Available online at www.sciencedirect.com
ScienceDirect
Advances in Water Resources 31 (2008) 132–146
Advances in Water Resources
www.elsevier.com/locate/advwatres

Statistical downscaling of GCM simulations to streamflow using relevance vector machine

Subimal Ghosh, P.P. Mujumdar *

Department of Civil Engineering, Indian Institute of Science, Bangalore, Karnataka 560 012, India

Received 21 February 2007; received in revised form 20 July 2007; accepted 25 July 2007
Available online 15 August 2007

Abstract

General circulation models (GCMs), the climate models often used in assessing the impact of climate change, operate on a coarse scale and thus the simulation results obtained from GCMs are not particularly useful in a comparatively smaller river basin scale hydrology. The article presents a methodology of statistical downscaling based on sparse Bayesian learning and Relevance Vector Machine (RVM) to model streamflow at river basin scale for monsoon period (June, July, August, September) using GCM simulated climatic variables. NCEP/NCAR reanalysis data have been used for training the model to establish a statistical relationship between streamflow and climatic variables. The relationship thus obtained is used to project the future streamflow from GCM simulations. The statistical methodology involves principal component analysis, fuzzy clustering and RVM. Different kernel functions are used for comparison purpose. The model is applied to Mahanadi river basin in India. The results obtained using RVM are compared with those of state-of-the-art Support Vector Machine (SVM) to present the advantages of RVMs over SVMs. A decreasing trend is observed for monsoon streamflow of Mahanadi due to high surface warming in future, with the CCSR/NIES GCM and B2 scenario.

Keywords: GCM; Statistical downscaling; Relevance vector machine; Streamflow

1. Introduction

Modeling hydrologic impacts of climate change involves simulation results from General Circulation Models (GCMs), which are the most credible tools designed to simulate time series of climate variables globally, accounting for the effects of greenhouse gases in the atmosphere. GCMs perform reasonably well in simulating climatic variables at larger spatial scale ($>10^4$ km^2), but poorly at the smaller space and time scales relevant to regional impact analyses [5]. Such poor performances of GCMs at local and regional scales have led to the development of Limited Area Models (LAMs) in which fine computational grid over a limited domain is nested within the coarse grid of a GCM [24]. This procedure is also known as dynamic downscaling. The major drawback of dynamic downscaling, which restricts its use in climate change impact studies, is its complicated design and high computational cost. Moreover, it is inflexible in the sense that expanding the region or moving to a slightly different region requires redoing the entire experiment [11]. Another approach to downscaling, termed statistical downscaling, involves deriving empirical relationships that transform large scale features of the GCM (Predictors) to regional scale variables (Predictands) such as precipitation and streamflow. There are three implicit assumptions involved in statistical downscaling [18]. Firstly, the predictors are variables of relevance and are realistically modeled by the host GCM. Secondly, the empirical relationship is valid also under altered climatic conditions. Thirdly, the predictors employed fully represent the climate change signal.

Statistical downscaling methodologies can be broadly classified into three categories [31,54]: weather generators, weather typing and transfer function. Weather generators

* Corresponding author. Tel.: +91 80 2360 0290; fax: +91 80 2360 0290.
E-mail addresses: subimal@civil.iisc.ernet.in (S. Ghosh), pradeep@civil.iisc.ernet.in (P.P. Mujumdar).

0309-1708/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2007.07.005

are statistical models of observed sequences of weather variables. They can also be regarded as complex random number generators, the output of which resembles daily weather data at a particular location [26]. There are two fundamental types of daily weather generators, based on the approach to model daily precipitation occurrence: the Markov chain approach [21–23,30] and the spell-length approach [56]. In the Markov chain approach, a random process is constructed which determines a day at a station as rainy or dry, conditional upon the state of the previous day, following given probabilities. In case of spell-length approach, instead of simulating rainfall occurrences day by day, spell-length models operate by fitting probability distribution to observed relative frequencies of wet and dry spell lengths. In either case, the statistical parameters extracted from observed data are used along with some random components to generate a similar time series of any length. Weather typing approaches [7] involve grouping of local, meteorological variables in relation to different classes of atmospheric circulation. Future regional climate scenarios are constructed either by resampling from the observed variable distribution (conditioned on the circulation pattern produced by a GCM), or by first generating synthetic sequences of weather pattern using Monte Carlo techniques and then resampling from the generated data. The mean, or frequency distribution of the local climate is then derived by weighting the local climate states with the relative frequencies of the weather classes. The most popular approach of downscaling is the use of transfer function which is a regression based downscaling method [11,9,53,42] that relies on direct quantitative relationship between the local scale climate variable (predictand) and the variables containing the large scale climate information (predictors) through some form of regression. Individual downscaling schemes differ according to the choice of mathematical transfer function, predictor variables or statistical fitting procedure. Todate, linear and nonlinear regression, Artificial Neural network (ANN), canonical correlation, etc. have been used to derive predictor–predit-and relationship. Among them, ANN based downscaling techniques have gained wide recognition owing to their ability to capture nonlinear relationships between predictors and predictand [11,19,45,41,51].

Despite a number of advantages, the traditional neural network models have several drawbacks including possibility of getting trapped in local minima and subjectivity in the choice of model architecture [43]. Recently, Vapnik [46,47] pioneered the development of a novel machine learning algorithm, called Support Vector Machine (SVM), which provides an elegant solution to these problems. The SVM has found wide range of applications in the fields of classification and regression analysis. SVM has some drawbacks of rapid increase of basis functions with the size of training data set and absence of probabilistic interpretation [15]. Recently Tipping [44] developed Relevance Vector Machine (RVM), a new methodology for classification and regression using the concept of probabilistic bayesian learning framework, which can predict accu-

