB2 scenario. (e) Comparison of mean rainfall for observed data and projected data along with the uncertainty in the form of minimum rainfall amount and maximum rainfall amount for 20C3M scenario. (f) Bar plot comparison of annual average of mean rainfall, maximum rainfall, and minimum rainfall for different time slices for SRESA2 scenario. (g) Bar plot comparison of annual average of mean rainfall, maximum rainfall, and minimum rainfall for different time slices for SRESA1B scenario. (h) Bar plot comparison of annual average of mean rainfall, maximum rainfall, and minimum rainfall for different time slices for SRESB1 scenario.

to the hypothesis that “wet is getting wetter and dry is getting drier.” However, a careful test of this hypothesis using multiple climate models coupled with a proper land surface model and hydrology may be considered for future research.

5. Summary and Concluding Remarks

[55] For the present study, we projected rainfall for the entire Indian landmass at a high resolution (0.5° latitude \times 0.5° longitude) for 20th and 21st century scenarios as determined by the IPCC. The task is achieved by applying the statistical downscaling technique presented by *Kannan and Ghosh* [2013] to the CCCma (CGCM3.1) GCM-simulated climatic variables as predictors and rainfall as a predictand. For all of the projections, we selected the same surface-level predictors

mentioned by *Kannan and Ghosh* [2013]. To compare the performance of high-level variables with those of surface-level variables for rainfall projections, we selected two additional high-level variables, u-wind and v-wind, in the form of wind shear along with the five surface-level variables and performed a single simulation of rainfall projections. Even after taking into consideration high altitude variables, no significant improvements are determined in the results. The predictors are bias-corrected using the quantile-based remapping method of *Li et al.* [2010]. To remove the possible correlation between the predictors and to reduce dimensionality, the PCA technique is employed. To capture cross correlation among the rainfall data (gridded data), the K-means clustering technique is adopted for generating the daily rainfall states for each zone. A CART analysis is employed that

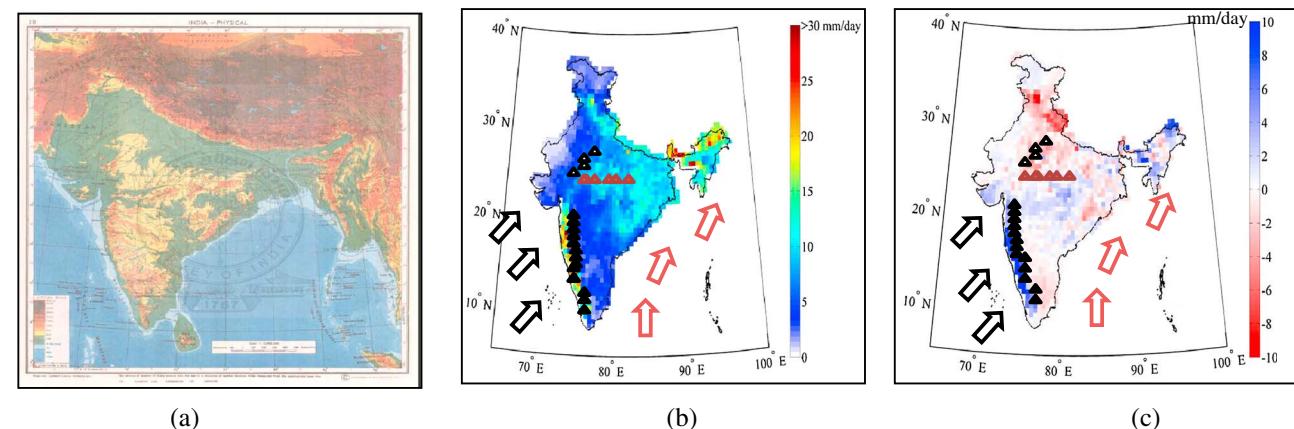


Figure 13. Relationship between orography and mean rainfall for India. The statistical downscaling technique seems to capture the orography quite well. There is clear distinction between the rainfall at windward side and leeward side for Western Ghats, Satpura range, and hilly region at northeast India. (a) Physical map of India (source: Survey of India). (b) Projected mean rainfall of India (20C3M). Black arrows show direction of SASM (Arabian sea branch). Black filled triangles show approximate alignment of Western Ghats, pink filled triangles show the approximate alignment of Satpura range, and black empty triangles show approximate alignment of Aravali range. Pink arrows show direction of SASM (Bay of Bengal branch). (c) The orography is linked to the changes in the mean rainfall for future.

