



Assessment and evaluation of potential climate change impact on monsoon flows using machine learning technique over Wainganga River basin, India

Jew Das & Umamahesh V. Nanduri

To cite this article: Jew Das & Umamahesh V. Nanduri (2018) Assessment and evaluation of potential climate change impact on monsoon flows using machine learning technique over Wainganga River basin, India, Hydrological Sciences Journal, 63:7, 1020-1046, DOI: [10.1080/02626667.2018.1469757](https://doi.org/10.1080/02626667.2018.1469757)

To link to this article: <https://doi.org/10.1080/02626667.2018.1469757>



Published online: 30 May 2018.



Submit your article to this journal



Article views: 1305



View related articles



View Crossmark data



Citing articles: 12 View citing articles



Assessment and evaluation of potential climate change impact on monsoon flows using machine learning technique over Wainganga River basin, India

Jew Das and Umamahesh V. Nanduri

Department of Civil Engineering, National Institute of Technology Warangal, Telangana, India

ABSTRACT

In this study, classification- and regression-based statistical downscaling is used to project the monthly monsoon streamflow over the Wainganga basin, India, using 40 global climate model (GCM) outputs and four representative concentration pathways (RCP) scenarios. Support vector machine (SVM) and relevance vector machine (RVM) are considered to perform downscaling. The RVM outperforms SVM and is used to simulate future projections of monsoon flows for different periods. In addition, variability in water availability with uncertainty and change point (CP) detection are accomplished by flow-duration curve and Bayesian analysis, respectively. It is observed from the results that the upper extremes of monsoon flows are highly sensitive to increases in temperature and show a continuous decreasing trend. Medium and low flows are increasing in future projections for all the scenarios, and high uncertainty is noticed in the case of low flows. An early CP is detected in the case of high emissions scenarios.

ARTICLE HISTORY

Received 2 July 2017
Accepted 8 March 2018

EDITOR

R. Woods

ASSOCIATE EDITOR

S. Huang

KEYWORDS

Monsoon streamflow;
machine learning; Bayesian
analysis; climate change;
Wainganga; India

1 Introduction

The changing global climate alters the hydrological cycle, which in response is causing variability in the frequency of the extreme events, availability of water, irrigation water use, and quality of freshwater resources (Simonovic 2017). Thus, the varying nature of the climate, due to the perturbations induced by human activities, draws significant attention in water resources and hydrology. To assess the induced impact in the future for better risk and resources management, global climate models (GCMs) are used as the most credible tools (Ghosh and Mujumdar 2008, Goharian *et al.* 2016). GCMs simulate the present and future climate under different climate scenarios by incorporating the changes in the atmospheric forcings on the horizontal (2° – 4° grid resolution) and for 10–20 layers in the vertical direction (Dibike and Coulibaly 2005). The climate variables are well simulated at coarser resolution, i.e. at continental and hemispherical scales; however, impact analysis at finer grid size, i.e. at the regional scale, requires variables at the sub-grid scale. Hence, a method called downscaling is used to bring GCM variables down from coarser resolution to finer resolution, as hydrological importance increases at a local scale but at the same time climate model ability decreases from global to local scale (Xu 1999). This also helps to generate variables that cannot be obtained

directly or are not well predicted by the GCMs. The downscaling techniques are classified as statistical and dynamic (Chu *et al.* 2010); statistical downscaling has gained popularity in hydrological studies because of its shorter computational time (Chen *et al.* 2010). The statistical downscaling technique involves processing of large-scale GCM variables (predictors), establishing a relationship between the historical predictors and hydrological variables (predictand), forecasting for future time steps, and finally evaluating the consequences relative to the present climate (Das and Umamahesh 2016).

Generally, impact studies in hydrology involve a two-step process also known as indirect downscaling (Joshi *et al.* 2013, 2016). Initially, the meteorological variables (e.g. precipitation, temperature) are down-scaled to the regional climate conditions from the large-scale GCM outputs. Next, the downscaled variables are used as inputs to the hydrological model to assess the hydrological changes over the region due to the future climate (Okkan and Fistikoglu 2014, Zhou *et al.* 2015, Kuo *et al.* 2017). However, the two-step process of impact studies is usually constrained by the existing gap between a GCM's ability and spatial resolution (Joshi *et al.* 2016), and also by the domain of calibration of the hydrological model (Tisseuil *et al.* 2010) along with the data requirements to fix the

model parameters (Eckhardt *et al.* 2005). Therefore, direct downscaling is carried out in the present study, as large ensembles of future projection can be obtained at low computational cost. Statistical downscaling and dynamic downscaling are used as direct downscaling methods to evaluate the possible changes in the hydrological variables incorporating future climate conditions. Dynamic downscaling uses complex algorithms to represent the atmospheric processes nested within the GCM outputs. However, statistical downscaling projects future scenarios based on the statistical relationship between large-scale atmospheric features and hydrological variables (Das and Umamahesh 2015). Therefore, due to the simplicity in computation, statistical downscaling is used in the present study. With the assumptions that the statistical relationships hold good in the future for changed climate scenarios, the selected predictors completely represent the changing climate signal, and the spatial variability is unchanged under different forcings (Diaz-Nieto and Wilby 2005, Ghosh and Mujumdar 2008), statistical downscaling is used extensively in hydrological impact analysis. In recent studies, statistical downscaling has been used to downscale precipitation (Raje and Mujumdar 2011, Hashmi *et al.* 2013, Okkan and Inan 2015a, Okkan and Kirdemir 2016, Singh and Goyal 2016), temperature (Goyal *et al.* 2012, Duhan and Pandey 2015, Pang *et al.* 2017) and evaporation (Yang *et al.* 2012, Sachindra *et al.* 2015, Sachindra and Perera 2016). In addition, a few studies have investigated the variation in streamflow patterns due to climate change using direct downscaling techniques. Examples include Landman *et al.* (2001), who analysed seasonal variations in the streamflow at the inlets of 12 dams in South Africa using canonical correlation analysis (CCA) and sea surface temperature fields as predictors; Cannon and Whitfield (2002), who used an ensemble neural network model to downscale streamflow in British Columbia, Canada; Ghosh and Mujumdar (2008) used machine learning techniques to downscale monsoon streamflow over the Mahanadi basin, India, and observed a decreasing trend for monsoon streamflow. Similar to the above studies, there are many recent studies that show applications of direct downscaling to streamflow. These include, but are not limited to, Tisseuil *et al.* (2010), Sachindra *et al.* (2013) and Okkan and Inan (2015b).

In this sense, different statistical downscaling techniques, such as linear regression, nonlinear regression or artificial neural network (ANN), are generally used to build the predictor-predictand relationship. However, downscaling techniques based on ANN have increased acknowledgment due to their ability to model the

nonlinear association between predictors and predictand (Wilby *et al.* 1998, Trigo and Palutikof 1999). However, regardless of various points of interest, ANN has major limitations, for instance being confined at local minima and the subjective nature of the model architecture (Suykens 2001). Vapnik (1995) proposed a novel machine learning algorithm known as support vector machines (SVMs) to strengthen the weakness of ANN, and this method has also been successfully applied on climate change impact analysis (Tripathi *et al.* 2006, Anandhi *et al.* 2008, Chen *et al.* 2010). Moreover, Srivastava *et al.* (2013) adopted this machine learning technique to downscale soil moisture using land surface temperature for hydrological applications and Chen *et al.* (2012) advocated that machine learning has better efficiency than the statistical downscaling model when examining the impact on runoff due to climate change. Regardless of the prevalent adaptation of the SVMs, this methodology additionally has a few downsides: for example, monotonically increasing support vectors with the training dataset, nonprobabilistic interpretation of the output, and no direct techniques to estimate the parameters (Ghosh and Mujumdar 2008). To overcome the limitations of SVM, Tipping (2001) proposed a relevance vector machine (RVM) using a probabilistic Bayesian learning framework, which uses fewer relevance vectors as compared to the SVM and can predict accurately with fewer parameters.

In addition to the future climate projections and downscaling techniques, various physical phenomena should be taken into consideration for better understanding of hydrological processes and selection of the predictors. The variability and frequency of the extreme hydrological events are aggravated by dynamic land use, deforestation and unsustainable farming practices, and this changeability will cause modifications in the long-term average climatic conditions (Githui *et al.* 2009). Moreover, Chiew and McMahon (2002) stated that variability of precipitation would have a direct effect on runoff, whilst at the same time increasing temperature may lead to decreases in runoff and soil moisture levels. These uncertainties will increase the difficulties for water managers in proper design and management practices. In addition, Ashofteh *et al.* (2013) stated that there is justification to evaluate the negative consequences of climate change on runoff from the perspectives of water resource planners and hydrologists. Therefore, it is important to consider water availability with associated uncertainty and the detection of change points in the hydro-meteorological time series to determine the changes in the statistical properties of time series data (Raje 2014).

In the present study, the primary objective is to project the monsoon streamflow over the Wainganga River basin in India using 40 GCMs and four different

representative concentration pathways (RCP2.0, RCP4.5, RCP6.0 and RCP8.5) scenarios (a list of the GCMs is provided in the Appendix, Table A1). The specific objectives of the investigation include: (1) performance testing of the two machine learning techniques (SVM and RVM), carried out based on the observed datasets; (2) use of the better performing model to downscale the streamflow using future projections of GCM outputs under different scenarios; (3) examination of water availability under different climate change scenarios through flow-duration curves (FDC) and evaluation of the associated uncertainty using a nonparametric bootstrapping technique; and (4) to inspect the change point in the future projected mean annual monsoon series using Bayesian change point analysis. A similar kind of approach was adopted by Ghosh and Mujumdar (2008) to downscale monsoon streamflow using only one GCM and one emissions scenario. Moreover, Mujumdar and Ghosh (2008) stated that a single trajectory obtained with a single GCM and a single scenario cannot alone represent future hydrology and may lead to inappropriate planning and management. However, to the knowledge of the authors, the present study is one of the first studies in India to incorporate 40 GCMs with four different scenarios to project streamflow variations. Moreover, the present study also analyses the uncertainty in FDC with change point detection using Bayesian analysis for different scenarios. The general characteristics of the study area and the data type are introduced in Section 2. The detailed methodology is illustrated in Section 3, followed by the results and discussion under Section 4. Finally, the summary and critical findings from the study are explained in Section 5.

2 Study area and data used

Wainganga River basin, with an area of about 49 695.40 km² situated between 19°30'–22°50'N and 78°0'–81°0'E, is the largest sub-basin of the River Godavari, India (Fig. 1). The length of the river is about 635.40 km, and the annual rainfall varies from 900 to 1600 mm, with most of the precipitation occurring during the monsoon season. The study area comes under the tropical climate with minimum temperature varying between 7 and 13°C and maximum temperature fluctuating between 39 and 47°C during summer. Agricultural and forestlands are the predominant land-use classifications over the study area. The decadal land-use classifications (1985, 1995 and 2005) were studied for the Wainganga basin. It is observed that there has been no significant change in land-use pattern over the years (see Appendix, Table A2). Therefore, the variability in streamflow due to dynamic vegetation is assumed to remain constant in the future. Moreover, there are no major flow-regulating structures over the Wainganga basin and hence the flows are unregulated.

The Wainganga is a rain-fed river and most of the rain is received during the monsoon season (June–October), therefore only the monsoon season is considered to study climate change impact. Monthly streamflow data (1971–2000) for the Wainganga basin were collected from the Central Water Commission (CWC), Hyderabad. In the downscaling techniques, selection of predictors plays an important role. Generally, streamflow is considered as resultant from precipitation and evaporation (Ghosh and Mujumdar 2008) and the previous study carried out by Das and Umamahesh (2016) to project the monsoon precipitation over the Godavari basin suggested that

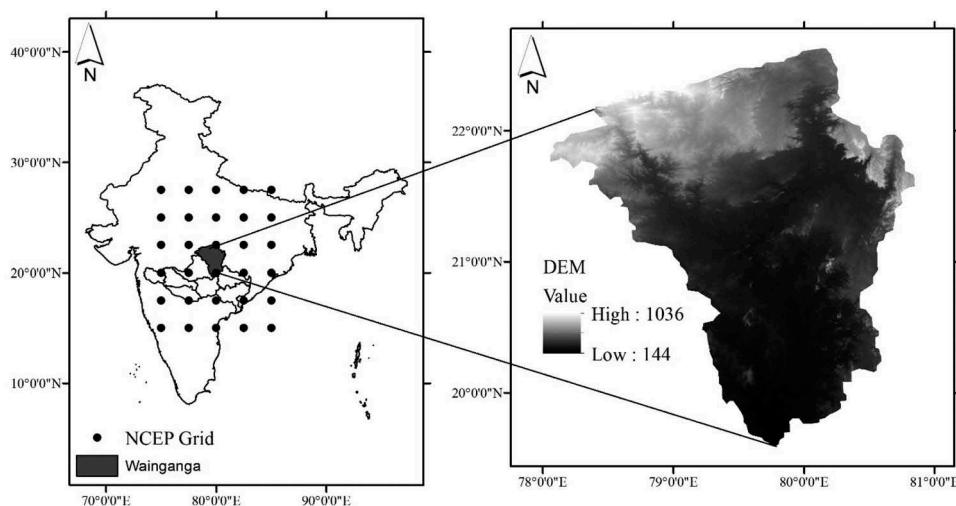


Figure 1. Location map of study area with NCEP grid points.

mean sea-level pressure, specific humidity and 500 hpa geopotential height as predictors were able to capture the monsoon precipitation over the basin. In addition, the evaporation rate is highly correlated with the temperature. Sachindra *et al.* (2013) performed a study to evaluate the best potential predictors to downscale the streamflow for different months. They also suggested that soil moisture content could be one of the potential predictors for June and October in addition to the geopotential height. However, the soil moisture content also depends on the texture of the soil. Larger texture soil infiltrates more and delays runoff generation. The soil classification of the present study area is dominated mainly by clay and clayey loam soil. The high clay content restricts the movement of water through the soil. Therefore, it is assumed that soil moisture has a very less significant role in runoff generation over the present study area. Moreover, the variables with hydrological importance, such as runoff, evapotranspiration and soil moisture, are not well represented by the GCMs (Loaiciga *et al.* 1996). Hence, in the present analysis, mean sea-level pressure, specific humidity, 500 hpa geopotential height, and 2-m surface air temperature are considered as predictors to model the monsoon streamflow. The above-mentioned predictors for the historical period were collected from the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/

NCAR) re-analysis data (<https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html>), and for different climate models and future forcing scenarios were downloaded from the Coupled Model Intercomparison Project 5 (CMIP5) data archive (http://www.ipcc-data.org/sim/gcm_monthly/AR5/Reference-Archive.html). There are 26, 39, 18 and 36 GCM projections under the RCP2.6, RCP4.5, RCP6.0 and RCP8.5 scenarios, respectively. The extent of the NCEP data points covering the study area spanning 15°–27.5°N and 75°–85°E, with grid resolution of 2.5° × 2.5°, is presented in Figure 1.