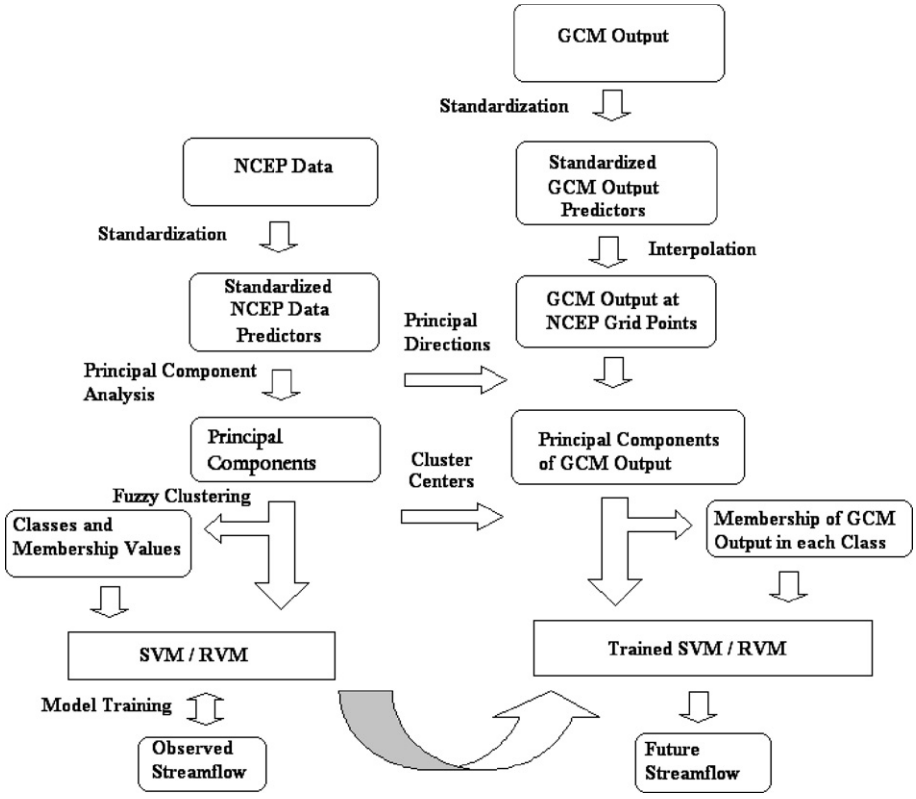

Fig. 1. Flowchart of the proposed model.

rately utilizing dramatically fewer basis functions than a comparable SVM while offering a number of additional advantages.

In a recent study [42], SVM has been used as a downscaling technique for predicting subdivisional precipitation of different regions in India. In that study, the GCM generated large scale output (predictors) are converted into principal components using Principal Component Analysis (PCA) and used directly as an input to SVM with Gaussian RBF as the kernel function. Ghosh and Mujumdar [13] found that a heuristic classification of large scale GCM outputs based on fuzzy clustering, prior to regression, improves the model performance and thus in the present study both SVM and RVM coupled with PCA and fuzzy clustering are used to downscale GCM output to streamflow. The flowchart of the model is presented in Fig. 1. The large scale GCM outputs are converted into principal components using PCA, which is further classified into fuzzy clusters using fuzzy c-mean clustering. The membership in each of the clusters along with the principal components is used as input to SVM/RVM. The relationship between the climate variables and streamflow is complex and nonlinear. Standard regression methods such as linear regression fail to model such nonlinear processes, and therefore SVM and RVM are used in the present study. Gaussian RBF, Laplacian RBF and heavy tailed RBF have been used as the kernel functions to compare the results. The National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis data have been used for training the downscaling model and GCM output is used for projecting future streamflow with the trained model. The performance of RVM is compared with SVM for downscaling in the present study. Results are obtained with different kernel functions. The model is applied to the case study of Mahanadi river basin in India to model the reservoir inflow to the Hirakud dam from large scale GCM output. Details of the case-study, data and the analysis performed prior to the training of SVM or RVM, is presented in the following section.

2. Data and input to vector machine

The Hirakud dam is located at Mahanadi River in Orissa at east coast of India (Fig. 2). The latitude and the longitude of the location are 21.32°N and 83.45°E, respectively. The monthly inflow to Hirakud dam from 1961 to 1990, is obtained from Department of Irrigation, Government of Orissa, India. Due to the absence of any major control structure upstream to Hirakud dam, the inflow to the dam is considered as unregulated flow. Mahanadi is a rain-fed river with high streamflow in monsoon (June, July, August and September) due to heavy rainfall and therefore the ground water component with infiltration is insignificant compared to the streamflow during the monsoon season. In the non-monsoon season, infiltration to ground water is quite significant in absence of rainfall, resulting in low streamflow in Mahanadi with almost dry conditions. Thus, only for the monsoon season the streamflow can be modeled with the climatological variables without considering ground water component. Therefore the monthly monsoon flow data of Mahanadi from year 1961 to year 1990 is used in the downscaling model as predictand. Selection of predictor is an important step in statistical downscaling. The predictors, used for downscaling [49,52] should be: (1) reliably simulated by GCMs, (2) readily available from archives of GCM outputs, and (3) strongly correlated with the surface variables of interest. Cannon and Whitfield [9] have used MSLP, 500 hPa geopotential height, 800 hPa specific humidity, and 100–500 hPa thickness field as the predictors for downscaling GCM output to streamflow. Monsoon streamflow can be considered broadly as the resultant of rainfall and evaporation. Rainfall is a consequence of Mean Sea Level Pressure

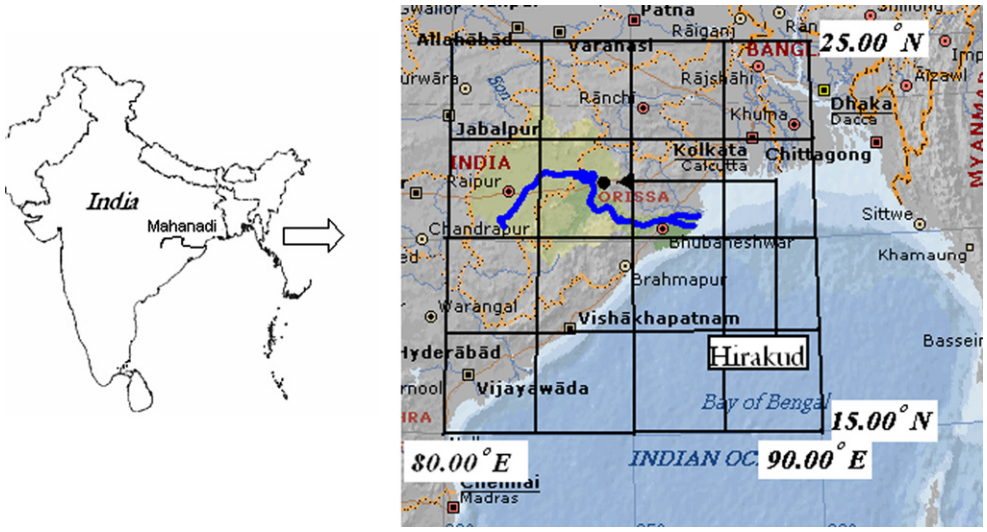

Fig. 2. NCEP grids superposed on Mahanadi River Basin.