utilized training period rainfall states and principal components in order to project future rainfall states. A nonparametric kernel regression is utilized to obtain the projected daily rainfall at each node. Rainfall projection results obtained for the 20th century for the 20C3M scenario provide a good match to observed data in terms of statistical properties (i.e., mean and standard deviation). Multisite downscaling techniques often fail to capture the cross correlation among rain gauge stations. The cross-correlation plot displays the strength of the model in capturing the influence of rainfall at nearby nodes on the node at which the rainfall is projected. The high-level variables used as predictors display improvement for the computation time by reducing the number of principal components for the same degree of variability. However, they do not show significant changes in accuracy as compared with the mean rainfall projected using surface-level variables.

[56] Future projections are performed for the 21st century for three scenarios A1B, B1, and A2. As compared to the 20C3M, the results indicated spatial nonuniformity for changes in mean rainfall. For some of the nodes, there is an increase in mean rainfall, while for other nodes, there is a decrease in mean rainfall. The magnitude of the change for mean rainfall (decrease or increase) indicates an increase with time (i.e., the change in mean rainfall varied as we move along the time slices from the 2020s to the 2080s). Also, the change increases with a change in the scenarios. The A2 scenario as the worst case indicated a maximum change. Afterward, the A1B and B1 scenarios are queued. However, for all three scenarios, the pattern of the change appears to remain the same. An increase in rainfall is projected for the western coast, some parts of central India, and the northeast part of India that actually receive higher rainfall as compared to other parts of India, whereas the northern part of India and the southeast coast display a decrease in mean rainfall that receive less rainfall. Consistency in the projection patterns may lead to a hypothesis of “wet

areas getting wetter and dry areas getting drier,” which needs to be validated and this will be taken care as a future scope of the present work. The influence of orography is nicely captured by the model and is clearly evident from the projected rainfall pattern that shows high rainfall for the windward side of the Western Ghat and the Satpura and lower rainfall for the leeward side. The presence of multiple simulations enabled us to perform an uncertainty assessment. On one hand, the uncertainty assessment indicates complete suppression of the monsoon in southern central India for the future as compared to the 20C3M, similar to the projection results by *Moetasim et al.* [2009]. On the other hand, it provides a nonuniform pattern (an overall increase) similar to the results obtained by *Krishna Kumar et al.* [2011]. Time series plots for the 20th century show that the model is not able to capture the variability shown by the observed rainfall time series. However, the mean is well captured. The projected rainfall time series for the future does not display any significant trend, indicating that, for the future, no major change in rainfall, as far as the magnitude is concerned, will occur, but that the spatial distribution will change. Even if the focus and the outcome of the current study is the high-resolution future rainfall projection obtained by applying the statistical downscaling technique proposed by *Kannan and Ghosh* [2013], for research purposes, the limitations of the overall approach and the future scope must be understood. The equidistant CDF matching technique used for the bias correction of the GCM-simulated variables assumes a time-invariant bias [*Li et al.*, 2010]. However, if the behavior of the bias is changed in the future as a result of changes in circulation patterns, the results will no longer be valid. The PCA technique that is used for dimensional reduction is more effective when linear relationships exist among the variables. Since nonlinear relationships are common among the climate variables, PCA may not produce good results. Multiple nonlinear dimensional reduction techniques such as locally linear embedding [*Roweis and Saul*, 2000]