3 Methodology

An overview of the proposed methodology is framed in the form of a flow chart and presented in Figure 2. This section illustrates the pre-processing of the predictors prior to the downscaling and training as well as testing of different machine learning techniques using various kernel functions. In addition, post-processing of future projected monsoon streamflow is accomplished by the equiprobability transformation. A nonparametric bootstrapping technique is chosen to quantify the uncertainty associated with the FDC, and change point detection in annual monsoon flow is achieved using a Bayesian approach.

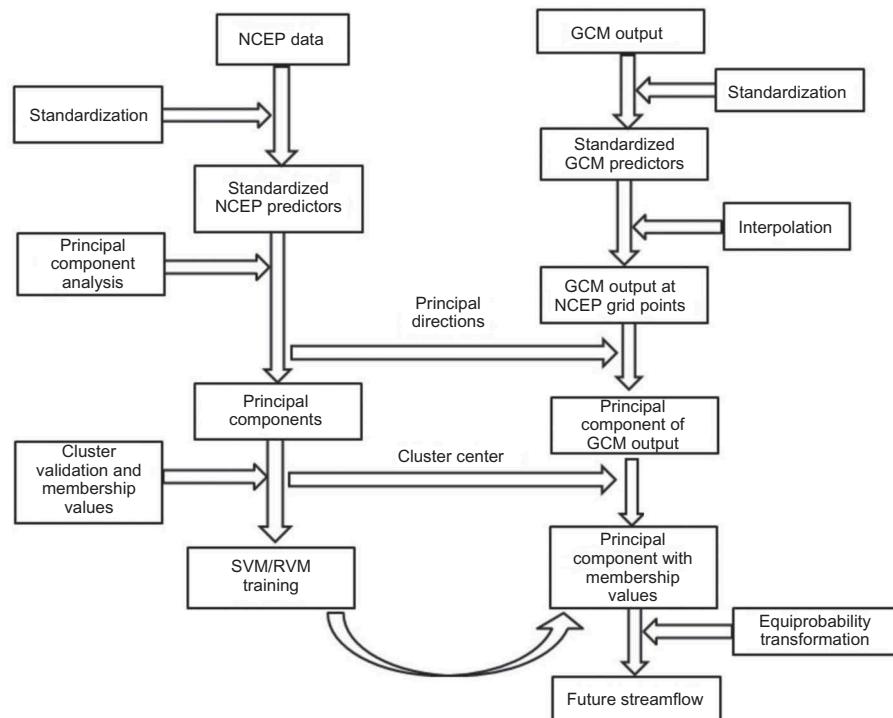


Figure 2. Flowchart of the proposed framework.

3.1 Pre-processing of predictors

Due to inadequate understanding of the fundamental physics, many processes are parameterized in GCM modelling, which leads to a mismatch between the GCM-simulated and observed climate variables (Mujumdar and Nagesh Kumar 2012). The mismatch may cause differences in magnitude of different inputs causing instability in the model. Hence, standardization is used to scale all the data to a uniform scale:

$$P_{\text{std}}(t) = \frac{P(t) - \mu_{P(\text{REF.PERIOD})}}{\sigma_{P(\text{REF.PERIOD})}} \quad (1)$$

where $P_{\text{std}}(t)$ refers to the standardized or bias-corrected predictor at time t , $P(t)$ is the original predictor value, $\mu_{P(\text{REF.PERIOD})}$ is the mean for the particular predictor for the reference/base-line period, and $\sigma_{P(\text{REF.PERIOD})}$ is the standard deviation calculated for the reference period. In the present study, the period 1971–2000 is considered as the base-line period. Since different GCMs have different spatial resolution and there is a disparity between the re-analysis grid points and GCM grid points, bilinear interpolation is carried out to project the GCM outputs at re-analysis grid points.

Due to the effects of overfitting, more computational time, and the existing correlations between the predictors, it is not advisable to perform downscaling methods with a large number of predictors (Mujumdar and Nagesh Kumar 2012). Therefore, to reduce the dimensionality and correlation among the predictors while retaining the variability of the original data, principal component analysis (PCA) is implemented. Principal components (PCs) are computed from the eigenvalues and eigenvectors of the covariance matrix of original data. Standardization before PCA makes all the predictors equally important by creating new variables having mean 0 and variance 1. PCA analysis is carried out for the selected potential predictors over the 30 grid points. It is noted that the first four PCs preserve 96.2% of the variability from the original dataset of the base period. The PCs obtained for the base-line period are divided into calibration and validation sets to develop SVM and RVM models. Then, the coefficient matrix (eigenvectors) obtained from the base-line period is used for deriving PCs from the predictors extracted from historical as well as future standardized GCM outputs (Sachindra *et al.* 2013). It is assumed that the PCs obtained for future scenarios using the coefficient matrix derived from the historical data are uncorrelated. This approach is widely used in the literature to assess the impact of climate change (Tripathi *et al.* 2006, Ghosh and Mujumdar 2008).

3.2 Classification of the PCs

Fuzzy c-means (FCM) clustering is used to classify the PCs into different clusters/classes. Unlike k-means clustering, c-means clustering assigns weighted membership to the data that may belong to more than one cluster. Moreover, c-means clustering provides better results for overlapped data and is comparatively better than k-means clustering. Hence, c-means clustering is used to classify the PCs in this study. In the FCM algorithm (Bezdek 1981), the input datasets are distributed into p -dimensional Euclidean space [$X = (x_1, x_2, \dots, x_n) \subset \Re^p$] in which the centres of the fuzzy clusters/classes are to be estimated. This is achieved by minimizing the objective function:

$$f(C) = \sum_{i=1}^k \sum_{j=1}^n \mu_{ij}^m \|x_j - c_i\|^2 \quad (2)$$

where C is the cluster centre matrix, k is the number of classes or clusters, n is the number of datasets, m is fuzzification parameter, and μ_{ij} is the degree of membership in cluster i of sample j .

The two important parameters of the FCM technique are the number of clusters and the fuzzification parameter, which controls the degree of fuzziness (Ghosh and Mujumdar 2008). The cluster validation problem was addressed by Roubens (1982) and suggested based on the fuzziness performance index (FPI) and normalized classification entropy (NCE). The FPI (Equation (3)) estimates the degree of fuzziness and NCE (Equation (4)) approximates the degree of disorganization produced by different clusters. The optimum numbers of classes and fuzzification parameters are considered based on minimizing FPI and NCE. In addition, Güler and Thyne (2004) suggested $\text{FPI} \cong 0.25$ for selection of cluster numbers and fuzzification parameter.

$$\text{FPI} = 1 - \frac{cF - 1}{c - 1}, \text{ where } F = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^n (\mu_{ij})^2 \quad (3)$$

$$\text{NCE} = \frac{H}{\log c}, \text{ where } H = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^n \mu_{ij} \log(\mu_{ij}) \quad (4)$$

Based on the cluster validation results using different indexes, the obtained optimum number of clusters and fuzzification parameter are 3 and 1.4, respectively. The results obtained from different numbers of clusters and fuzzification parameters are presented in Figure 3. Finally, PCs with membership functions are considered as the input matrix to the SVM and RVM.

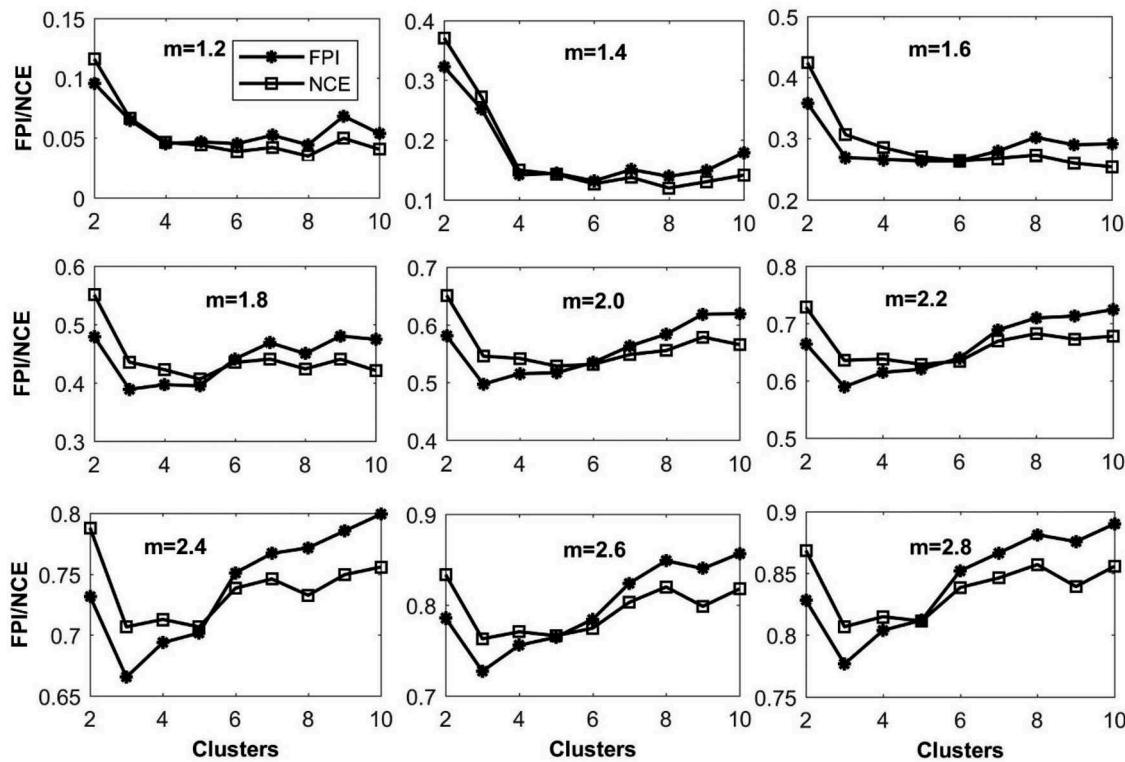


Figure 3. FPI and NCE values for different cluster and fuzzification parameters.

3.3 Support vector machine

The SVM is a data-driven model, generally used for classification and regression operations, developed by Vapnik (1995). It undergoes training and testing processes to classify or predict data and the regression-based support vector machine is generally known as support vector regression (SVR) (Aboutalebi *et al.* 2016). In the training (calibration) process, after achieving the weights and bias according to the training dataset (input and output), SVR tries to approximate the output from the given testing input dataset. Basically, SVR performs two functions where in the first case (training process) it estimates the errors of prediction and secondly it computes the output values based on the weights, bias and input data. The error is estimated using an epsilon-insensitive loss function given by:

$$|y - f(x)| = \begin{cases} 0 & \text{if } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon = \xi & \text{otherwise} \end{cases} \quad (5)$$

in which x is the input vector, y is the output vector, $f(x)$ is output computed by SVR, ε is the sensitivity of prediction error, and ξ is the penalty for the values outside the range $[-\varepsilon$ to $+\varepsilon]$. A schematic diagram of the epsilon-insensitive loss function is presented in Figure 4.

After nonlinear mapping of input data onto a high-dimensional feature space, the linear model is built in this feature space using the following equation:

$$\mathbf{y} = f(\mathbf{x}) = \sum_{i=1}^n w_i \times K(\mathbf{x}_i, \mathbf{x}) + b \quad (6)$$

where w_i represents weights, $K(\mathbf{x}_i, \mathbf{x})$ is the kernel function and b is bias. Linear regression is not suitable in most hydrological applications and therefore nonlinearity is introduced using the kernel function onto the high-dimensional feature space (Ghosh and Mujumdar 2008). The bias and weights are computed by solving an optimization problem with minimum value of loss function given by:

$$\text{minimize} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n (\xi_i^- + \xi_i^+) \quad (7)$$

subject to:

$$y_i - \langle w, x \rangle - b \leq \varepsilon + \xi_i^+ \quad (8)$$

$$\langle w, x \rangle + b - y_i \leq \varepsilon + \xi_i^- \quad (9)$$

$$\xi_i^-, \xi_i^+ \geq 0 \quad (10)$$

where c is the penalty coefficient. The optimization problem can be explained using Lagrange multipliers, and for the detailed methodology readers are advised to follow

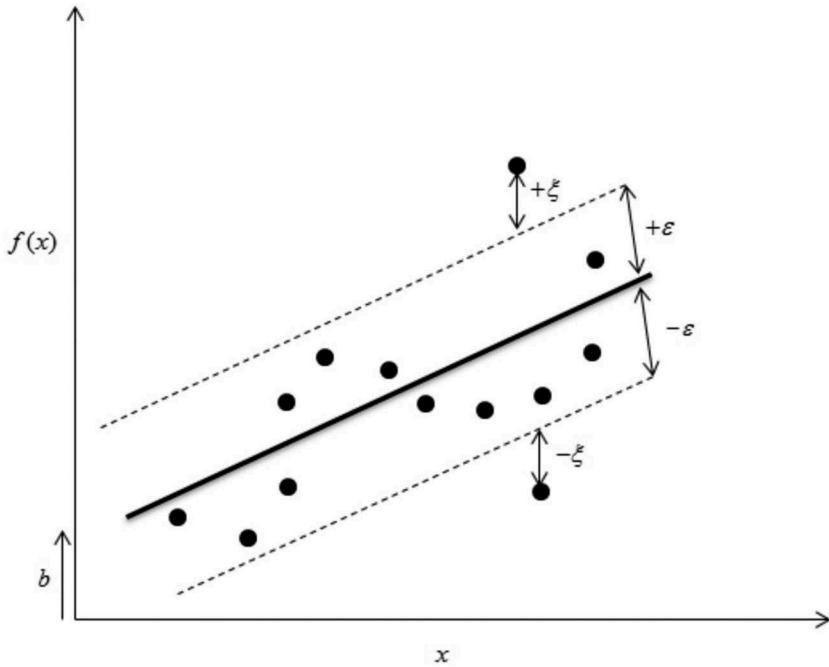


Figure 4. Architecture of epsilon-insensitive function.

Ghosh and Mujumdar (2008) and Aboutalebi *et al.* (2016). The computed values of bias and weights are replaced in Equation (6) to get the output from the SVR. Therefore, C and ϵ are the two important parameters of SVR.

3.4 Relevance vector machine

The relevance vector machine (RVM) adopts a Bayesian approach, which does not have constraints like SVM, as discussed in Section 1. The RVM implements a probabilistic background and implements priors over the model weights controlled by a set of hyperparameters, whose plausible values are estimated iteratively from the data (Tipping 2001). The most convincing aspect of the RVM is that it predicts accurately with fewer relevance vectors and the performance can be comparable with an equivalent SVM. Posterior distributions of the majority of weights peak around zero, and the training vectors associated with the non-zero weights are the “relevance vectors”.

In RVM, with given input (x) to predict output (y) based on $y = f(x) + \epsilon_n$, in which the error term $\epsilon \equiv N(0, \sigma^2)$ is from a Gaussian process and weights are $w = (w_1, w_2, \dots, w_n)^T$. The likelihood function of the dataset can be written as:

$$p(y|w, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\|y - \Phi w\|^2}{2\sigma^2}\right) \quad (11)$$

where $\Phi = [\phi(x_1), \phi(x_2), \dots, \phi(x_n)]^T$ with $\phi(x_i) = [1, K(x_i, x_1), \dots, K(x_i, x_n)]^T$.