(MSLP) [2,3,22,49], geopotential height and humidity whereas evaporation is mainly guided by temperature and humidity. Therefore, The present study considers 2 m surface air temperature, MSLP, 500 hPa geopotential height and surface specific humidity as the predictors for modeling Mahanadi streamflow in monsoon season. It is worth mentioning that land use is the single most important factor in generating the flow from the rainfall. In the present study, land use pattern is assumed to remain the same in future and therefore the statistical relationship between the predictors and the streamflow will remain unaltered in future. Gridded climate variables are obtained from the National Center for Environmental Prediction/ National Center for Atmospheric Research (NCEP/ NCAR) reanalysis project [25] (http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml). Reanalysis data are outputs from a high resolution atmospheric model that has been run using data assimilated from surface observation stations, upper-air stations, and satellite-observing platforms. Results obtained using these fields therefore represent those that could be expected from an ideal GCM [9]. Monthly climatological data from 1961 to 1990 were obtained for a region spanning 15°N–25°N in latitude and 80°E–90°E in longitude. Fig. 2 shows the NCEP grid points superposed on the map of Mahanadi river basin. A statistical relationship based on fuzzy clustering and vector machine is developed between large scale climatic variables and inflow to Hirakud dam, with reanalysis data as regressor and observed streamflow as regressand. This relationship is used to model the future streamflow using GCM output. GCM developed by Center for Climate System Research/ National Institute for Environmental Studies (CCSR/NIES), Japan, with B2 scenario is used for projection of future streamflow. The grid size of the GCM is 5.5° latitude × 5.625° longitude. The monthly output for B2 scenario is extracted for CCSR–NIES GCM for the region of interest covering all the NCEP grid points extending from 13.8445°N to 30.4576°N in latitude and 78.7500°E to 95.6250°E in longitude from IPCC data distribution center (http://www.mad.zmaw.de/IPCC_DDC/html/ddc_gcmdata.html).

Standardization [54] is used prior to statistical downscaling to reduce systematic biases in the mean and variances of GCM outputs relative to the observations or NCEP/NCAR data. The procedure typically involves subtraction of mean and division by standard deviation of the predictor variable for a predefined baseline period for both NCEP/NCAR and GCM output. The period 1961–1990 is used as a base-line because it is of sufficient duration to establish a reliable climatology, yet not too long, nor too contemporary to include a strong global change signal [54]. A major limitation of standardization is that it considers the bias in only mean and variance. There is a possibility that the reanalysis data and GCM output may deviate from normal distribution, and there may exist bias in other statistical parameters. For Mahanadi river basin, four predictor variables (MSLP, 2 m surface air temperature, specific humidity, and 500hPa geopotential height) at 25 NCEP grid points with a dimensionality of 100, are used which are highly correlated with each other. Principal Component Analysis (PCA) [21,13] is performed to transform the set of correlated N-dimensional predictors ($N = 100$) into another set of N-dimensional uncorrelated vectors (called principal components) by linear combination, such that most of the information content of the original data set is stored in the first few dimensions of the new set. In the present study, it is observed that first 10 Principal Components (PCs) represent 98.1% of the information content (or variability) of the original predictors, and therefore they are used in downscaling. The advantage of PCA is that it reduces the dimensionality of the predictors and at the same time there is no redundant information and correlation among the predictors, which may lead to multicollinearity. Fuzzy clustering is used to classify the principal components into classes or clusters. Fuzzy clustering assigns membership values of the classes to various data points, and it is more generalized and useful to describe a point not by a crisp cluster, but by its membership values in all the clusters [35,16].

The important parameters required for fuzzy clustering algorithm are number of clusters (c) and fuzzification parameter (m). Fuzzification parameter controls the degree of fuzziness of the resulting classification, which is the degree of overlap between clusters. The minimum value of m is 1 which implies hard clustering. Number of clusters and fuzzification parameter are determined from cluster validity indices like Fuzziness Performance Index (FPI) and Normalized Classification Entropy (NCE) [36]. FPI estimates the degree of fuzziness generated by a specified number of classes and is given by

$$\mathrm{FPI} = 1 - \frac{cF - 1}{c - 1} \tag{1}$$

where

$$F = \frac{1}{T}\sum_{i=1}^{c}\sum_{t=1}^{T}(\mu_{it})^2 \tag{2}$$

μ_{it} is the membership in cluster i of the principal components in month t. NCE estimates the degree of disorganization created by a specified number of classes and given as

$$\mathrm{NCE} = \frac{H}{\log c} \tag{3}$$

where

$$H = \frac{1}{T}\sum_{i=1}^{c}\sum_{t=1}^{T} -\mu_{it} \times \log(\mu_{it}) \tag{4}$$

The optimum number of classes/clusters is established on the basis of minimizing these two measures given by Eqs. (1)–(3). The FPI and NCE attain their minimum values when the number of clusters is 3 for almost all cases with different m values. The value of FPI should be chosen in such a way that the resulting clustering is neither too fuzzy

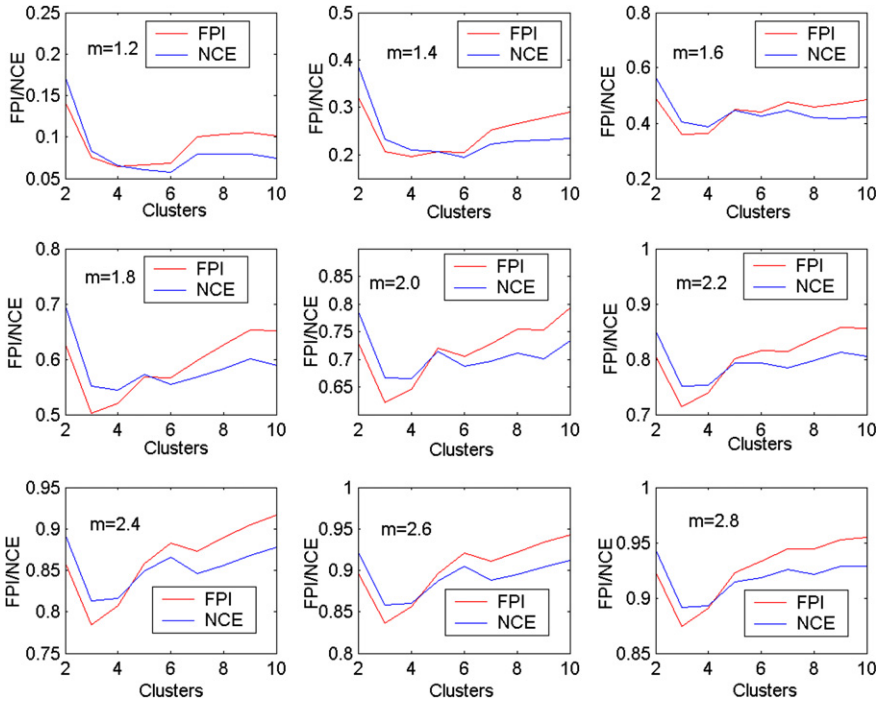

Fig. 3. Cluster validity test.