can be adopted to take care of nonlinearities. Also, principal components, neglected for CART analysis on the basis of low variability explanations, may actually influence rainfall to a greater extent and must be taken into account while establishing statistical relationships using supervised dimensionality reduction techniques such as a local Fisher discriminant analysis [Sugiyama et al., 2010]. The methodology adopted for our study involved the establishment of a statistical relationship between the predictors and the predictand using observed period data that is assumed to be stationary for the time period when the projections are obtained. The hypothesis of stationarity for the established statistical relationship over a period of 100 years must be tested in future work. As a result of the application of a nonparametric kernel regression technique that represents conditional expectations, the model did not fare well in capturing extreme events. Hence, the mean conditions are well captured, but extreme conditions are not well simulated. The methodology must be modified for extreme events. Various approaches can be adopted for downscale extremes [Mannhardt-Shamseldin et al., 2010]. Another reason for the lack of extreme event prediction may be the use of principal components, essentially an unsupervised dimensionality reduction method, that neglected some of the components representing a small percentage of the original variability but that may be key elements for explaining extremes. A detailed study is required in terms of detailed statistical hypothesis testing and the use of supervised dimensionality reduction methods. The use of high altitude predictors may not improve the mean performance of downscaling models. However, they may have impacts on extremes and may improve extreme downscaling performance, which must be studied in detail. For the current study, all of the projections are obtained using GCM-simulated data published in the third phase of the Coupled Model Intercomparison Project (CMIP3). The experimental results for the fifth phase of the Coupled Model Intercomparison Project (CMIP5) are now published. The CMIP5 experiments indicated advantages in terms of assessing the mechanisms responsible for model differences in poorly understood feedbacks associated with the carbon cycle and clouds by exploring the ability of models in predicting climate on decadal time scales. The CMIP5 data consisted of more climatic variables projected on a much longer time scale (up to 2300 years) and a greater number of simulations. The same analysis is performed using CMIP5 data, and the projection results can be compared for the future.

[57] The overall contribution of this work is a high-resolution rainfall projection for the entire 21st century. For the first time, a rainfall projection is performed for all of India at a 0.5° resolution using a statistical downscaling technique. High accuracy in the projection of mean rainfall, replicating the spatial pattern of both the mean and standard deviation, capturing the cross correlation, and an understanding of orography are some of the key strengths of the adopted methodology. Rainfall projections for the entire 21st century can be used for calculating average water availability at the regional level. The use of a water availability scenario, with possible water demands, per population projections, will provide regional water storage requirements. Additionally, the present study may provide information for planning interbasin water transfer. The other potential application of the present analysis

is the use of projections for agricultural water management and hydraulic structure design.