Furthermore, Tipping (2001) recommended some prior constraints, w , by imposing a complexity penalty on the likelihood or the error function. Tipping (2001) preferred the popular choice of a zero-mean Gaussian prior distribution over W which can be defined as:

$$p(w|\alpha) = \prod_{i=0}^n N(w_i|0, \alpha_i^{-1}) \quad (12)$$

The hyperparameter α controls the deviation of each weight from zero.

Finally, Bayes' rule is used to obtain the posterior distribution of all unknowns using a non-informative prior distribution (Ghosh and Mujumdar 2008, Okkan and Inan 2015a):

$$p(w, \alpha, \sigma^2 | y) = \frac{p(y|w, \alpha, \sigma^2)p(w, \alpha, \sigma^2)}{\int p(y|w, \alpha, \sigma^2)p(w, \alpha, \sigma^2) \partial w \partial \alpha \partial \sigma^2} \quad (13)$$

The predictive distribution of weight can be defined by:

$$\begin{aligned} p(w|y, \alpha, \sigma^2) &= \frac{p(y|w, \sigma^2)p(w|\alpha)}{p(y|\alpha, \sigma^2)} \\ &= (2\pi)^{-n/2} \left| \sum \right|^{-1/2} \exp \left[-\frac{1}{2} (w - \mu)^T \sum^{-1} (w - \mu) \right] \end{aligned} \quad (14)$$

where the posterior mean and covariance can be represented by:

$$\mu = \sigma^{-2} \sum^{\Phi^T} y \text{ and } \sum = (\sigma^{-2} \Phi^T \Phi + A)^{-1} \quad (15)$$

where $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_n)$.

Having a uniform hyperprior over α and σ^2 , the term $p(y|\alpha, \sigma^2)$ can be maximized and given by:

$$\begin{aligned} p(y|\alpha, \sigma^2) &= \int p(y|w, \sigma^2) p(w|\alpha) dw \\ &= (2\pi)^{-n/2} |\sigma^2 I + \Phi A^{-1} \Phi^T|^{1/2} \\ &\times \exp \left[-\frac{1}{2} y^T (\sigma^2 I + \Phi A^{-1} \Phi^T)^{-1} y \right] \end{aligned} \quad (16)$$

Tipping (2001) stated that the estimation of the hyper-parameter is accomplished with an iterative formula such as a gradient ascent on the objective function. Due to the consequences of this optimization technique, many elements of α will go to infinity, so W will have a few non-zero weights which will be considered as relevant vectors. In this way, the model appreciates the properties of SVM and simultaneously provides estimation uncertainty bounds (Samui and Dixon 2012).

Both SVM and RVM are performed in the R environment using the “e1071” (<https://cran.r-project.org/package=e1071>) and “kernlab” (<https://cran.r-project.org/package=kernlab>) packages. In the present analysis, we consider the Gaussian radial basis function (RBF) (Equation (17)) for SVM and both RBF and Laplacian RBF (Equation (18)) for RVM:

$$k(x, x') = \exp(-\sigma \|x - x'\|^2) \quad (17)$$

$$k(x, x') = \exp(-\sigma \|x - x'\|) \quad (18)$$

3.5 Training and testing of the models

In the present regression-based statistical downscaling, a relationship between the processed predictors (PCs with membership function) and predictand (Wainganga monsoon streamflow) is established using different kernel functions. To evaluate the training and testing performance, a K -fold cross-validation test is used, which divides the training and testing data in to K sets and the performance is evaluated based on average R (correlation coefficient) obtained from each set (Ghosh and Mujumdar 2008). In the present case, we have assigned $K = 10$.

Deo and Samui (2017) advocated that the choice of appropriate training and testing data to construct and evaluate the performance of an appropriate model is a major task. According to the literature review, there is no basic rule for partitioning the complete dataset into

training and test subsets, e.g. 70% training and 30% testing (Samui and Dixon 2012); 69% training (Pal 2006); 63% training (Kurup and Dudani 2002). Generally, training/calibration of a nonlinear model requires a greater number of datasets. In this analysis, only the monsoon season is considered, which includes five months (i.e. June, July, August, September and October) for the period 1971–2000. Hence, the total length of the dataset consists of 150 data points. Therefore, based on the data availability and large number of datasets required to train the nonlinear model, 90% of the dataset is used as the training subset and the remaining 10% as the test subset. Due to the large numeric ranges in the predictor, it is always advisable to scale the dataset between 0 and 1 to avoid numerical problems (Hsu *et al.* 2003) and speed up the calculations (Deo and Samui 2017). Hence, scaling is carried out for the predictors before regression analysis. Gaussian and Laplacian RBFs are considered as kernel functions while performing machine learning regression, as the RBF has good potential to execute nonlinear mapping in a high-dimensional feature space (Hsu *et al.* 2003) and has better performance than the linear kernel (Keerthi and Lin 2003).

Performance evaluation of both SVM and RVM for different kernel functions is carried out using the Nash-Sutcliffe efficiency criterion (NSE) (Nash and Sutcliffe 1970), percent bias (PBIAS), and root mean square error observations standard deviation ratio (RSR), as mentioned by Moriasi *et al.* (2007) for monthly streamflow. The NSE assesses the model's capability to mimic the output data away from the mean statistics, PBIAS measures the over- or underestimation of modelled data with respect to the observed (Gupta *et al.* 1999), and the RSR evaluates the lower root mean square error (RMSE) with respect to the observational standard deviation.

3.6 Equiprobability transformation as bias correction

It is not possible to capture the entire variance of the predictand through regression-based downscaling and the resulting bias significantly affects the extreme events (Ghosh and Mujumdar 2008). In addition, it is difficult to incorporate the biases in large-scale atmospheric patterns or unrealistic inter-variable relationships in GCMs. Hence, bias correction is essential before assessing the climate change impact (Ahmed *et al.*, 2013), or else it will propagate in the computations of subsequent years (Das and Umamahesh 2016). Therefore, to correct the bias, GCM-simulated monsoon streamflows are mapped onto the observed

streamflows through a distribution function. Ines and Hansen (2006) used a gamma distribution to correct the bias in daily rainfall; similarly, Das and Umamahesh (2016) studied the monsoon rainfall by fitting a suitable distribution based on a chi-squared test. However, in the present analysis we use the Weibull distribution to correct the bias associated with the GCM-simulated monsoon streamflow. The detailed methodology consists of the following steps:

- Initially, cumulative distribution functions (cdfs) for the GCM downscaled and observed monsoon streamflow for 1971–2000 are computed using Weibull plotting positions.
- For a given value of monsoon streamflow simulated for 1971–2000 by the GCM, the corresponding cdf (cdf_{GCM}) is computed.
- For the particular cdf_{GCM} , the equivalent observed streamflow is obtained. The GCM-simulated streamflow data are replaced with the observed streamflow data and the correction factor is computed for the particular cdf.
- The first three steps are repeated with the all GCM-simulated monsoon streamflow data, considering 1971–2000 as the base-line period. The obtained correction is applied to the future simulated monsoon streamflow data.

The correction for bias that is derived from the base-line period is used to derive a bias-free estimate of monsoon streamflow in future scenarios. This method of bias correction is applied based on percentiles. For a particular percentile of observed (1971–2000) and simulated (1971–2000) monsoon streamflow, the correction is computed, and for the particular percentile in the future, the computed correction is applied. It should be noted that the bias in the streamflow is corrected based on the base-line period, and it is assumed that the bias will remain the same in the future for different scenarios (stationary bias). However, due to climate change, the higher percentile values of the base-line period may be projected as lower percentile values in the future or *vice versa* (Das and Umamahesh 2017b). Hence, this type of bias correction and the inherent assumptions will create additional uncertainty in the streamflow values with higher percentiles in the future. Since a large number of GCM projections are considered in the present study, this type of uncertainty is neglected. Moreover, due to its popularity in application, the equiprobability transformation technique is used in the present study. The correction is computed for all the percentiles and GCMs using Weibull plotting positions. The

correction is computed separately for different GCMs and applied separately for the respective GCMs. The transformation is defined as:

$$S_o = F_o^{-1}(F_m(S_m)) \quad (19)$$

where S_o and S_m are observed and modelled streamflow, respectively; F_m is the cdf of S_m and F_o^{-1} is the inverse cdf corresponding to S_o .

3.7 Nonparametric bootstrapping

Bootstrap resampling is widely used to perform uncertainty analysis in various fields of research (Lall and Sharma 1996, Mudelsee *et al.* 2003, Korenaga 2013). This resampling technique is easy to implement and it samples a sample population of fixed size with replacement and generates multiple realizations to compute the statistical properties (mean, variance, confidence interval, etc.) of interest (Schnaitt and Heinson 2015). Let a dataset X with random samples $X = (x_1, x_2, \dots, x_n)$ and empirical distribution \hat{F} have a probability of $1/n$ on each random observation. Then, the bootstrap sample can be defined to be a random sample of size n , i.e. $X^* = (x_{1*}, x_{2*}, \dots, x_{n*})$ drawn from \hat{F} . The bootstrap dataset $(x_{1*}, x_{2*}, \dots, x_{n*})$ consists of the member of the original dataset (x_1, x_2, \dots, x_n) with variability in the probability of occurrence of each random sample during resampling (i.e. some appearing zero times, some appearing once, some appearing twice, etc.) (Efron and Tibshirani 1993). In brief, bootstrapping is a technique that repeatedly simulates data from a given dataset to assess the uncertainty associated with the data samples, with the fundamental assumption that each data point in the dataset has a probability of recurrence in future (Yu *et al.* 2002). In the present application, bootstrapping is applied to the different monsoon streamflow percentiles ($Q_{0.1}$, Q_{10} , Q_{20} , Q_{30} , Q_{40} , Q_{50} , Q_{60} , Q_{70} , Q_{80} , Q_{90} , Q_{100}) to assess the associated uncertainty of the historical period. Here, 1000 realizations of bootstrap samples are generated using the “bootstrap” R package (<https://cran.r-project.org/package=bootstrap>) for each percentile and the 95% confidence interval is obtained from the resampled dataset to represent the uncertainty.

3.8 Change point detection using a Bayesian approach

In the context of climate change, identification of the time point with respect to the change in the statistical properties in hydro-meteorological time series is important (Perreault *et al.* 2000, Raje 2014). Moreover, shifting

from the traditional stationary assumptions due to the consequences of climate change as stated by the hydrologist (Milly *et al.* 2008), it is essential to consider change point (CP) analysis for better planning and optimal management of water resources (Raje 2014). In the present study, change point detection in mean annual monsoon streamflow series is executed using a Bayesian approach with the analysis proposed by Barry and Hartigan (1993). In the series of independent observations, whose density varies from each other depending on a parameter $\theta_i \in \Theta$. In this case, the product partition model developed by Hartigan (1990) is used to partition the data into several sets/blocks. The model assumes that the observations are independent and identically distributed with normal distribution, having constant means within the blocks and constant variance throughout each sequence. The specified uniform prior distribution for the partition probability ρ is updated through a Markov chain Monte Carlo (MCMC) algorithm using a Bayesian approach to obtain the posterior distribution of the same form (Crowley 1997), and there is an equal probability of drawing parameters from any location without affecting the likelihood function. Moreover, there are several classical methods for CP detection, such as the standard normal homogeneity (SNH) test and Wilcoxon's nonparametric rank sum test (Reeves *et al.* 2007), which provides a single year for CP location. The Bayesian approach has several advantages, such as it encompasses the uncertainty associated with parameters and data, includes the prior information, and overcomes the limitation of the frequentist approach (Raje 2014). Hence, it is considered to perform the task of CP detection. In the present situation, this method is applied on the future simulated mean annual monsoon streamflow under different climate change scenarios. The R package "bcp" (<https://cran.r-project.org/package=bcp>) is used to perform the analysis.

4 Results and discussion

To accomplish the objective of the present study, statistical downscaling was performed using machine learning techniques (SVM and RVM) and the most efficient technique was selected based on the performance criteria to project monsoon

streamflow over the Wainganga basin. An equiprobability transformation using the Weibull distribution was introduced to correct the post-modelling bias correction. Finally, the impact assessments on water availability with uncertainty analysis and change point detection using the Bayesian approach were executed.

4.1 Performance evaluation of SVM and RVM

As discussed in Section 3.5, the ability of the SVM and RVM to capture the historical streamflow using NCEP/NCAR data as predictors is evaluated using various statistical efficiencies, such as NSE, PBIAS and RSR. The present section compares the performance of SVM and RVM during the calibration and validation period.

As discussed in Section 3.3, for the support vector regression problem the insensitive zone width is controlled by ε and the flexibility of the model is regulated by c (regularization parameter). The parameter ε is assigned a value of 0.01 and c is defined according to Cherkassky and Mulier (2007) i.e. $(Y_{\max} - Y_{\min})$. Similarly, the hyperparameter in RVM is fixed by manual iteration to obtain a better performance index, and the assigned values for the hyperparameter are 2.10 and 1.85 for the Laplacian and Gaussian functions, respectively. The model efficiency obtained using SVM and RVM during the training and testing periods is presented in Table 1. From Table 1, it can be seen that RVM with Laplacian RBF gives comparatively better correlation coefficients (R) during training and testing periods. In addition, the overall NSE value in the case of RVM with Laplacian RBF is higher than that for the other models. Similarly, the training and testing PBIAS values are categorized as very good under performance-rating criteria. Moreover, the computational time and the possibility of overtraining are minimized due to the parsimony of RVM (Okkan and Inan 2015b).

In addition, scatter and histogram plots for the calibration and validation periods using SVM and RVM with different kernel functions, as well as the Taylor diagram, are depicted in Figure 5. It may be seen from Figure 5 that RVM with Laplacian RBF performs better than the other models and is able to capture the low and high values of

Table 1. Evaluation of model performance. VG: Very Good, G: Good, S: Satisfactory, US: Unsatisfactory, as suggested by Moriasi *et al.* (2007). Bold indicates the best performance.