nor too hard. The clustering becomes non-fuzzy when FPI = 0 and fully fuzzy when FPI = 1 [16]. Guler and Thyne [16] have recommended an FPI value of 0.25 for the purpose of selection of number of clusters and fuzzification parameter in fuzzy clustering. In this work, FPI and NCE are plotted with number of clusters c, for different values of fuzzification parameter, m (Fig. 3). It is found that FPI value of almost 0.25 is achieved for $m = 1.4$ and $c = 3$. These values are used for fuzzy clustering. The sum of the membership of a data point in three clusters is equal to 1 and thus the membership of only two clusters will automatically fix the other and are sufficient to be used as an input to vector machine. Thus, the number of input variables used in the SVM and RVM is 12 (10 principal components along with two memberships). The following section presents the Support Vector (SV) regression used for statistical downscaling with training and testing.

3. Training and testing with support vector machine

The foundations of Support Vector Machine (SVM) have been developed by Vapnik [46] and are gaining popularity due to many attractive features, and promising empirical performance. The formulation embodies Structural Risk Minimization (SRM) principle, which has been proved to be superior [17] to traditional Empirical Risk Minimization (ERM) principle, employed by conventional neural networks. SRM minimizes an upper bound on the expected risk, as opposed to ERM that minimizes the error on the training data. This difference equips SVM with a greater ability to generalize, which is the goal of statistical learning. A brief introduction to statistical learning with the concept of SRM may be found in Vapnik [47] and Dibike et al. [12].

Given a training data $\{(x_1, y_1), \ldots, (x_l, y_l), X \in \Re^n, Y \in \Re\}$, the SV regression equation can be given by [37]

$$y = f(x) = \sum_{i=1}^{l} w_i \times K(x_i, x) + b \tag{5}$$

where $K(x_i,x)$ and w_i are the kernel functions and the corresponding weights used in the SV regression. b is a constant known as bias. The ith input x_i for training is called support vector if $w_i \neq 0$ for that particular i. Naturally, in Eq. (5) the inputs other than support vectors will be vanished. The architecture of an SVM is presented in Fig. 4. The loss function considered for SVM is an ε-insensitive loss function (Fig. 5) described as

$$|\xi|_\varepsilon = |y - f(x)|_\varepsilon = \begin{cases} 0 & \text{if } |y - f(x)| \leqslant \varepsilon \\ |y - f(x)| - \varepsilon & \text{otherwise} \end{cases} \tag{6}$$

The methodology for computation of weights and bias is presented in Appendix 1. Gaussian, Laplacian and heavy tailed Radial Basis Functions (RBF) are used as kernel in the present study. Details of these kernel functions are presented in Appendix 2. Selection of a suitable RBF is an important task in SVM as it has a high sensitivity on model performance [10]. All the above mentioned kernels are used in the present analysis to compare and select the SVM regression model with the best kernel for downscaling purpose.

SVM regression models with all the kernels are trained to determine the relationship between NCEP/NCAR out-

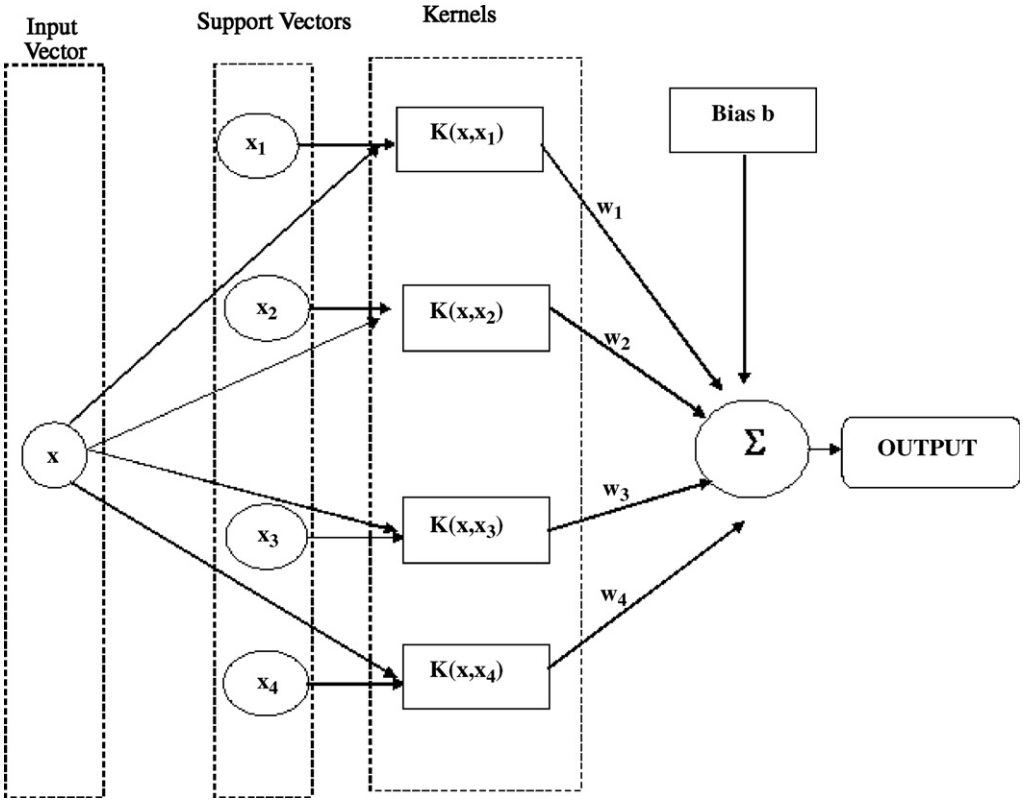

Fig. 4. Architecture of an SVM.