[58] The model failed to simulate extremes. However, this result may be considered as a common limitation of any transfer function-based statistical downscaling model [Wilby et al., 2004]. With recent statistical developments [Benestad, 2010; Mannhardt-Shamseldin et al., 2010], future applications based on downscaling techniques for extremes may yield better products for understanding and simulating heavy rainfall events. This may be considered as a potential future research area.

[59] **Acknowledgments.** We sincerely thank anonymous reviewers for reviewing the manuscript and providing critical comments to improve the quality of our work. Our work is partially supported by the Space Technology Cell, Indian Institute of Technology, Bombay, the Indian Space Research Organization and APN through project no. ARCP2012-13NMV-DeCosta.

References

- Anandhi, A., V. V. Srinivas, R. S. Nanjundiah, and D. N. Kumar (2008), Downscaling precipitation to river basin for IPCC SRES scenarios using support vector machines, *Int. J. Climatol.*, 28(3), 401–420.10.1002/joc.1529.
- Asselin, R. (1972), Frequency filter for time integrations, *Mon. Weather Rev.*, 100, 487–490.
- Bardossy, A., Plate, E. (1991), Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence, *J. Hydrol.* 122(1–4), 33–47.
- Benestad, R. (2010), Downscaling precipitation extremes, *Theor. Appl. Climatol.*, 100, 1–21, doi:10.1007/s00704-009-0158-1.
- Boer, G. J., N. A. McFarlane, R. Laprise, J. D. Henderson, and J. P. Blanchet (1984), The Canadian Climate Centre spectral atmospheric general circulation model, *Atmos. Ocean*, 22(4), 397–429.
- Charles, S. P., B. C. Bates, and J. P. Hughes (1999), A spatio-temporal model for downscaling precipitation occurrence and amounts, *J. Geophys. Res.*, 104(D24), 31657–31669.
- Dhar, O. N., and S. Nandargi (2003), Hydrometeorological aspects of floods in India, *Nat. Hazards*, 28, 1–33.
- Dobler, A., and B. Ahrens (2011), Four climate change scenarios for the Indian summer monsoon by the regional climate model COSMO-CLM, *J. Geophys. Res.*, 116(D24), 1–13. doi:10.1029/2011JD016329.
- Druryan, L. M., M. Fulakeza, and P. Lonergan (2002), Dynamic downscaling of seasonal climate predictions over Brazil, *J. Climate*, 15 (23), 3411–3426.
- Efroimovich, S. (1999), *Nonparametric Curve Estimation: Methods, Theory and Applications*, Springer, New York.
- Fowler, H. J., M. Ekstrom, and C. G. Kilsby (2005), New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Assessment of control climate, *J. Hydrol.*, 300(1–4), 212–233.
- Ghosh, S., and Mujumdar, P. P. (2006), Future rainfall scenario over Orissa with GCM projections by statistical downscaling, *Curr. Sci. India*, 90(3), 396–404.
- Ghosh, S. (2010), SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output, *J. Geophys. Res.*, 115, D22102, doi:10.1029/2009JD013548.
- Goswami, B. N., V. Venugopal, D. Sengupta, M. S. Madhusoodanan, and P. K. Xavier (2006), Increasing trend of extreme rain events over India in a warming environment, *Science*, 314, 1442, doi:10.1126/science.1132027.
- Groppelli, B., D. Bocchiola, and R. Rosso (2011), Spatial downscaling of precipitation from GCMs for climate change projections using random cascades: a case study in Italy, *Water Resour. Res.*, doi:10.1029/2010WR009437.
- Hewitson, B. C. and R. G. Crane (1992), Large-scale atmospheric controls on local precipitation in tropical Mexico, *Geophys. Res. Lett.*, 19(18), 1835–1838.
- Hewitson, B. C., and R. G. Crane (1996), Climate downscaling: techniques and application, *Clim. Res.*, 7, 85–95.
- Hughes, J. P., D. P. Lettenmaier, and P. Guttorp (1993), A stochastic approach for assessing the effect of changes in synoptic circulation patterns on gauge precipitation, *Water Resour. Res.*, 29(10), 3303–3315.
- Johnson, F., and A. Sharma (2009), Measurements of GCM skill in predicting variables relevant for hydroclimatological assessments, *J. Clim.*, 22, 4373–4382, doi:10.1175/2009JCLI2681.1.
- Johnson, F., and A. Sharma (2012), A nesting model for bias correction of variability at multiple time scales in general circulation model