Model	Subsets	R	PBIAS	RSR	NSE	Overall NSE
SVM _{Gaussian}	Training	0.796	-6.70 (VG)	0.78 (US)	0.628 (G)	0.604 (S)
	Testing	0.630	-16.4 (G)	0.61 (S)	0.346 (US)	
RVM _{Gaussian}	Training	0.852	-19 (S)	0.67 (S)	0.721 (G)	0.704 (G)
	Testing	0.751	-6.90 (VG)	0.52 (G)	0.508 (S)	
RVM _{Laplacian}	Training	0.944	0.40 (VG)	0.331 (VG)	0.889 (VG)	0.857 (VG)
	Testing	0.713	-0.80 (VG)	0.677 (S)	0.515 (S)	

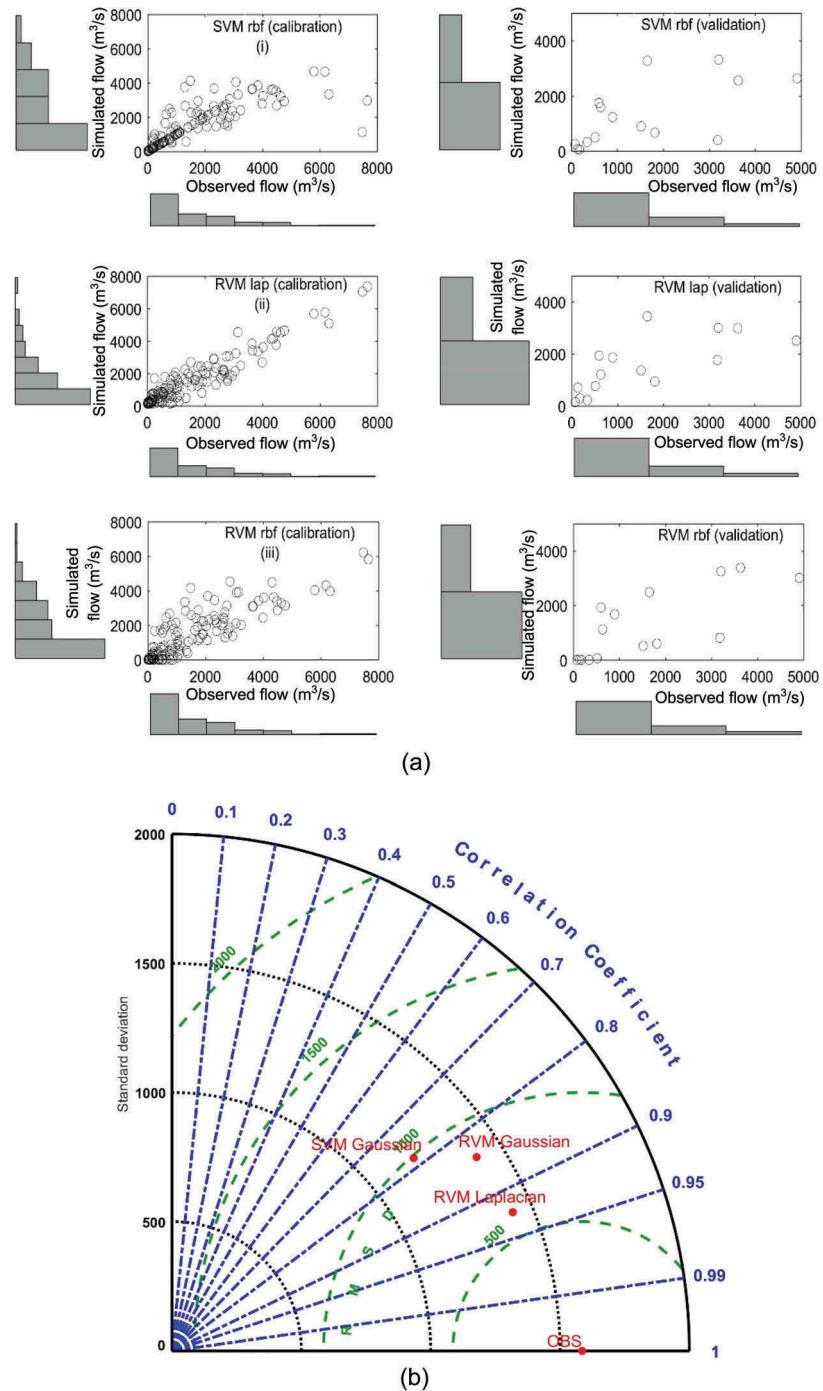


Figure 5. (a) Scatter and histogram plots between observed and NCEP for different kernel functions for the calibration and validation periods: (i) SVM with Gaussian, (ii) RVM with Laplacian, (iii) RVM with Gaussian; and (b) comparison using Taylor diagram.

monsoon streamflow. Hence, this model is selected to project the future streamflow under different scenarios.

4.2 Future projection of monsoon streamflow and analysis

To obtain the future projection of monsoon streamflow, pre-processing of the future predicted predictors is carried out. PCA analysis and fuzzy clustering of the

predictors are accomplished based on the obtained principal directions or eigenvectors and number of clusters with fuzzification coefficient during the NCEP predictors' pre-processing, respectively. Then, after selecting the downscaling model, the processed future projected predictors under different RCP scenarios from 40 different GCMs are given as input to the calibrated downscaling model. The results obtained for the future projection are divided into 30-year time

slices and are presented in the Appendix (Figs A1–A4). It can be observed from the downscaled monsoon streamflow for different scenarios that the projections are not similar, and streamflow varies significantly from model to model. These deviations could result from the use of different parameterization schemes, dissimilar model structure, and response of the GCMs to the atmospheric forcings (Teng *et al.* 2012). However, the inclusion of multi-model combinations will reduce the uncertainty associated with single model (Li and Sankarasubramanian 2012). In addition, combined multi-model projections improve the predictability, capturing the strength of single models (Ajami *et al.* 2007, Duan *et al.* 2007). Therefore, an ensemble average of all the models for different scenarios is considered to assess climate change impact on monsoon flow. The cdfs for ensemble monthly monsoon flows using Weibull plotting positions for the periods 1971–2000, 2011–2040, 2041–2070 and 2071–2100 are presented in Figure 6.

It is observed that the value of streamflow at which the cdf approaches 1 for the future time slices under different scenarios is significantly lower than that of the base-line period. This implies a reduction in the probability of occurrence of high-flow events in the future and hence there is likely to be a decreasing trend in the monthly peak flow during the monsoon season. Though there is a decreasing trend in the peak flows, the probability of occurrence of flows having cdfs of between 0.1 and 0.7 is likely to increase in future under different scenarios. In addition, the cdfs for all the time

slices under RCP2.6 do not have significant variations. However, for the high emissions scenarios there is remarkable variation in the cdfs for different time slices. The probability of occurrence of a complete range of monthly monsoon flows is likely to decrease in the future period under RCP4.5, RCP6.0 and RCP8.5.

The monthly variations in streamflow are computed and presented as violin plots in Figure 7. (A violin plot is a combination of a box plot and a density plot.) The violin plots are plotted for individual months and the white circles represent the mean values. It is observed that there is large variation in the monsoon flows during different months in the monsoon season for the base-line period. However, the variations in the future periods under different scenarios are significantly decreasing. For July, there is no significant difference in the mean flow with respect to the observed series for different scenarios. However, the mean monsoon flow for July is significantly decreasing as compared to the base-line period. In the case of October, the mean monthly streamflow for different periods under different scenarios is likely to increase. However, June and September exhibit a similar kind of pattern for different scenarios. Except for July, mean monthly flows in the monsoon season decrease gradually from the period 2011–2040 to 2071–2100. In the present study streamflow is projected based on the predictand (as discussed in Section 2), so it is difficult to attribute the change in the streamflow with respect to precipitation. However, in a recent study, Das and

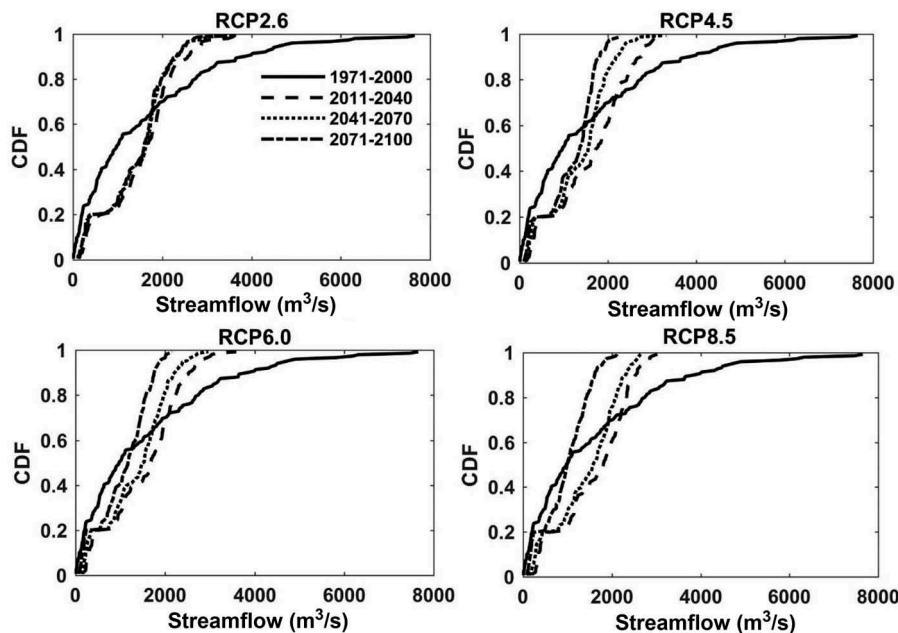


Figure 6. CDF plots for monthly monsoon flows under different scenarios.

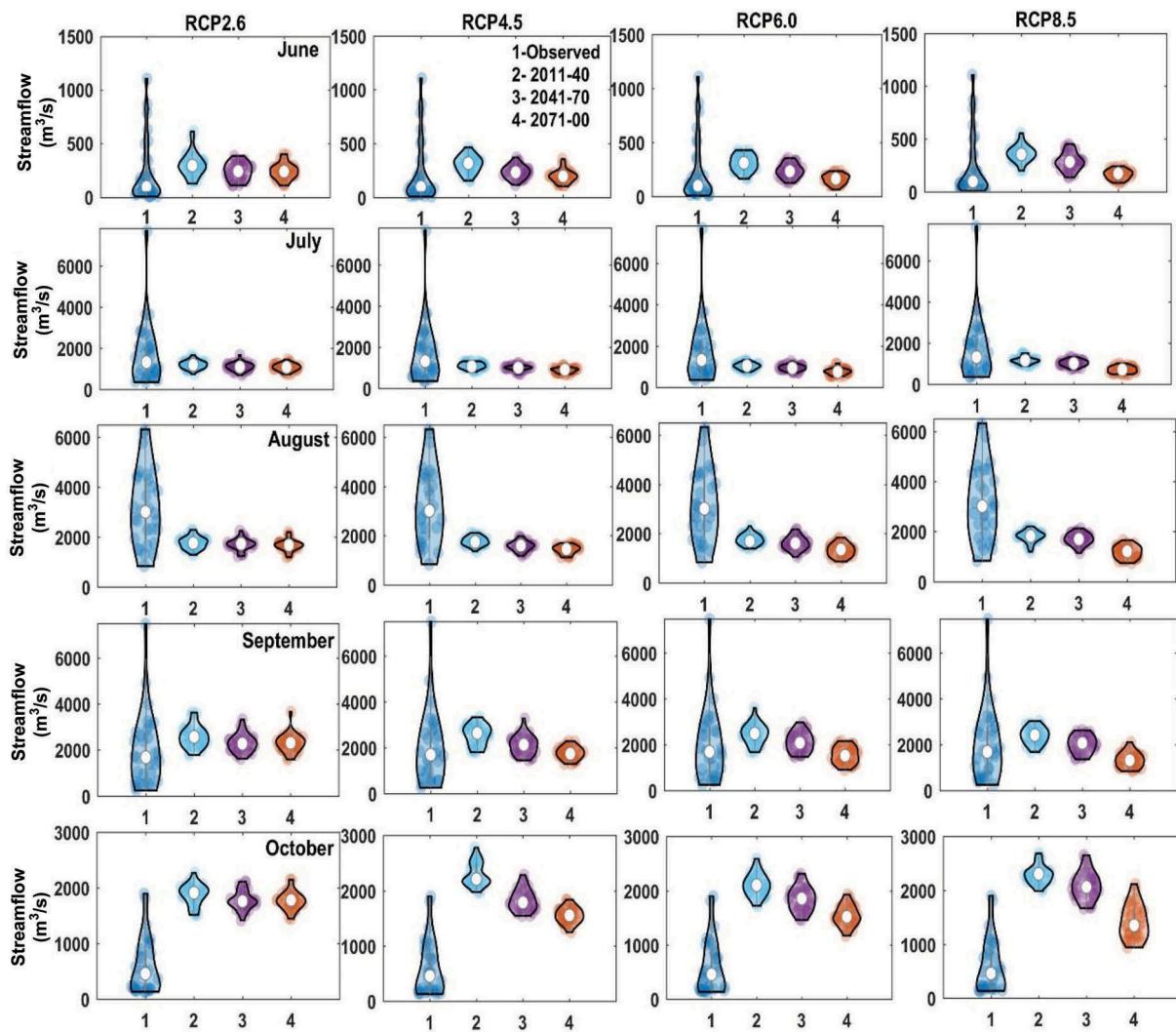


Figure 7. Month-wise violin plots under different scenarios.

Umamahesh (2017a) analysed the spatio-temporal variability of water balance components using a hydrological model over the Wainganga basin. They observed that monsoon rainfall during the month of October is likely to increase more significantly than during the other monsoon months. Hence, the increase in the streamflow in October can be attributed to the change in precipitation. In addition, Hengade and Eldho (2016) studied the climate change impact on the hydrology of the Wainganga basin using a hydrological model and advocated that change in the streamflow is more sensitive to the change in precipitation over the basin.

Moreover, the annual monsoon flows are also analysed and presented in Figure 8 as box plots. It can be seen from Figure 8 that under RCP2.6 there is no significant difference in the mean annual monsoon flows for all time slices. Moreover, the mean annual monsoon flow for 2011–2040 has no substantial variation

compared to the base-line period under all forcing scenarios. However, mean annual monsoon flows for the periods 2041–2070 and 2071–2000 are likely to decrease compared to the base-line period. Towards the end of the 21st century (i.e. 2071–2100), the decrease in the mean annual monsoon flow is more significant. Tang *et al.* (2012) analysed the sensitivity of streamflow over the Salmon River basin, Idaho, USA, and observed a 2–6% and 3–8% decline in mean annual streamflow with a temperature increase of 2 and 3°C, respectively.

Hegerl *et al.* (2007) stated that the global climate change resulting from long-term anthropogenic perturbations is generally quantified by the change in the temperature profile. In addition, according to the Intergovernmental Panel on Climate Change (IPCC 2007) the average surface temperature over India could rise by between 2 and 5°C. Hence, an attempt was made to analyse the sensitivity of monsoon streamflow over the Wainganga basin to the

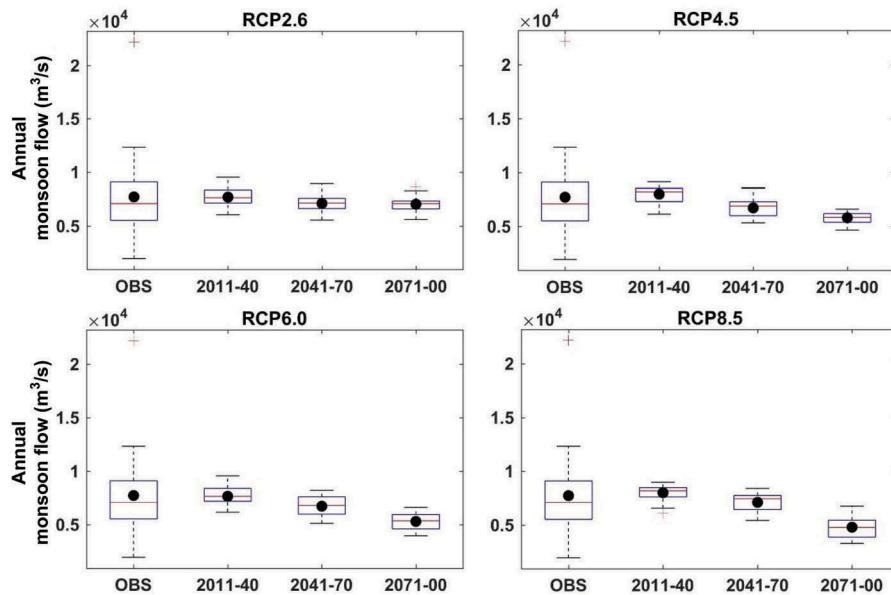


Figure 8. Box plots for annual monsoon flows under different scenarios. Black circle represents mean value.

increasing temperature profile, and the results are presented in [Figure 9](#).