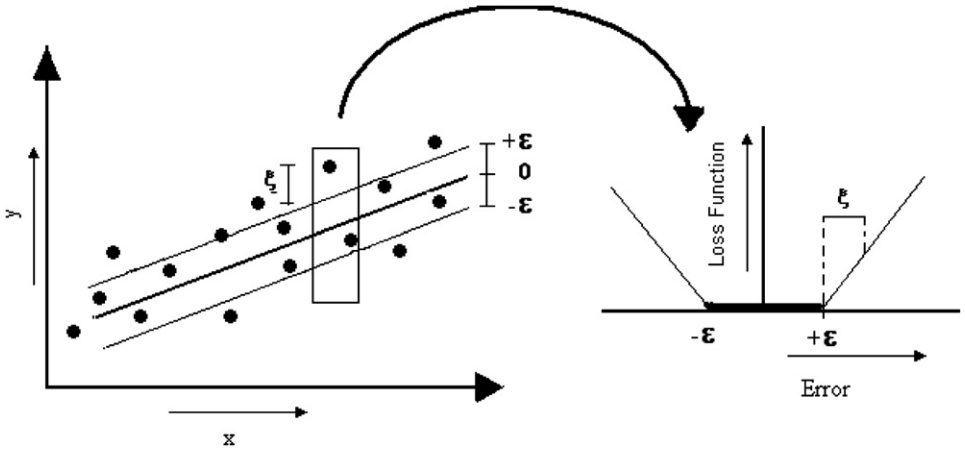

Fig. 5. ε-Insensitive loss function.

put of large scale climate variables and Mahanadi monsoon streamflow where the principal components, membership in fuzzy clusters and seasonal components are used as input. For training and testing K-fold cross validation ($K = 10$) procedure is used. According to this methodology, the training set is partitioned into K disjoint sets. The model is trained, for a chosen kernel, on all the subsets except for one, which is left for testing. The procedure is repeated for a total of K trials, each time using a different subset for testing. The average R values (Correlation coefficient between observed and predicted streamflow) over all the training and testing are considered as the goodness of the model fit to assess the performance of the model. The model resulting in the highest R value for testing is considered as the best model with a minimum chance of overfitting. Before training, both the input and output data are standardized to [0, 1]. The parameters of SVM, C, σ and ε (Appendices 1 and 2) is considered as 100, 1 and 0.01 for training of the standardized data. The results obtained for the developed SVM model with different kernels are presented in Table 1, which shows that SVM with heavy tailed RBF kernel produces maximum R value in testing

Table 1
Results obtained from training and testing of SVM based downscaling model

Kernel used	R value for training	R value for testing	Number of support vectors (% of training data set)
Gaussian RBF	0.9975	0.5572	96.20
Laplacian RBF	0.9992	0.6133	93.61
Heavy tailed RBF	0.9995	0.6173	93.24

involving minimum support vectors. Based on this criterion the SVM with heavy tailed RBF kernel is selected as the best SVM regression model. It is observed that for all the kernels the model suffers from severe overfitting and therefore it is not acceptable. It should be noted that rigorous post-modeling sensitivity analysis by perturbing all the three parameters, C, σ and ε may reduce the overfitting, but this is computationally expensive. To improve the performance by reducing the overfitting, RVM is used for statistical downscaling. The following section presents the mathematical development of RVM.

4. Relevance vector machine

Despite of excellent modeling performance of SVM, it has some practical and significant drawbacks. They are [44]:

- Although relatively sparse, SVMs make unnecessarily liberal use of basis functions since the number of support vectors required typically grows linearly with the size of the training set. Some form of post-processing is often required to reduce computational complexity [8].
- Predictions are not probabilistic. In regression the SVM outputs a point estimate, whereas the conditional distribution of target given input ($p(t|x)$) is desired.
- There is no straightforward method to estimate C and ϵ. Sometimes cross validation is used to estimate them which is wasteful for both data and computation.
- The kernel function $K(x, x_i)$ must satisfy Mercer's condition.

Relevance Vector Machine developed by Tipping [44] is a Bayesian treatment of Eq. (5) which does not suffer from any of the limitations stated above. In RVM, a fully probabilistic framework is adopted and introduced a priori over the model weights governed by a set of hyperparameters, associated with weights, whose most probable values are iteratively estimated from the data. Sparsity is achieved because in practice the posterior distributions of many of the weights are sharply (indeed infinitely) peaked around zero. The remaining training vectors associated with non-zero weights are termed as relevance vectors. The most compelling feature of the RVM is that, while capable of generalization performance comparable to an equivalent SVM, it typically utilizes dramatically fewer kernel functions. Following Tipping [44], the mathematical background of RVM is presented here.

RVMs have identical functional form as SVMs (Eq. (5)), but use kernel terms that correspond to fixed nonlinear basis function [44,28,27]. Seeking to forecast y for given x according to $y = f(x) + \epsilon_n$, involving weights $\mathbf{w} = (w_0, w_1, \ldots, w_l)^{\mathrm{T}}$, where $\epsilon_n \sim N(0, \sigma^2_{\epsilon_n})$, the likelihood of the complete data set can be written as

$$p(\mathbf{y}|\mathbf{w}, \sigma^2_{\epsilon_n}) = (2\pi\sigma^2_{\epsilon_n})^{-l/2} \exp\left\{-\frac{1}{2\sigma^2_{\epsilon_n}}\|\mathbf{y} - \Phi_{\mathbf{w}}\|^2\right\} \tag{7}$$

where $\Phi(x_i) = [1, K(x_i, x1), K(x_i, x_2), \ldots, K(x_i, x_l)]^{\mathrm{T}}$. Maximum likelihood estimation of $\mathbf{w}$ and $\sigma^2_{\epsilon_n}$ often results in severe overfitting. Tipping [44] suggested imposition of some prior constraints on the parameters, $\mathbf{w}$, by adding a complex penalty to the likelihood or the error function. This is a prior information that controls the generalization ability of the learning system. Typically, higher-level hyperparameters are used to constrain an explicit zero-mean Gaussian prior probability distribution over the weights, $\mathbf{w}$:

$$p(\mathbf{w}|\alpha) = \prod_{i=0}^{l} N(w_i|0, \alpha_i^{-1}) \tag{8}$$

where α is a hyperparameter vector that controls how far from zero each weight is allowed to deviate. For completion of hierarchical prior specifications, hyperpriors over α and the noise variance, $\sigma^2_{\epsilon_n}$, are defined. Consequently, using Bayes rule, the posterior overall unknowns could be computed given the defined noninformative prior distributions:

$$p(\mathbf{w}, \alpha, \sigma^2_{\epsilon_n}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{w}, \alpha, \sigma^2_{\epsilon_n}) \cdot p(\mathbf{w}, \alpha, \sigma^2_{\epsilon_n})}{\int p(\mathbf{y}|\mathbf{w}, \alpha, \sigma^2_{\epsilon_n}) p(\mathbf{w}, \alpha, \sigma^2_{\epsilon_n}) \mathrm{d}\mathbf{w}\, \mathrm{d}\alpha\, \mathrm{d}\sigma^2_{\epsilon_n}} \tag{9}$$