- precipitation simulations, *Water Resour. Res.*, 48, W01504, doi:10.1029/2011WR010464, 2012.
- Juang, H.-M. H., and M. Kanamitsu (1994), The NMC nested regional spectral model, *Mon. Weather Rev.*, 122, 3–26.
- Kalnay, E., M. Kanamitsu, R. Kistler et al. (1996), The NCEP/NCAR 40-years reanalysis project, *Bull. Am. Meteor. Soc.*, 77(3), 437–471.
- Kannan, S., and S. Ghosh (2011), Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output, *Stochastic Environ. Res. Risk Assess.*, doi:10.1007/s00477-010-0415-y.
- Kannan, S., and S. Ghosh (2013), A nonparametric Kernel regression model for downscaling multisite daily precipitation in the Mahanadi basin, *Water Resour. Res.*, 49, doi:10.1002/wrcr.20118.
- Kim, J., N. L. Miller, and J. D. Farrara (2000), A seasonal precipitation and stream flow hindcast and prediction study in the western United States during the 1997/98 winter season using a dynamic downscaling system, *J. Hydrometeorol.*, 1(4), 311–329.
- Kripalani, R. H., J. H. Oh, and H. S. Chaudhari (2007), Response of the East Asian summer monsoon to doubled atmospheric CO₂: coupled climate model simulations and projections under IPCC AR4, *Theor. Appl. Climatol.*, 87, 1–28.
- Krishna Kumar, K., S. K. Patwardhan, A. Kulkarni, K. Kamala, K. K. Rao, and R. Jones (2011), Simulated projections for summer monsoon climate over India by a high-resolution regional climate model (PRECIS), *Curr. Sci.*, 101(3).
- Kundu, P. K., and R. K. Siddani (2011), Scale dependence of spatiotemporal intermittence of rain, *Water Resour. Res.*, 47(8), 1–17, doi:10.1029/2010WR010070.
- Lall, U., and A. Sharma (1996), A nearest neighbor boots tarp for time series resampling, *Water Resour. Res.*, 32, 679–693.
- Li, H., J. Sheffield, and Eric, F. W. (2010), Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching, *J. Geophys. Res.*, doi:10.1029/2009JD012882.
- Mannhardt-Shamseldin E. C., R. L. Smith, S. R. Sain, L. O. Mearns and D. Cooley (2010), Downscaling extremes: a comparison of extreme value distributions in point-source and gridded precipitation data, *Ann. Appl. Stat.*, 4(1), 484–502.
- Mehrotra, R., and A. Sharma (2005), A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences, *J. Geophys. Res.*, 110, D16108, doi:10.1029/2004JD005677.
- Mehrotra, R., and A. Sharma (2006), A nonparametric stochastic downscaling framework for daily rainfall at multiple locations, *J. Geophys. Res.*, 111, D15101.
- Mehrotra, R., and A. Sharma (2010), Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment, *Water Resour. Res.*, 46, W07526, doi:10.1029/2009WR008423.
- Mehrotra, R., A. Sharma, and I. Cordery (2004), Comparison of two approaches for downscaling synoptic atmospheric pattern to multisite precipitation occurrence, *J. Geophys. Res.*, 109, D14107, doi:10.1029/2004JD004823.
- Misra, V., P. A. Dirmeyer, and B. P. Kirtman (2003), Dynamic downscaling of seasonal simulations over South America, *J. Climate*, 16(1), 103–117.
- Moetasim, A., Y. Shi, W.-w. Tung, R. J. Trapp, X. Gao, J. S. Pal, and N. S. Diffenbaugh (2009), Suppression of south Asian summer monsoon precipitation in the 21st century, *Geophys. Res. Lett.*, 36(1), doi:10.1029/2008GL036500.
- Nadaraya, E. A. (1964), On estimating regression, *Theor. Probab. Appl.* 10, 186–190.
- Parthasarathy, B., K. Rupakumar, and A. A. Munot (1996), Homogeneous regional summer monsoon rainfall over India: interannual variability and teleconnections, Research report no. RR-070.
- Raje, D., and P. P. Mujumdar (2009), A conditional random field-based downscaling method for assessment of climate change impact on multisite daily precipitation in the Mahanadi basin, *Water Resour. Res.*, 45, W10404, doi:10.1029/2008WR007487.
- Rajeevan, M., J. Bhate (2008), A high resolution daily gridded rainfall data set (1971–2005) for mesoscale meteorological studies, NCC Research Report, India Meteorological Department, Pune, India.
- Robert, A. J., J. Henderson, and C. Turnbull (1972), An implicit time integration scheme for baroclinic modes of the atmosphere, *Mon. Weather Rev.*, 100, 329–335.
- Roweis, S. T., and L. K. Saul (2000), Nonlinear dimensionality reduction by locally linear embedding, *Science*, 290(5500), 2323–2326, doi:10.1126/science.290.5500.2323.
- Rupa Kumar, K. R., A. K. Sahai, K. K. Kumar, S. K. Patwardhan, P. K. Mishra, J. V. Revadekar, and K. Kamala (2006), High-resolution climate change scenarios for India for the 21st century, *Curr. Sci.*, 90(3).
- Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons, New York.
- Sharma, A. (2000), Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: part 1 – a strategy for system predictor identification, *J. Hydrol.*, 239, 232–239.
- Shepard, D. (1968), A two-dimensional interpolation function for irregularly spaced data, Proc. 1968 ACM Nat. Conf., 517–524.
- Sugiyama, M., T. Ide, S. Nakajima, and J. Sese (2010), Semi-supervised local Fisher discriminant analysis for dimensionality reduction, *Mach. Learn.* 78, 35–61, doi:10.1007/s10994-009-5125-7.
- Wand, M. P., and M. C. Jones (1995), *Kernel Smoothing*, Chapman and Hall, London.
- Wang, Q. J., and Robertson, D. E. (2011), Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences, *Water Resour. Res.*, 47(2), 1–19, doi:10.1029/2010WR009333.
- Wang, Q. J., Robertson, D. E., and Chiew, F. H. S. (2009), A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, *Water Resour. Res.*, 45, W05407, doi:10.1029/2008WR007355.
- Wilby, R. L., Hey, L. E., and Leavesly, G. H. (1999), A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River Basin, Colorado, *J. Hydrol.*, 225, 67–91.
- Wilby, R. L., S. P. Charles, E. Zorita, et al. (2004), The guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change (IPCC), prepared on behalf of Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGICA). [Available at <http://ipccddc.cru.uea.ac.uk/guidelines/StatDownGuide.pdf>.]
- Wilks, D. S. (1999), Multisite downscaling of daily precipitation with a stochastic weather generator, *Clim. Res.*, 11, 125–136.
- Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek (2003), A technique for generating regional climate scenarios using a nearest neighbor algorithm, *Water Resour. Res.*, 39(7), 1199, doi:10.1029/2002WR001769.