It can be observed from [Figure 9](#) that, with a continuous increasing trend in the average temperature, the range of monthly streamflow decreases significantly as compared to the base-line period. In the box plot, the upper and lower edges of the box give the 75th and 25th percentiles of the dataset, respectively. The middle line suggests the median value. The 25th percentile of

the monsoon flow is likely to increase for all the time slices and scenarios. However, the 75th percentile flow is likely to decrease as compared to the base-line period. Hence, the variability or range of the monthly streamflow is expected to decrease. Although there is an increase in the median of the monsoon flow, the occurrence of peak flows will reduce significantly. A similar kind of study was carried out by Ghosh and Mujumdar ([2008](#)) over the Mahanadi basin, India, and

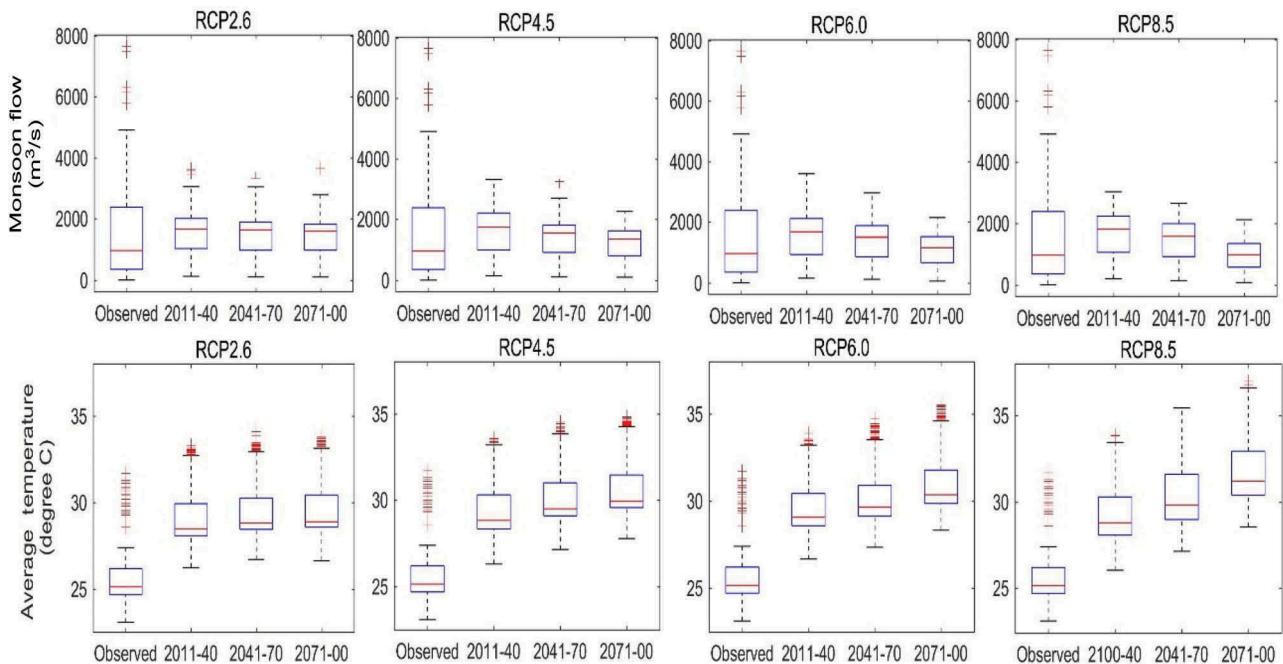


Figure 9. Box plots of monthly monsoon streamflow (upper panels) and change in the average temperature (lower panels) for different scenarios.

they attributed the decrease in the peak flow to high surface warming. In addition, Ghosh and Mujumdar (2008) stated that due to the groundwater component the low flows are insensitive to the climatic variables, and as a result effects on only high/peak flows are projected in the outcomes. Studies have been carried out using indirect downscaling methods to project the streamflow. Ashofteh *et al.* (2013) used the IHACRES hydrological model to project changes in the average annual runoff over the Aidoghoush River basin and observed a decreasing trend, mainly related to the high flows. Similarly, Yilmaz and Imteaz (2011) used the HEC-HMS hydrological model to evaluate possible climate change impacts on streamflow and found a substantial decrease in the runoff during summer and spring seasons over the Karasu basin, Turkey. Though peak flow is likely to decrease, the medium and low flows are likely to increase and will increase water availability. Detailed analysis is presented in the following section.

This study investigates the possible future changes in the monsoon flows using regression-based models. However, quantification of the streamflow based on the other water balance components is not carried out in the present study. According to Legesse *et al.* (2010), the increase in temperature will increase the potential evapotranspiration, which will result in decreased streamflow. Moreover, the potential evapotranspiration is very much sensitive to the change in the temperature rather than other climatic variables such as relative humidity, solar radiation and wind speed (Guo *et al.* 2017). In this sense, an increase in temperature over the present study area will likely increase the potential evapotranspiration, as most of the area is covered with forest and cropland, which will probably decrease the peak flows during the monsoon season.

4.3 FDC and uncertainty analysis

To understand water availability in the future, flow-duration curves (FDC) are constructed for the observed as well as for the projected monsoon flows under different scenarios. Mostly, the uncertainty analysis of FDC is carried out while performing the regionalization of the FDC at ungauged sites with observational uncertainties (Yu *et al.* 2002, Westerberg *et al.* 2014). However, in the present case, we analyse the uncertainty associated with the monsoon streamflow percentiles using the ensemble outputs from the GCMs under different scenarios. The uncertainty bound (95% CI) for the observed period is computed using the

nonparametric bootstrapping technique. To compute the 95% confidence interval, the 2.5 and 97.5 percentiles of the bootstrap samples are considered as the lower and upper bounds, respectively. The uncertainty bound for the observed period is plotted to find out whether the future projected FDC for different scenarios and time slices are falling within the uncertainty bound or not. Moreover, the uncertainties associated with the different scenarios are also discussed. Since large numbers of future realizations are obtained from the outputs of GCMs, the bootstrapping method is not used during the future periods.

Figure 10 depicts the comparison between the observed FDC with uncertainty bounds and GCM ensemble mean with uncertainty bounds for different future scenarios and time slices. The ordinate of Figure 10 is in log scale. It is observed from Figure 10 that the future projected GCM ensemble mean flows with exceedence probability higher than 0.4 are falling outside the confidence interval of the historical period.

In addition, Figure 10 depicts that the high flows from the GCM ensemble mean significantly decrease with respect to the observed period, whereas the medium and low flows are likely to increase. In the present study, the probabilities of exceedence in the ranges 0–0.4, 0.41–0.80 and 0.81–1.0 are referred as high, medium and low flows, respectively. This will also increase the water availability over the basin. The possible reason may be attributed to the significant contribution of the groundwater component to low flows, as indicated by Ghosh and Mujumdar (2008). In addition, Pechlivanidis *et al.* (2016) analysed the impact of climate change on future hydrological fluxes in the Indian subcontinent using the HYPE hydrological model. Based on the outcomes from the study, they stated that the water availability of Indian basins is likely to increase, except in northwestern basins due to their significant spatial variability. Similarly, Islam *et al.* (2014) advocated an increase in water yield in most of the Indian river systems for the 2030s and 2050s.

It is also clear from Figure 10 that a high uncertainty bound is noticed during the high flows and the uncertainty bound is decreasing towards the medium and low flows for historical and future periods. Moreover, the uncertainty bound obtained from the ensemble of GCMs is significantly larger than from the historical period. It is observed that the upper bound of the GCM uncertainty for the high flows projects very close to the upper bound of the historical uncertainty in most

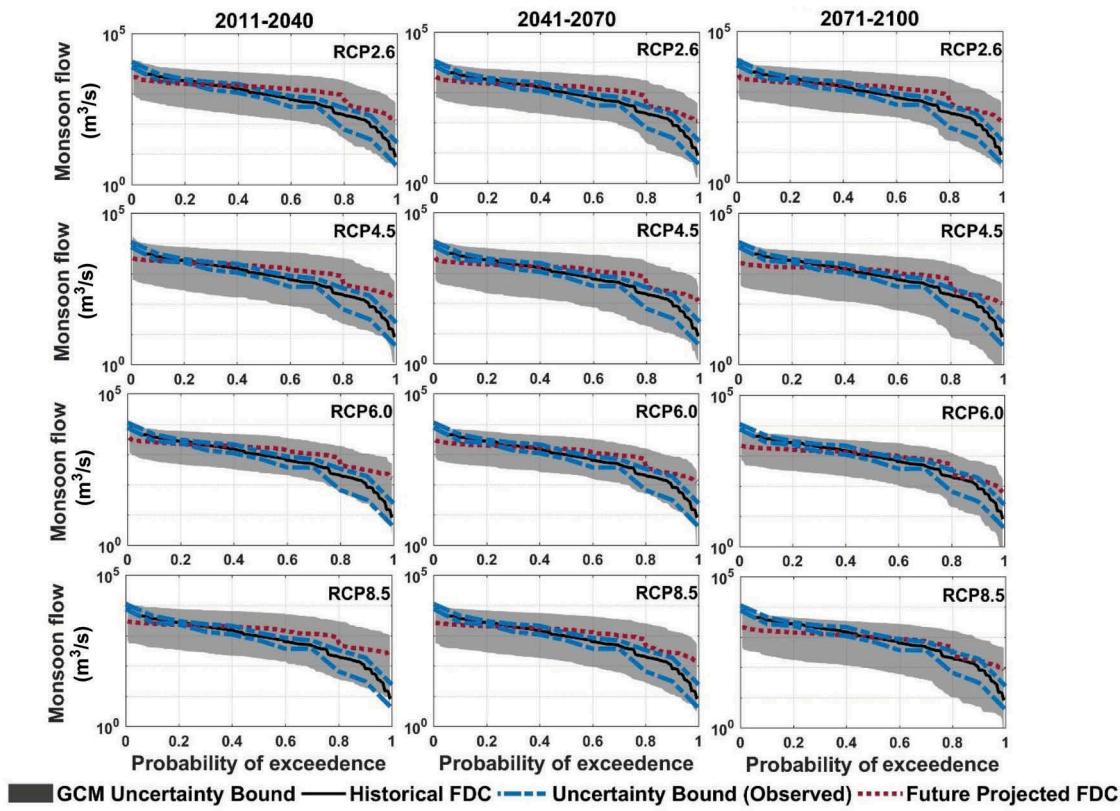


Figure 10. Uncertainty bounds associated with FDC for historical and different RCP scenarios: (left) 2011–2040; (middle) 2041–2070; and (right) 2071–2100 for all scenarios. Grey indicates the GCM uncertainty bound, the GCM ensemble mean flow is represented by dotted (red) lines and the confidence intervals for the historical period are represented by dashed (blue) lines.

cases. Therefore, except for very high flows and some cases with very low flows, the GCM uncertainty bound is significantly larger than the observed bound.

4.4 Change point analysis

With the increasing climate change variability, it has become indispensable to identify the change point in a hydrological series, as discussed previously (Section 3.7). With normally distributed annual monsoon streamflow having partition probability following a uniform distribution, 5000 realizations of parameter samples were obtained using an MCMC simulation. Before the MCMC simulation, the mean annual monsoon streamflows were checked for normal distributions using normal probability plots. The posterior density for CP location obtained from the Bayesian analysis is presented in Figure 11.

During historical period, the probability of change point in mean annual monsoon streamflow is high in 1993 and 1994. Moreover, under RCP2.6 scenario the high probability of change point is

detected in 2070. However, in case of RCP4.5, 6.0, and 8.5 higher probability is observed around 2056.

5 Conclusions

In this study, a classification and regression-based downscaling technique was used to project the monthly monsoon streamflow over the Wainganga River basin using the outputs of 40 GCMs (predictors) and four different RCP scenarios. Initial bias correction was carried out using the standardization technique with PCA to reduce the dimensionality and fuzzy clustering to classify the PCs. A machine learning technique was used as the downscaling procedure with better model efficiencies, and the ensemble average of multiple GCMs was considered to minimize the uncertainty associated with a single model. The investigation resulted in the following critical findings:

- The RVM outperforms the SVM in capturing the observed monsoon streamflow during the calibration and validation periods based on the different model performance criteria.

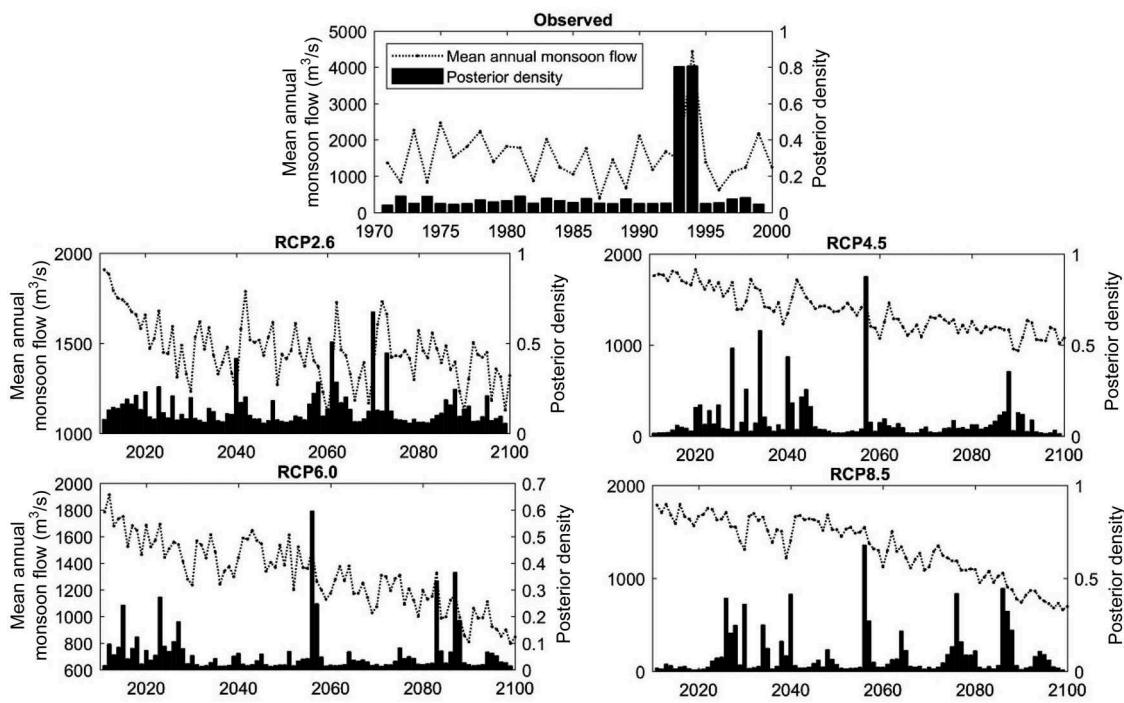


Figure 11. Posterior probability of CP location for observed and RCP scenarios.