Computation of $p(\mathbf{w}, \alpha, \sigma^2_{\epsilon_n}|\mathbf{y})$ in Eq. (9) is not possible directly as the integral in the right-hand side can not be performed. Instead the posterior can be decomposed as

$$p(\mathbf{w}, \alpha, \sigma^2_{\epsilon_n}|\mathbf{y}) = p(\mathbf{w}|\mathbf{y}, \alpha, \sigma^2_{\epsilon_n}) p(\alpha, \sigma^2_{\epsilon_n}|\mathbf{y}) \tag{10}$$

The posterior distribution of the weight can be given by

$$p(\mathbf{w}|\mathbf{y}, \alpha, \sigma^2_{\epsilon_n}) = \frac{p(\mathbf{y}|\mathbf{w}, \sigma^2_{\epsilon_n}) \cdot p(\mathbf{w}|\alpha)}{p(\mathbf{y}|\alpha, \sigma^2_{\epsilon_n})} = (2\pi)^{-l/2}|\Sigma|^{-1/2} \times \exp\left\{-\frac{1}{2}(\mathbf{w} - \mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{w} - \mu)\right\} \tag{11}$$

where the posterior covariance and mean are respectively:

$$\Sigma = (\sigma^{-2}_{\epsilon_n}\Phi^{\mathrm{T}}\Phi + \mathbf{A})^{-1} \tag{12}$$

$$\mu = \sigma^{-2}_{\epsilon_n}\Sigma\Phi^{\mathrm{T}}\mathbf{y} \tag{13}$$

with $A = \mathrm{diag}(\alpha_0, \ldots, \alpha_l)$. Therefore, machine learning becomes a search for the hyperparameter posterior most probable, i.e., the maximization of $p(\alpha, \sigma^2_{\epsilon_n}|\mathbf{y}) \propto p(y|\alpha, \sigma^2_{\epsilon_n})$

$p(\alpha)p(\sigma^2_{\epsilon_n})$ with respect to α and $\sigma^2_{\epsilon_n}$. For uniform hyperpriors, it is required to maximize the term $p(\mathbf{y}|\alpha, \sigma^2_{\epsilon_n})$, which is computable and given by

$$p(\mathbf{y}|\alpha, \sigma^2_{\epsilon_n}) = \int p(\mathbf{y}|\mathbf{w}, \sigma^2_{\epsilon_n})p(\mathbf{w}|\alpha)\mathrm{d}\mathbf{w} = (2\pi)^{-l/2}|\sigma^2_{\epsilon_n} I + \Phi A^{-1}\Phi^{\mathrm{T}}|^{1/2} \times \exp\{-\frac{1}{2}\mathbf{y}^{\mathrm{T}}(\sigma^2_{\epsilon_n} I + \Phi A^{-1}\Phi^{\mathrm{T}})^{-1}\mathbf{y}\} \quad (14)$$

Tipping [44] contended that all the evidence from several experiments suggests that this predictive approximation is very effective. Bayesian models refer to Eq. (9) as the marginal likelihood, and its maximization is known as the type II-maximum likelihood method [6,48]. As argued by Tipping [44], MacKay [29] refers to this term as the evidence for hyperparameter and its maximization as the evidence procedure. Hyperparameter estimation is typically carried out with an iterative formula such as a gradient ascent on the objective function [44,29].

At convergence of the hyperparameter estimation procedure, predictions can be made based on the posterior distribution over the weights, conditioned on the maximized most probable values of α and $\sigma^2_{\epsilon_n}$, α_{MP} and σ^2_{MP} respectively. The predictive distribution for a given x_* can be computed using Eq. (11):

$$p(y_*|\mathbf{y}, \alpha_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}}) = \int p(y_*|\mathbf{w}, \sigma^2_{\mathrm{MP}})p(\mathbf{w}|\mathbf{y}, \alpha_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}})\mathrm{d}\mathbf{w} \quad (15)$$

Since both terms in the integrand are Gaussian, this can be readily computed, giving

$$p(y_*|\mathbf{y}, \alpha_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}}) = N(y_*|t_*, \sigma^2_*) \quad (16)$$

with

$$t_* = \mu^{\mathrm{T}}\Phi(x_*), \quad (17)$$

$$\sigma^2_* = \sigma^2_{\mathrm{MP}} + \Phi(x_*)^{\mathrm{T}}\Sigma\Phi(x_*) \quad (18)$$

The outcome of the optimization involved in RVM (i.e. maximization of $p(\mathbf{y}|\alpha, \sigma^2_{\epsilon_n})$), is that many elements of α go to infinity such that w will have only a few nonzero weights that will be considered as relevant vectors. The relevant vectors (RVs) can be viewed as counterparts to support vectors (SVs) in SVMs; therefore, the resulting model enjoys the properties of SVMs (i.e., sparsity and generalization) and, in addition, provides estimates of uncertainty bounds in the predictions they make [28].

4.1. Training and testing with RVM

Principal components and fuzzy cluster memberships derived from NCEP/NCAR reanalysis data are used as input to RVM. Similar to SVM, Gaussian RBF, Laplacian RBF and Heavy tailed RBF are used as kernels in the RVM regression model for downscaling with K-fold cross validation. The results obtained from training and testing are presented, with $\sigma = 1$, in Table 2. Compared to SVM, RVM involves very few relevant vectors for the regression with all the kernels and thus minimizing the possibility of overtraining as well as computational time. This is reflected in the differences between R values for training and testing with all the kernels. A comparatively small difference between the R value of training and testing shows the reduction of overtraining which is not achieved by SVM. Among the RVM kernels the model with heavy tailed RBF shows the highest R value for testing among all the RVM models and thus selected as the best model for downscaling. The selection of the width of the kernel is one of the major criterion in selecting the appropriate model. The kernel width can not be computed with the Bayesian treatment of RVM and therefore a post-modeling sensitivity analysis is required to compute kernel width that results in minimum overfitting. Sensitivity analysis of the training and testing R values and the number of RVs involved in the model is carried out, with variation in the kernel width, and presented in Fig. 6. As RVM involves only kernel width as a parameter, the computational effort of post-modeling sensitivity analysis is significantly less compared to SVM. It is observed that the testing R value achieved its maximum at a kernel width of 1.9, involving minimum number of RVs. The training and testing R values are obtained as 0.7745 and 0.7256, respectively, using only 7.41% of the data set as relevant vectors. Therefore RVM using heavy tailed RBF with the width of 1.9 is used for statistical downscaling in the present study. After the selection of model the whole data set is trained using RVM based regression with heavy tailed RBF as the kernel. The overall R value is obtained as 0.8226. The observed and predicted monsoon streamflow from June 1961 to August 1990 with scatter plot are presented in Fig. 7. It is clear that even RVM is not able to mimic the extreme rainfall observed in the record. Possibly this could be because regression based statistical downscaling models often cannot explain entire variance of the downscaled variable [54,42]. The goodness of fit of the model is also tested with Nash–Sutcliffe coefficient [32], which has been recommended by ASCE Task Committee on definition of Criteria for evaluation of watershed models of the watershed management committee, Irrigation and Drainage Division [1]. The Nash–Sutcliffe coefficient (E) is given by

Table 2
Results obtained from training and testing of RVM based downscaling model

Kernel used	R value for training	R value for testing	Number of relevant vectors (% of training data set)
Gaussian RBF	0.9423	0.6019	71.30
Laplacian RBF	0.8417	0.6418	25.56
Heavy tailed RBF	0.7937	0.6998	8.06

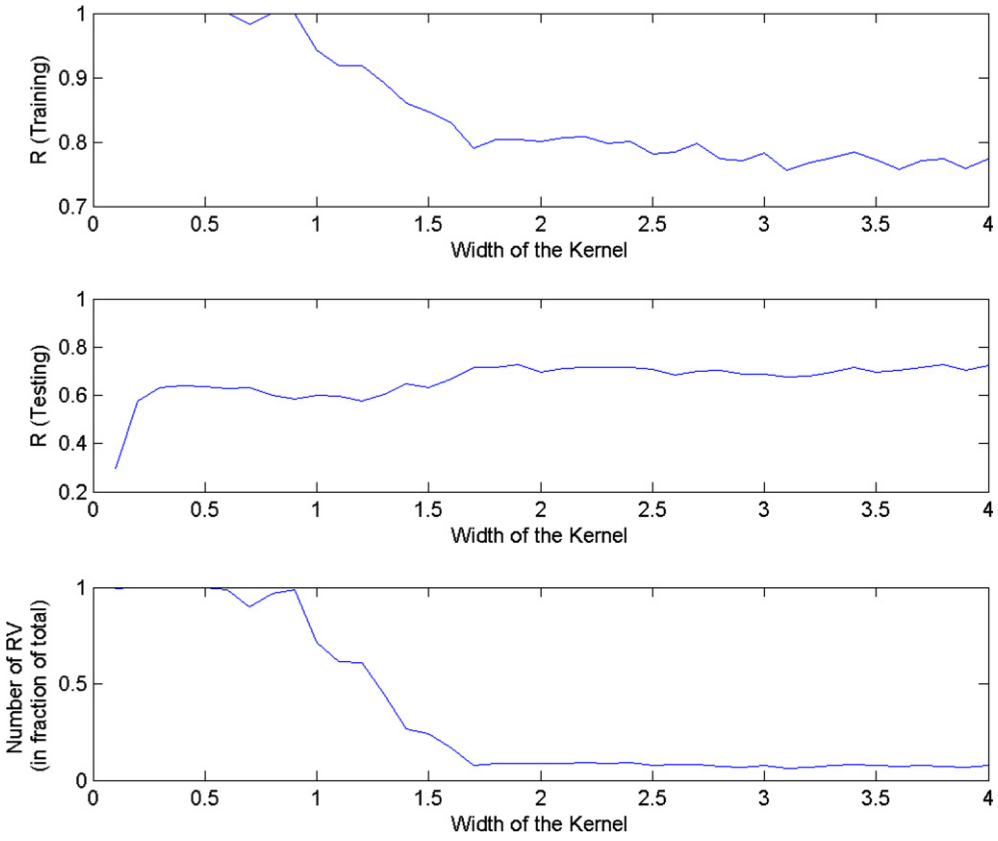

Fig. 6. Sensitivity analysis with the width of kernel in RVM.

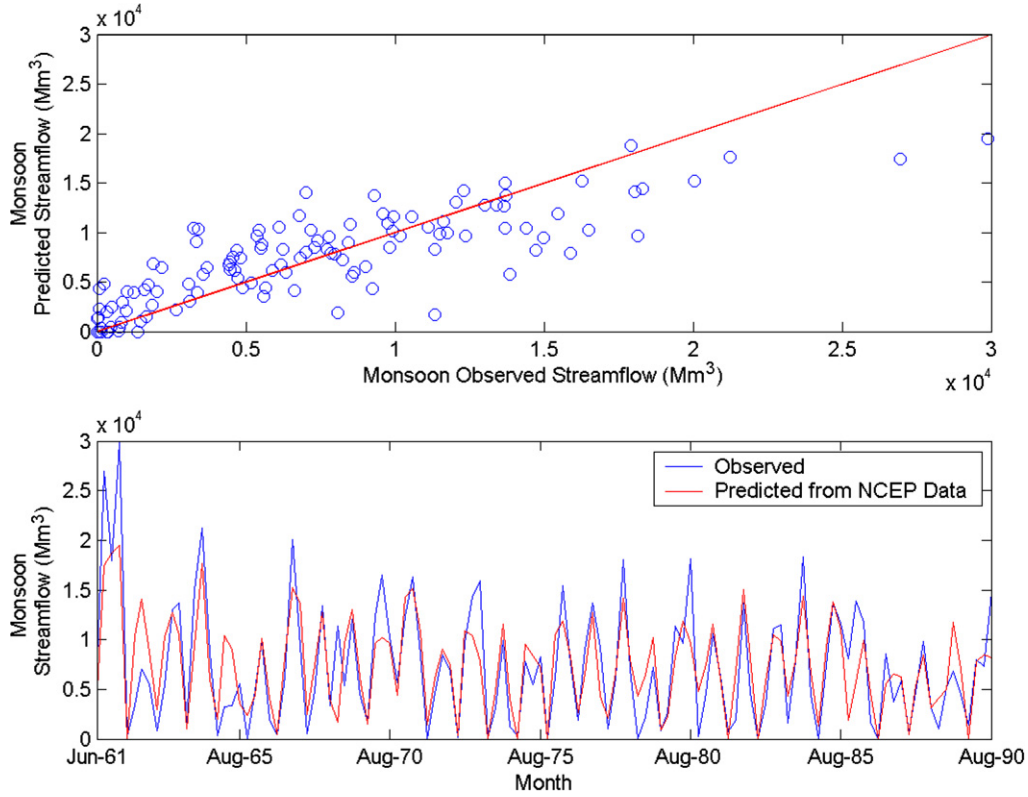

Fig. 7. Observed and RVM modeled monthly streamflow of Mahanadi River.