- With an increase in the span of the future time period, the peak monsoon flow shows a continuous decreasing trend and the highest value is observed in the case of the high emissions scenario, RCP8.5.
- An increase in the surface temperature will significantly affect high flows in the Wainganga basin, whereas medium to low flows are expected to increase considerably with respect to the historical observations.
- The uncertainty associated with the FDC for the future projection reveals a high uncertainty bound for high flows, and the uncertainty bound is decreasing for the medium and low flows.
- Early detection of a CP (middle of the 21st century) is noticed in the case of the RCP4.5, RCP6.0 and RCP8.5 scenarios; however, in the case of RCP2.6, the year of the CP is detected in the latter part of the 21st century, i.e. around 2070.

In implementing the large number of GCMs and future scenarios, the outcomes should be considered as indicative of the response of the hydrological regime to climate change, rather than being conclusive. Encompassing the GCM and climate change scenario uncertainties, an attempt of this kind can result in more reliable outcomes to improve the robustness of planning and management decisions. In future, if any hydraulic structures are built over the basin to regulate the streamflow, then indirect

downscaling should be performed by incorporating such human interventions for better simulation of streamflow under a changing climate. Moreover, the adopted methodology of this study could be implemented to evaluate the variability in the streamflow to the inflow of a reservoir. This would help to modify the operation rules under climate change to improve the allocation and utilization of surface water resources for irrigation and hydropower generation.

According to the IPCC, the global climate has changed significantly since pre-industrial times, due to predominant anthropogenic interventions, and is threatening the sustainability of natural resources at the regional scale. Therefore, the present methodology can be used to analyse the impact on other hydro-climatic variables (e.g. rainfall, temperature, soil moisture, groundwater fluctuation, evaporation) for proper management at the regional scale. Here, GCM uncertainty was minimized by considering an ensemble average. However, uncertainty analysis can be improved by accounting a weighted ensemble average based on the performance criteria for each GCM. Further improvement may be achieved by encompassing scenario uncertainty, which will help to improve risk minimization practices.

Acknowledgements

The authors gratefully acknowledge the anonymous reviewers and Associate Editor (Dr Shaochun Huang) for



insightful and constructive comments to improve the manuscript significantly.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Aboutalebi, M., Bozorg-Haddad, O., and Loáiciga, H.A., 2016. Simulation of methyl tertiary butyl ether concentrations in river-reservoir systems using support vector regression. *Journal of Irrigation and Drainage Engineering*, 142 (6), 1–10. doi:[10.1061/\(ASCE\)IR.1943-4774.0001007](https://doi.org/10.1061/(ASCE)IR.1943-4774.0001007)
- Ahmed, K.F., et al., 2013. Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast. *Global and Planetary Change*, 100, 320–332. Elsevier B.V. doi:[10.1016/j.gloplacha.2012.11.003](https://doi.org/10.1016/j.gloplacha.2012.11.003)
- Ajami, N.K., Duan, Q., and Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, 43 (1), 1–19. doi:[10.1029/2005WR004745](https://doi.org/10.1029/2005WR004745)
- Anandhi, A., et al., 2008. Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology*, 28 (3), 401–420. doi:[10.1002/joc.1529](https://doi.org/10.1002/joc.1529)
- Ashofteh, P.S., Bozorg Haddad, O., and Mariño, M.A., 2013. Scenario assessment of streamflow simulation and its transition probability in future periods under climate change. *Water Resources Management*, 27 (1), 255–274. doi:[10.1007/s11269-012-0182-2](https://doi.org/10.1007/s11269-012-0182-2)
- Barry, D. and Hartigan, J.A., 1993. A Bayesian analysis for change point problems. *Journal of the American Statistical Association*, 88 (421), 309. doi:[10.2307/2290726](https://doi.org/10.2307/2290726)
- Bezdek, J., 1981. *Pattern recognition with fuzzy objective function algorithm*. New York: Plenum.
- Cannon, A.J. and Whitfield, P.H., 2002. Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models. *Journal of Hydrology*, 259 (1–4), 136–151. doi:[10.1016/S0022-1694\(01\)00581-9](https://doi.org/10.1016/S0022-1694(01)00581-9)
- Chen, H., Xu, C.-Y., and Guo, S., 2012. Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. *Journal of Hydrology*, 434–435, 36–45. Elsevier B.V. doi:[10.1016/j.jhydrol.2012.02.040](https://doi.org/10.1016/j.jhydrol.2012.02.040)
- Chen, S.-T., Yu, P.-S., and Tang, Y.-H., 2010. Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology*, 385 (1–4), 13–22. doi:[10.1016/j.jhydrol.2010.01.021](https://doi.org/10.1016/j.jhydrol.2010.01.021)
- Cherkassky, V. and Mulier, F., 2007. *Learning from data – concepts, theory and methods*. 2nd edition. Hoboken, NJ: A John Wiley & Sons, Inc. doi:[10.1198/tech.2001.s558](https://doi.org/10.1198/tech.2001.s558)
- Chiew, F.H.S. and McMahon, T.A., 2002. Modelling the impacts of climate change on Australian streamflow. *Hydrological Processes*, 16 (6), 1235–1245. doi:[10.1002/hyp.1059](https://doi.org/10.1002/hyp.1059)
- Chu, J.T., et al., 2010. Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China. *Theoretical and Applied Climatology*, 99 (1–2), 149–161. doi:[10.1007/s00704-009-0129-6](https://doi.org/10.1007/s00704-009-0129-6)
- Crowley, E.M., 1997. Product partition models for normal means. *Journal of the American Statistical Association*, 92 (437), 192–198. doi:[10.1080/01621459.1997.10473616](https://doi.org/10.1080/01621459.1997.10473616)
- Das, J. and Umamahesh, N.V., 2015. Multisite downscaling of monsoon precipitation over the Godavari River Basin under the RCP 4.5 Scenario. In: K. Karvazy and V.L. Webster, eds. *World environmental and water resources congress 2015*. American Society of Civil Engineers, 1066–1075. doi:[10.1061/9780784479162.105](https://doi.org/10.1061/9780784479162.105)
- Das, J. and Umamahesh, N.V., 2016. Downscaling monsoon rainfall over river Godavari Basin under different climate-change scenarios. *Water Resources Management*, 30 (15), 5575–5587. doi:[10.1007/s11269-016-1549-6](https://doi.org/10.1007/s11269-016-1549-6)
- Das, J. and Umamahesh, N.V., 2017a. Spatio-temporal variation of water availability in a river basin under CORDEX simulated future projections. *Water Resources Management*. doi:[10.1007/s11269-017-1876-2](https://doi.org/10.1007/s11269-017-1876-2)
- Das, J. and Umamahesh, N.V., 2017b. Uncertainty and non-stationarity in streamflow extremes under climate change scenarios over a river basin. *Journal of Hydrologic Engineering*, 22 (10), 4017042. doi:[10.1061/\(ASCE\)HE.1943-5584.0001571](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001571)
- Deo, R.C. and Samui, P., 2017. Forecasting evaporative loss by least-square support-vector regression and evaluation with genetic programming, Gaussian process, and minimax probability machine regression: case study of Brisbane City. *Journal of Hydrologic Engineering*, 22 (6), 5017003. doi:[10.1061/\(ASCE\)HE.1943-5584.0001506](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001506)
- Diaz-Nieto, J. and Wilby, R.L., 2005. A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the River Thames, United Kingdom. *Climatic Change*, 69 (2–3), 245–268. doi:[10.1007/s10584-005-1157-6](https://doi.org/10.1007/s10584-005-1157-6)
- Dibike, Y.B. and Coulibaly, P., 2005. Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. *Journal of Hydrology*, 307 (1–4), 145–163. doi:[10.1016/j.jhydrol.2004.10.012](https://doi.org/10.1016/j.jhydrol.2004.10.012)
- Duan, Q., et al., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources*, 30 (5), 1371–1386. doi:[10.1016/j.advwatres.2006.11.014](https://doi.org/10.1016/j.advwatres.2006.11.014)
- Duhan, D. and Pandey, A., 2015. Statistical downscaling of temperature using three techniques in the Tons River basin in Central India. *Theoretical and Applied Climatology*, 121 (3–4), 605–622. doi:[10.1007/s00704-014-1253-5](https://doi.org/10.1007/s00704-014-1253-5)
- Eckhardt, K., Fohrer, N., and Frede, H.G., 2005. Automatic model calibration. *Hydrological Processes*, 19 (3), 651–658. doi:[10.1002/hyp.5613](https://doi.org/10.1002/hyp.5613)
- Efron, B. and Tibshirani, R.J., 1993. *An introduction to the bootstrap*. Boca Raton, London, Newyork, Washington, D.C.: Chapman & Hall/CRC.
- Ghosh, S. and Mujumdar, P.P., 2008. Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources*, 31 (1), 132–146. doi:[10.1016/j.advwatres.2007.07.005](https://doi.org/10.1016/j.advwatres.2007.07.005)

- Githui, F., et al., 2009. Climate change impact on SWAT simulated streamflow in western Kenya. *International Journal of Climatology*, 29 (12), 1823–1834. doi:10.1002/joc.1828
- Goharian, E., et al., 2016. Incorporating potential severity into vulnerability assessment of water supply systems under climate change conditions. *Journal of Water Resources Planning and Management*, 142 (2), 4015051. doi:10.1061/(ASCE)WR.1943-5452.0000579
- Goyal, M.K., Ojha, C.S.P., and Burn, D.H., 2012. Nonparametric statistical downscaling of temperature, precipitation, and evaporation in a semiarid region in India. *Journal of Hydrologic Engineering*, 17 (5), 615–627. doi:10.1061/(ASCE)HE.1943-5584.0000479
- Güler, C. and Thyne, G.D., 2004. Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy c-means clustering. *Water Resources Research*, 40 (12), 1–11. doi:10.1029/2004WR003299
- Guo, D., Westra, S., and Maier, H.R., 2017. Sensitivity of potential evapotranspiration to changes in climate variables for different Australian climatic zones. *Hydrology and Earth System Sciences*, 21 (4), 2107–2126. doi:10.5194/hess-21-2107-2017
- Gupta, H.V., Sorooshian, S., and Yapo, P.O., 1999. Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4 (2), 135–143. doi:10.1061/(ASCE)1084-0699(1999)4:2(135)
- Hartigan, J.A., 1990. Partition models. *Communications in Statistics – Theory and Methods*, 19 (8), 2745–2756. doi:10.1080/03610929008830345
- Hashmi, M.Z., Shamseldin, A.Y., and Melville, B.W., 2013. Statistically downscaled probabilistic multi-model ensemble projections of precipitation change in a watershed. *Hydrological Processes*, 27 (7), 1021–1032. doi:10.1002/hyp.8413
- Hegerl, G.C., et al., 2007. Understanding and attributing climate change. In: S. Solomon, et al., eds *Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge, UK and New York, NY, USA: Cambridge University Press. doi:10.1007/s10584-006-9107-5
- Hengade, N. and Eldho, T.I., 2016. Assessment of LULC and climate change on the hydrology of ashti catchment, India using VIC model. *Journal of Earth System Science*, 125 (8), 1623–1634. doi:10.1007/s12040-016-0753-3
- Hsu, C.W., Chang, C.C., and Lin, C.J., 2003. *A practical guide to support vector classification*. Taipei, Taiwan: National Taiwan University.
- Ines, A.V.M. and Hansen, J.W., 2006. Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and Forest Meteorology*, 138 (1–4), 44–53. doi:10.1016/j.agrformet.2006.03.009
- IPCC, 2007. Climate change 2007: impacts, adaptation and vulnerability. In: Parry, O.F. et al., eds., *Contribution of working group II to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Islam, A., et al., 2014. Modeling water management and food security in India under climate change. In: L.R. Ahuja, L. Ma, and R.J. Lascano, eds. *Practical applications of agricultural system models to optimize the use of limited water*. Madison, WI: American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc., 267–316. doi:10.2134/advagricsystmodel.c11
- Joshi, D., et al., 2013. Databased comparison of Sparse Bayesian learning and multiple linear regression for statistical downscaling of low flow indices. *Journal of Hydrology*, 488, 136–149. Elsevier B.V. doi:10.1016/j.jhydrol.2013.02.040
- Joshi, D., et al., 2016. Comparison of direct statistical and indirect statistical-deterministic frameworks in downscaling river low-flow indices. *Hydrological Sciences Journal*, 61 (11), 1996–2010. doi:10.1080/02626667.2014.966719
- Keerthi, S.S. and Lin, C.-J., 2003. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, 15 (7), 1667–1689. doi:10.1162/089976603321891855
- Korenaga, J., 2013. Stacking with dual bootstrap resampling. *Geophysical Journal International*, 195 (3), 2023–2036. doi:10.1093/gji/ggt373
- Kuo, C.C., Gan, T.Y., and Higuchi, K., 2017. Evaluation of future streamflow patterns in lake simcoe subbasins based on ensembles of statistical downscaling. *Journal of Hydrologic Engineering*, 22 (9), 4017028. doi:10.1061/(ASCE)HE.1943-5584.0001548
- Kurup, P.U. and Dudani, N.K., 2002. Neural networks for profiling stress history of clays from PCPT data. *Journal of Geotechnical and Geoenvironmental Engineering*, 128 (7), 569–579. doi:10.1061/(ASCE)1090-0241(2002)128:7(569)
- Lall, U. and Sharma, A., 1996. A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, 32 (3), 679–693. doi:10.1029/95WR02966
- Landman, W.A., et al., 2001. Statistical downscaling of GCM simulations to streamflow. *Journal of Hydrology*, 252, 221–236. doi:10.1016/S0022-1694(01)00457-7
- Legesse, D., et al., 2010. Streamflow sensitivity to climate and land cover changes: Meki River, Ethiopia. *Hydrology and Earth System Sciences*, 14 (11), 2277–2287. doi:10.5194/hess-14-2277-2010
- Li, W. and Sankarasubramanian, A., 2012. Reducing hydrologic model uncertainty in monthly streamflow predictions using multimodel combination. *Water Resources Research*, 48 (12), 1–17. doi:10.1029/2011WR011380
- Loaiciga, H.A., et al., 1996. Global warming and the hydrologic cycle. *Journal of Hydrology*, 174 (1–2), 83–127. doi:10.1016/0022-1694(95)02753-X
- Milly, P.C.D., et al., 2008. Climate change. Stationarity is dead: whither water management? *Science (New York, N.Y.)*, 319 (5863), 573–574. doi:10.1126/science.1151915
- Moriasi, D.N., et al., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50 (3), 885–900. doi:10.13031/2013.23153

- Mudelsee, M., et al., 2003. No upward trends in the occurrence of extreme floods in central Europe. *Nature*, 425 (6954), 166–169. doi:[10.1038/nature01928](https://doi.org/10.1038/nature01928)
- Mujumdar, P.P. and Ghosh, S., 2008. Modeling GCM and scenario uncertainty using a probabilistic approach: application to the Mahanadi River, India. *Water Resources Research*, 44, 6. doi:[10.1029/2007WR006137](https://doi.org/10.1029/2007WR006137)
- Mujumdar, P.P. and Nagesh Kumar, D., 2012. *Floods in a changing climate: hydrologic modeling*. New York: Cambridge University Press.
- Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I – a discussion of principles. *Journal of Hydrology*, 10 (3), 282–290. doi:[10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Okkan, U. and Fistikoglu, O., 2014. Evaluating climate change effects on runoff by statistical downscaling and hydrological model GR2M. *Theoretical and Applied Climatology*, 117 (1–2), 343–361. doi:[10.1007/s00704-013-1005-y](https://doi.org/10.1007/s00704-013-1005-y)
- Okkan, U. and Inan, G., 2015a. Bayesian learning and relevance vector machines approach for downscaling of monthly precipitation. *Journal of Hydrologic Engineering*, 20 (4), 4014051. doi:[10.1061/\(ASCE\)HE.1943-5584.0001024](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001024)
- Okkan, U. and Inan, G., 2015b. Statistical downscaling of monthly reservoir inflows for Kemer watershed in Turkey: use of machine learning methods, multiple GCMs and emission scenarios. *International Journal of Climatology*, 35 (11), 3274–3295. doi:[10.1002/joc.4206](https://doi.org/10.1002/joc.4206)
- Okkan, U. and Kirdemir, U., 2016. Downscaling of monthly precipitation using CMIP5 climate models operated under RCPs. *Meteorological Applications*, 23 (3), 514–528. doi:[10.1002/met.1575](https://doi.org/10.1002/met.1575)
- Pal, M., 2006. Support vector machines-based modelling of seismic liquefaction potential. *International Journal for Numerical and Analytical Methods in Geomechanics*, 30 (10), 983–996. doi:[10.1002/nag.509](https://doi.org/10.1002/nag.509)
- Pang, B., et al., 2017. Statistical downscaling of temperature with the random forest model. *Advances in Meteorology*, 2017. doi:[10.1155/2017/7265178](https://doi.org/10.1155/2017/7265178)
- Pechlivanidis, I., et al., 2016. Multi-basin modelling of future hydrological fluxes in the Indian Subcontinent. *Water*, 8 (5), 177. doi:[10.3390/w8050177](https://doi.org/10.3390/w8050177)
- Perreault, L., et al., 2000. Retrospective multivariate Bayesian change-point analysis: a simultaneous single change in the mean of several hydrological sequences. *Stochastic Environmental Research and Risk Assessment*, 14 (4), 0243–0261. doi:[10.1007/s004770000051](https://doi.org/10.1007/s004770000051)
- Raje, D., 2014. Changepoint detection in hydrologic series of the mahanadi river basin using a fuzzy Bayesian approach. *Journal of Hydrologic Engineering*, 19 (4), 687–698. doi:[10.1061/\(ASCE\)HE.1943-5584.0000844](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000844)
- Raje, D. and Mujumdar, P.P., 2011. A comparison of three methods for downscaling daily precipitation in the Punjab region. *Hydrological Processes*, 25 (23), 3575–3589. doi:[10.1002/hyp.8083](https://doi.org/10.1002/hyp.8083)
- Reeves, J., et al., 2007. A review and comparison of change-point detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, 46 (6), 900–915. doi:[10.1175/JAM2493.1](https://doi.org/10.1175/JAM2493.1)
- Roubens, M., 1982. Fuzzy clustering algorithms and their cluster validity. *European Journal of Operational Research*, 10 (3), 294–301. doi:[10.1016/0377-2217\(82\)90228-4](https://doi.org/10.1016/0377-2217(82)90228-4)
- Sachindra, D.A., et al., 2013. Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. *International Journal of Climatology*, 33 (5), 1087–1106. doi:[10.1002/joc.3493](https://doi.org/10.1002/joc.3493)
- Sachindra, D.A., et al., 2015. Potential improvements to statistical downscaling of general circulation model outputs to catchment streamflows with downscaled precipitation and evaporation. *Theoretical and Applied Climatology*, 122 (1–2), 159–179. doi:[10.1007/s00704-014-1288-7](https://doi.org/10.1007/s00704-014-1288-7)
- Sachindra, D.A. and Perera, B.J.C., 2016. Annual statistical downscaling of precipitation and evaporation and monthly disaggregation. *Theoretical and Applied Climatology*. doi:[10.1007/s00704-016-1968-6](https://doi.org/10.1007/s00704-016-1968-6)
- Samui, P. and Dixon, B., 2012. Application of support vector machine and relevance vector machine to determine evaporative losses in reservoirs. *Hydrological Processes*, 26 (9), 1361–1369. doi:[10.1002/hyp.8278](https://doi.org/10.1002/hyp.8278)
- Schnaidt, S. and Heinson, G., 2015. Bootstrap resampling as a tool for uncertainty analysis in 2-D magnetotelluric inversion modelling. *Geophysical Journal International*, 203 (1), 92–106. doi:[10.1093/gji/ggv264](https://doi.org/10.1093/gji/ggv264)
- Simonovic, S.P., 2017. Bringing future climatic change into water resources management practice today. *Water Resources Management*, 31 (10), 2933–2950. doi:[10.1007/s11269-017-1704-8](https://doi.org/10.1007/s11269-017-1704-8)
- Singh, V. and Goyal, M.K., 2016. Analysis and trends of precipitation lapse rate and extreme indices over north Sikkim eastern Himalayas under CMIP5ESM-2M RCPs experiments. *Atmospheric Research*, 167, 34–60. Elsevier B.V. doi:[10.1016/j.atmosres.2015.07.005](https://doi.org/10.1016/j.atmosres.2015.07.005)
- Srivastava, P.K., et al., 2013. Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. *Water Resources Management*, 27 (8), 3127–3144. doi:[10.1007/s11269-013-0337-9](https://doi.org/10.1007/s11269-013-0337-9)
- Suykens, J.A.K. (2001) Nonlinear modelling and support vector machines. In: IMTC 2001. proceedings of the 18th IEEE instrumentation and measurement technology conference. Rediscovering measurement in the age of informatics (Cat. No.01CH 37188). IEEE, pp. 287–294. doi: [10.1109/IMTC.2001.928828](https://doi.org/10.1109/IMTC.2001.928828).
- Tang, C., et al., 2012. Assessing streamflow sensitivity to temperature increases in the Salmon River Basin, Idaho. *Global and Planetary Change*, 88–89, 32–44. Elsevier B.V. doi:[10.1016/j.gloplacha.2012.03.002](https://doi.org/10.1016/j.gloplacha.2012.03.002)
- Teng, J., et al., 2012. Estimating the relative uncertainties sourced from GCMs and hydrological models in modeling climate change impact on runoff. *Journal of Hydrometeorology*, 13 (1), 122–139. doi:[10.1175/JHM-D-11-058.1](https://doi.org/10.1175/JHM-D-11-058.1)
- Tipping, M., 2001. Sparse Bayesian learning and the relevance vector mach. *Journal of Machine Learning Research*, 1, 211–244. doi:[10.1162/15324430152748236](https://doi.org/10.1162/15324430152748236)
- Tisseuil, C., et al., 2010. Statistical downscaling of river flows. *Journal of Hydrology*, 385 (1–4), 279–291. Elsevier B.V. doi:[10.1016/j.jhydrol.2010.02.030](https://doi.org/10.1016/j.jhydrol.2010.02.030)
- Trigo, R.M. and Palutikof, J.P., 1999. Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach. *Climate Research*, 13 (1), 45–59. doi:[10.3354/cr013045](https://doi.org/10.3354/cr013045)

- Tripathi, S., Srinivas, V.V., and Nanjundiah, R.S., 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology*, 330 (3-4), 621–640. doi:10.1016/j.jhydrol.2006.04.030
- Vapnik, V.N., 1995. *The nature of statistical learning theory*. New York: Springer Verlag.
- Westerberg, I.K., et al., 2014. Regional water balance modeling using flow-duration curves with observational uncertainties. *Hydrology and Earth System Sciences*, 18 (8), 2993–3013. doi:10.5194/hess-18-2993-2014
- Wilby, R.L., et al., 1998. Statistical downscaling of general circulation model output: a comparison of methods. *Water Resources Research*, 34 (11), 2995–3008. doi:10.1029/98WR02577
- Xu, C., 1999. Climate change and hydrologic models: a review of existing gaps and recent research developments. *Water Resources Management*, 13, 369–382. doi:10.1023/A:1008190900459
- Yang, T., et al., 2012. Statistical downscaling of extreme daily precipitation, evaporation, and temperature and construction of future scenarios. *Hydrological Processes*, 26 (23), 3510–3523. doi:10.1002/hyp.8427
- Yilmaz, A.G. and Imteaz, M.A., 2011. Impact of climate change on runoff in the upper part of the Euphrates basin. *Hydrological Sciences Journal*, 56 (7), 1265–1279. doi:10.1080/02626667.2011.609173
- Yu, P.-S., Yang, T.-C., and Wang, Y.-C., 2002. Uncertainty analysis of regional flow duration curves. *Journal of Water Resources Planning and Management*, 128 (6), 424–430. doi:10.1061/(ASCE)0733-9496(2002)128:6(424)
- Zhou, J., et al., 2015. Integrated SWAT model and statistical downscaling for estimating streamflow response to climate change in the Lake Dianchi watershed, China. *Stochastic Environmental Research and Risk Assessment*, 29 (4), 1193–1210. doi:10.1007/s00477-015-1037-1

Appendix

Here we provide details of the GCMs and the future projections obtained under each GCM ([Table A1](#)) and the percentage decadal land-use change ([Table A2](#)). Figures A1–A4 present the downscaled future monthly monsoon streamflow under RCP2.6, RCP4.5, RCP6.0 and RCP8.5, respectively. The three panels in each figure show projections for: (top) 2011–2040; (middle) 2041–2070; (bottom) 2071–2100.

Table A1. Details of GCM and future projections considered for each GCM.

Model name (Latitude × Longitude)	Modelling centre	RCP2.6	RCP4.5	RCP6.0	RCP8.5
ACCESS1.0 (A1) 1.25 × 1.875	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia	✓	✓		
ACCESS1.3 (A2) 1.25 × 1.875	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia	✓	✓		
BCC-CSM1.1 (B1) 2.767 × 2.813	Beijing Climate Center, China Meteorological Administration	✓	✓	✓	✓
BCC-CSM1.1(m) (B2) 1.113 × 1.125	Beijing Climate Center, China Meteorological Administration	✓	✓	✓	✓
BNU-ESM (B3) 2.767 × 2.813	College of Global Change and Earth System Science, Beijing Normal University	✓	✓		
CanESM2 (C1) 2.767 × 2.813	Canadian Centre for Climate Modelling and Analysis	✓	✓		
CCSM4 (C2) 0.942 × 1.25	National Center for Atmospheric Research	✓	✓	✓	✓
CESM1(WACCM) (C3) 1.895 × 2.5	National Center for Atmospheric Research	✓	✓		
CESM1(BGC) (C4) 0.942 × 1.25	National Center for Atmospheric Research	✓	✓		
CESM1(CAM5) (C5) 0.942 × 1.25	National Center for Atmospheric Research	✓	✓	✓	✓
CMCC-CESM (C6) 3.68 × 3.75	Centro Euro-Mediterraneo per i Cambiamenti Climatici				✓
CMCC-CM (C7) 0.742 × 0.75	Centro Euro-Mediterraneo per i Cambiamenti Climatici	✓	✓		
CMCC-CMS (C8) 1.849 × 1.875	Centro Euro-Mediterraneo per i Cambiamenti Climatici	✓	✓		
CNRM-CM5 (C9) 1.389 × 1.406	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancées en Calcul Scientifique	✓	✓		
CSIRO-MK3-6-0 (C10) 1.849 × 1.875	Commonwealth Scientific and Industrial Research Organization	✓	✓	✓	✓
EC-EARTH (E1) 1.113 × 1.125	Irish Centre for High-End Computing	✓	✓		
FGOALS-g2 (F1) 6.131 × 2.813	Institute of Atmospheric Physics, Chinese Academy of Sciences	✓	✓		
FIO-ESM (F2) 2.767 × 2.813	The First Institute of Oceanography, SOA	✓	✓	✓	✓
GFDL-CM2.1 (G1) 1.516 × 2.5	NOAA Geophysical Fluid Dynamics Laboratory	✓			
GFDL-CM3 (G2) 2 × 2.5	NOAA Geophysical Fluid Dynamics Laboratory	✓	✓		
GISS-E2-H (G3) 2 × 2.5	NASA Goddard Institute for Space Studies	✓	✓	✓	✓
GISS-E2-H-CC (G4) 2 × 2.5	NASA Goddard Institute for Space Studies	✓			
GISS-E2-R (G5) 2 × 2.5	NASA Goddard Institute for Space Studies	✓	✓	✓	✓
GISS-E2-R-CC (G6) 2 × 2.5	NASA Goddard Institute for Space Studies	✓			
HadGEM2-AO (H1) 1.25 × 1.875	Met Office Hadley Centre	✓	✓	✓	✓
HadGEM2-CC (H2) 1.25 × 1.875	Met Office Hadley Centre	✓			
HadGEM2-ES (H3) 1.25 × 1.875	Met Office Hadley Centre	✓	✓	✓	✓
INMCM4.0 (I1) 1.5 × 2	Institute for Numerical Mathematics	✓			
IPSL-CM5A-LR (I2) 1.895 × 3.75	Institut Pierre-Simon Laplace	✓	✓	✓	✓
IPSL-CM5A-MR (I3) 1.268 × 2.5	Institut Pierre-Simon Laplace	✓	✓	✓	✓

(Continued)

Table A1. (Continued).

Model name (Latitude × Longitude)	Modelling centre	RCP2.6	RCP4.5	RCP6.0	RCP8.5
IPSL-CM5B-LR (I4) 1.835 × 3.75	Institut Pierre-Simon Laplace	✓	✓	✓	✓
MIROC-ESM (M1) 2.767 × 2.813	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, and National Institute for Environmental Studies	✓	✓	✓	✓
MIROC5 (M2) 1.389 × 1.406	Atmosphere and Ocean Research Institute, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	✓	✓	✓	✓
MIROC4h (M3) 0.557 × 0.563	Atmosphere and Ocean Research Institute, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology		✓		
MIROC-ESM-CHEM (M4) 2.767 × 2.813	Atmosphere and Ocean Research Institute, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	✓	✓	✓	✓
MPI-ESM-LR (M5) 1.849 × 1.875	Max Planck Institute for Meteorology	✓	✓	✓	✓
MPI-ESM-MR (M6) 1.849 × 1.875	Max Planck Institute for Meteorology	✓	✓	✓	✓
MRI-CGCM3 (M7) 1.113 × 1.125	Meteorological Research Institute	✓	✓	✓	✓
NorESM1-M (N1) 1.895 × 2.5	Bjerknes Centre for Climate Research, Norwegian Meteorological Institute	✓	✓	✓	✓
NorESM1-ME (N2) 1.895 × 2.5	Bjerknes Centre for Climate Research, Norwegian Meteorological Institute	✓	✓	✓	✓

Table A2. Decadal LULC classification (%).

Classification	1985	1995	2005
Deciduous broadleaf forest	26.11	26.96	24.46
Cropland	49.30	47.66	48.94
Built-up land	0.69	0.72	0.73
Mixed forest	1.05	1.06	1.08
Shrubland	3.20	3.17	3.00
Barren land	0.01	0.02	0.02
Fallow land	3.53	5.49	6.03
Wasteland	0.63	0.56	0.74
Water bodies	2.68	2.82	1.09
Plantations	0.26	0.26	0.28
Grassland	0.001	0.001	0.001
Deciduous needleleaf forest	12.49	11.22	13.59

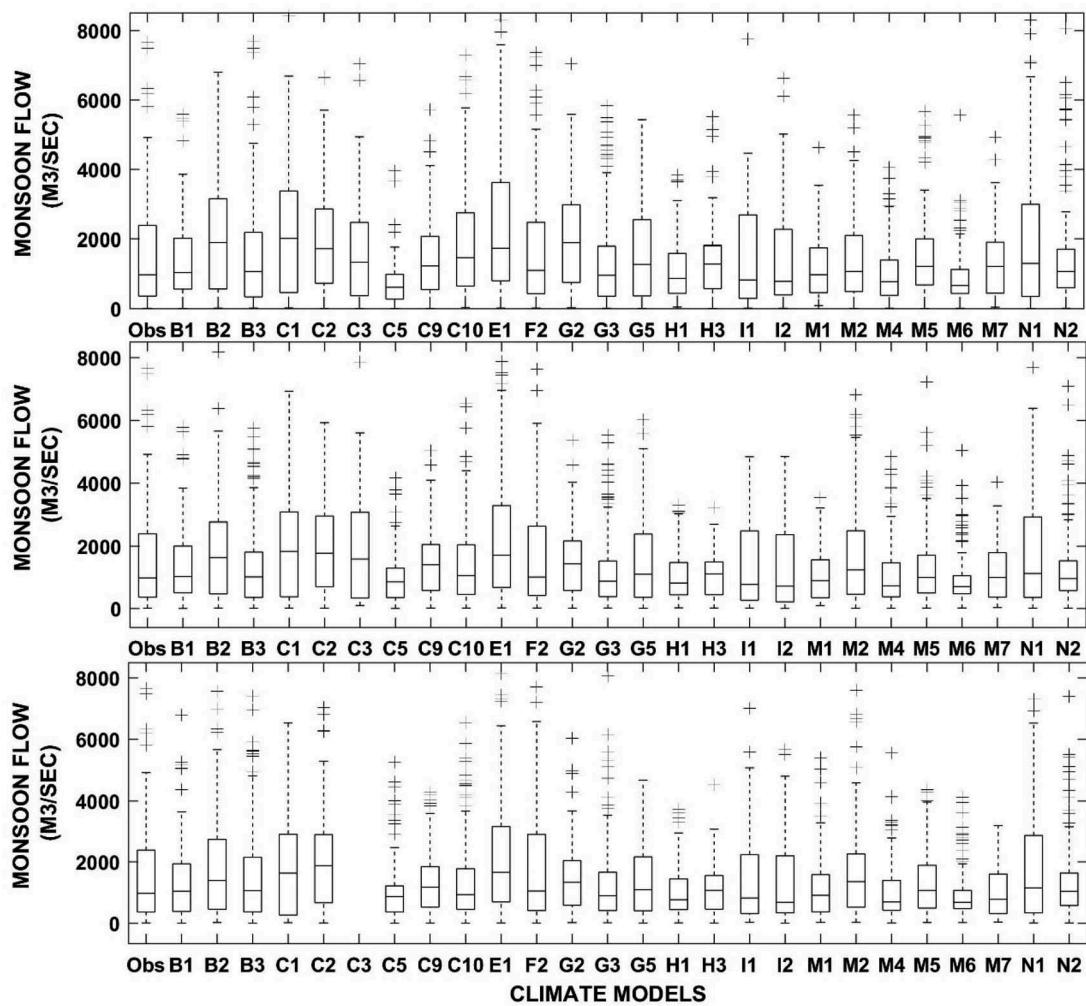


Figure A1. Downscaled monthly monsoon streamflow under RCP2.6.

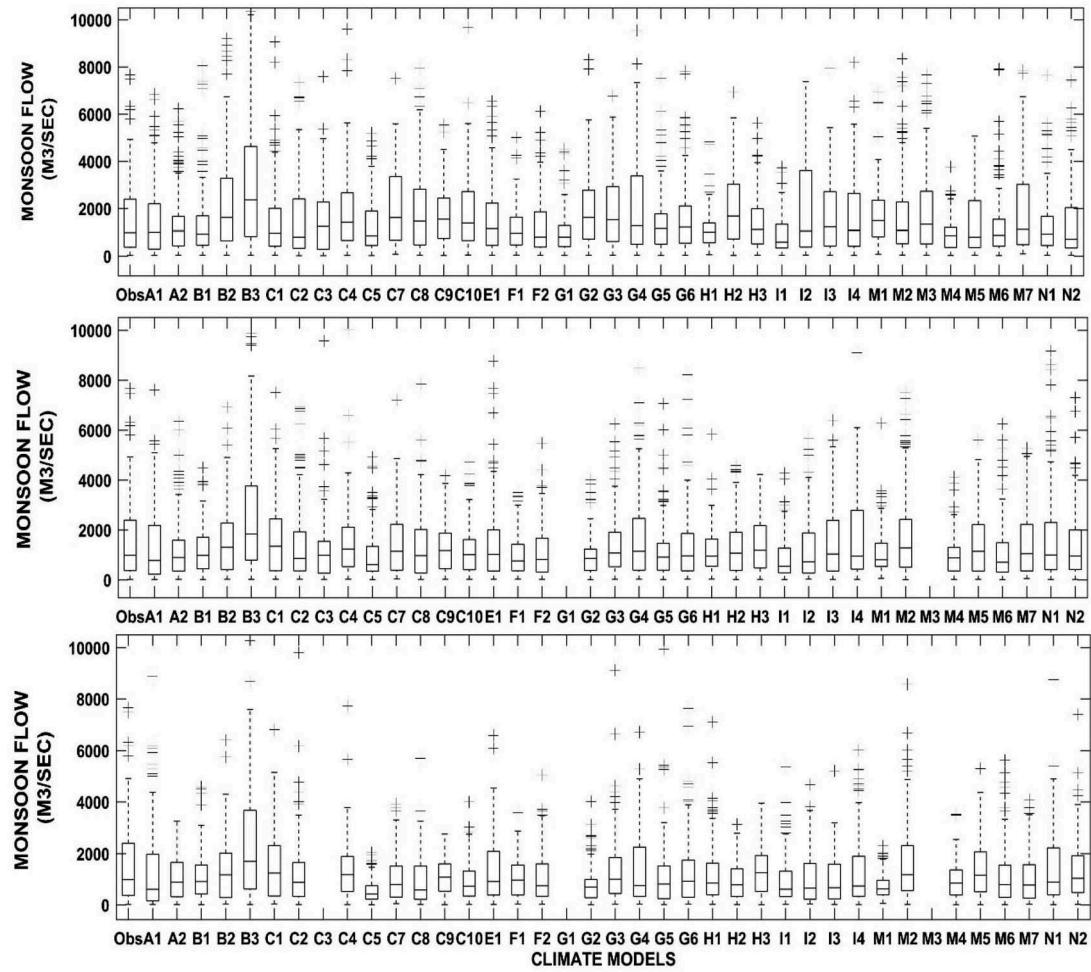


Figure A2. Downscaled monthly monsoon streamflow under RCP4.5.

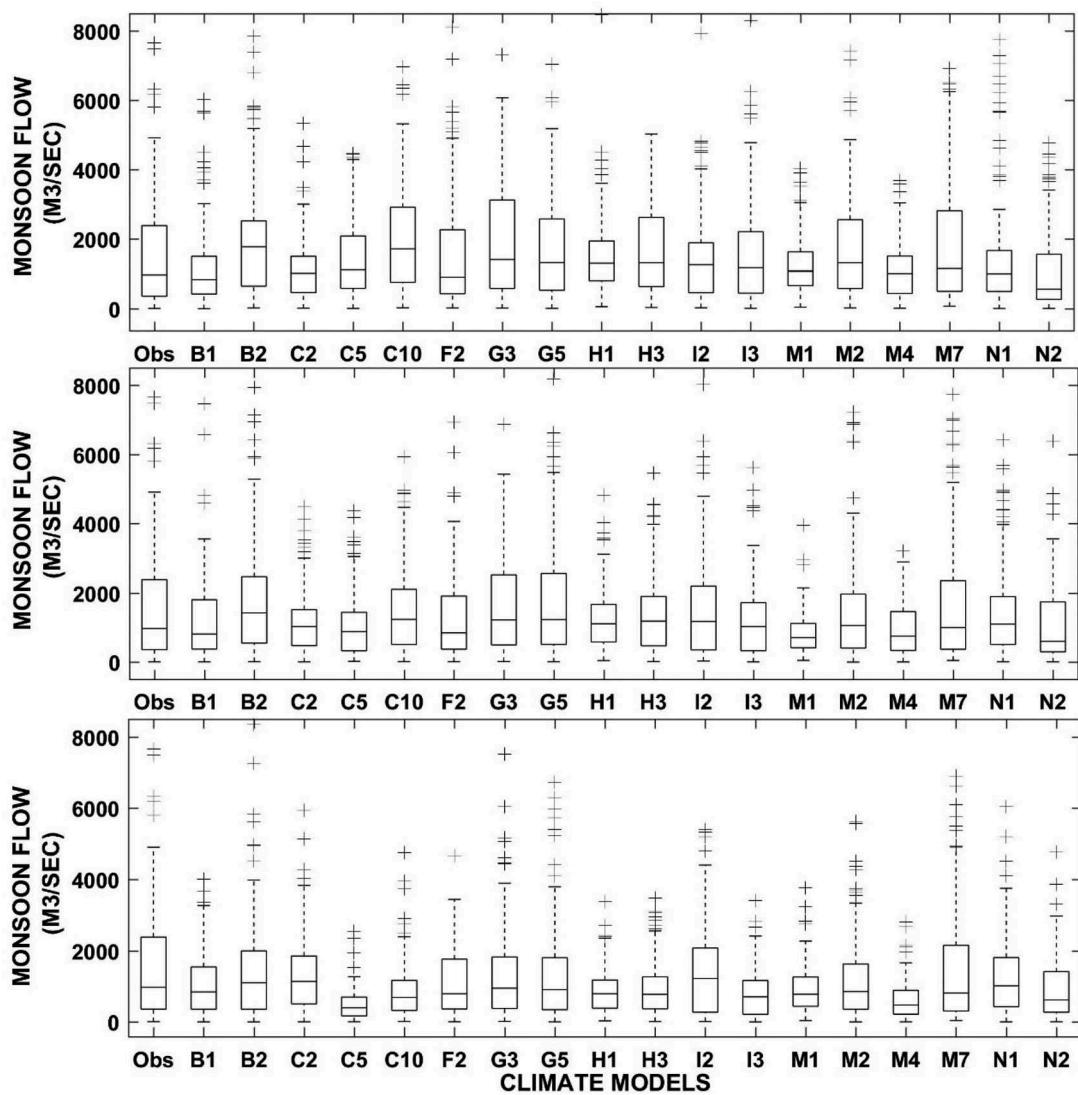


Figure A3. Downscaled monthly monsoon streamflow under RCP6.0.

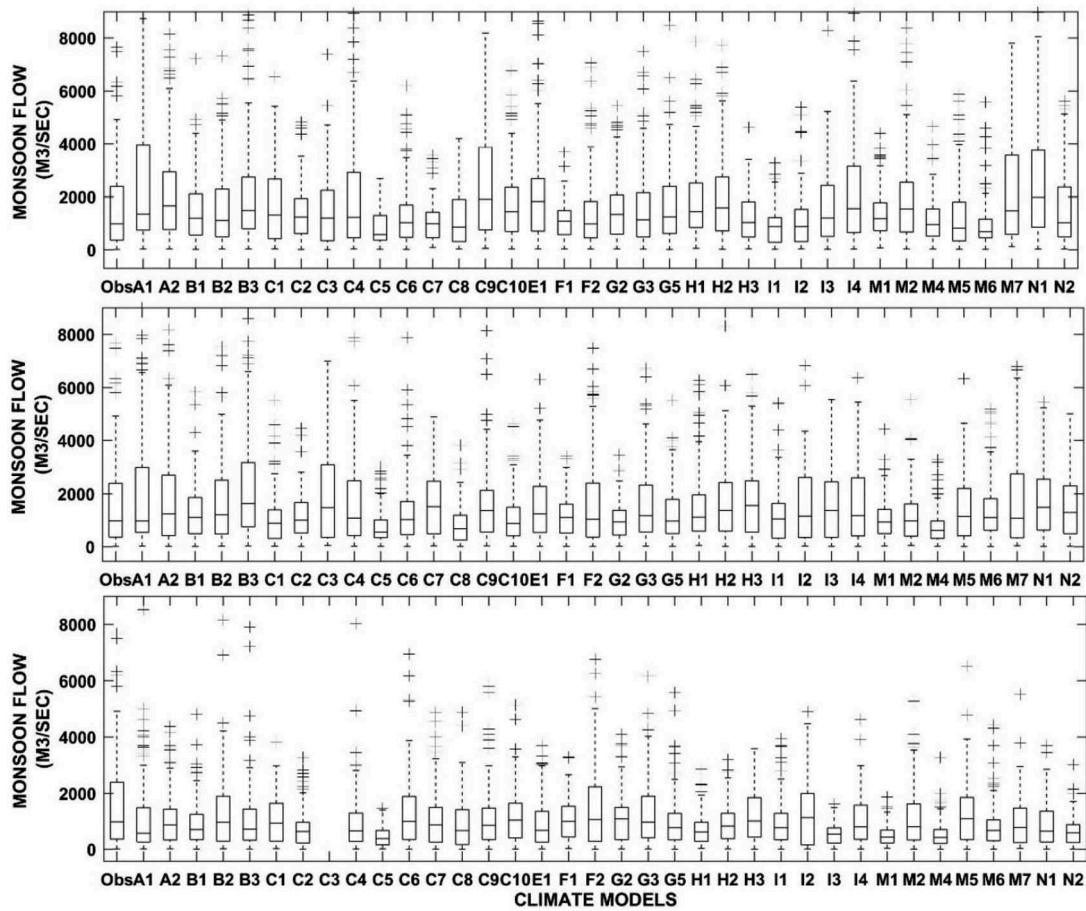


Figure A4. Downscaled monthly monsoon streamflow under RCP8.5.