$$E = 1 - \frac{\sum_t (Q_{ot} - Q_{pt})^2}{\sum_t (Q_{ot} - \overline{Q_o})^2} \quad (19)$$

where Q_{ot} and Q_{pt} are the observed and predicted streamflow in time t, and $\overline{Q_o}$ is the mean observed streamflow. Nash–Sutcliffe coefficient can vary from 0 to 1 with 0 indicating that the model predics no better than the average of the observed data, and 1 indicating a perfect fit. It is obtained as 0.67 for the present model which is satisfactory. Wetterhall et al. [49] have tested the long term seasonal mean, and standard deviation for verification of a downscaling model. In the present analysis also, similar test has been performed. The long term mean and standard deviation of observed streamflow are 7332.0 Mm^3 and 5995.6 Mm^3 and those of predicted streamflow are 7384.1 Mm^3 and 4607.6 Mm^3, which shows a good match in mean but difference in standard deviation. This may be because the regression based statistical downscaling models often cannot explain entire variance of the downscaled variable [54] and therefore the present model can not mimic the high streamflow in 1961. Other limitation of the method is that assuming constant error variance (homoscedasity) may prove to be a limitation when streamflow is the response variable as it is often observed that streamflow error variance is related to the magnitude of the flow, and, the hierarchical structure cannot accommodate Markovian dependence in the flows easily. After the verification, the RVM regression model is used for modeling of future streamflow time series from the predictor variables as projected by GCM developed by CCSR/NIES with B2 scenarios.

5. Future streamflow projection

GCM developed by Center for Climate System Research/National Institute for Environmental Studies (CCSR/NIES), Japan, with B2 scenario is used for projection of future streamflow. The grid size of the GCM is 5.5° latitude × 5.625° longitude. The monthly output for B2 scenario is extracted for CCSR/NIES GCM for the region of interest covering all the NCEP grid points extending from 13.8445°N to 30.4576°N in latitude and 78.7500°E to 95.6250°E in longitude from IPCC data distribution center. GCM grid points do not match with NCEP grid points and thus interpolation is required to obtain the GCM output at NCEP grid points. Interpolation is performed with a linear inverse square procedure using spherical distances [57]. The predictor variables for CCSR/NIES GCM are then interpolated to the 25 NCEP grid points. Using the principal directions or eigen vectors obtained from PCA of NCEP data, principal components are obtained for the GCM output. The membership of the principal components of GCM output in each of the fuzzy clusters are then computed using the cluster centers obtained from fuzzy clustering. Principal components and cluster membership of GCM output are then used in the developed RVM regression model to project the monsoon streamflow of Mahanadi for future.

For validation purpose, the monsoon streamflow is also computed for the base-line period of years 1961–1990 with the GCM output. The CDFs obtained from NCEP data, GCM output and the observed data, using Weibull's probability plotting formula, are presented in Fig. 8(a). Although the CDF obtained from GCM matches quite

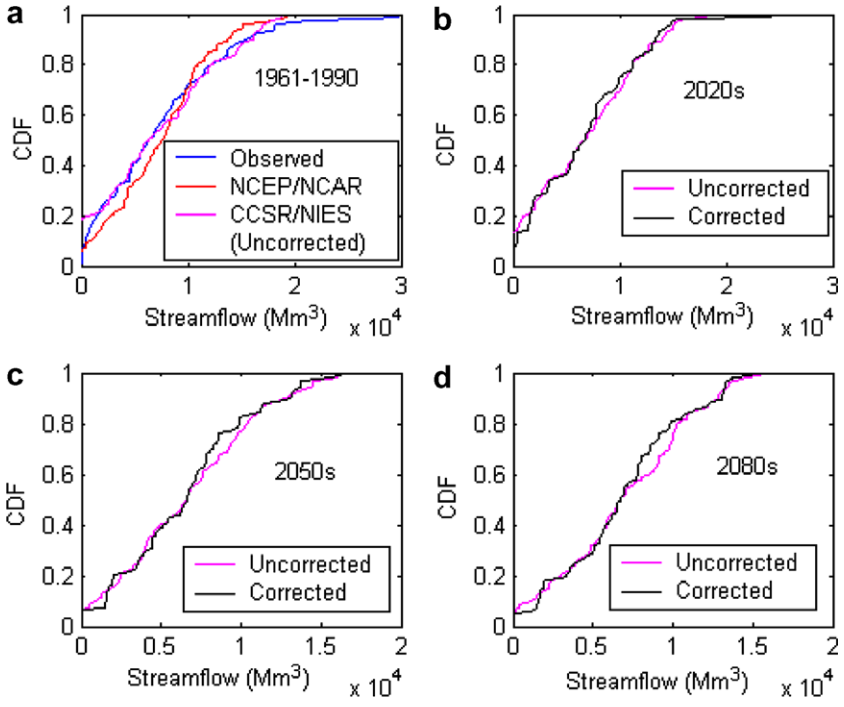

Fig. 8. Projected future Streamflow for CCSR/NIES GCM with B2 scenario using RVM based downscaling.

well, there is considerable bias near zero flow values and at the extreme cases. This is because standardisation may reduce the bias in the mean and variance of the predictor variable but it is much harder to accommodate the biases in large scale patterns of atmospheric circulation in GCMs (e.g. shifts in the dominant storm track relative to observed data) or unrealistic inter-variable relationships [50]. Moreover, regression based statistical downscaling models often cannot explain entire variance of the downscaled variable, which is also reflected in terms of bias near zero flow and high flow conditions. While modeling monsoon streamflow such biases should be taken care otherwise it will propagate in the computation of future seasons [14]. To remove such bias from a given downscaled output the following methodology is used:

- CDFs are obtained with the downscaled GCM generated and observed streamflow for the years 1961–1990 using Weibull's probability plotting position formula.
- For a given value of GCM generated streamflow (X_{GCM}), the value of CDF (CDF_{GCM}) is computed.
- Corresponding to CDF_{GCM} the observed streamflow value is obtained from the CDF of observed data.
- The GCM generated streamflow is replaced by the observed data, thus computed, having the same CDF value.
- The CDFs of GCM generated and observed streamflow, obtained for the years 1961–1990, act as reference, and based on them the correction is applied to the streamflow values obtained from GCM for future.

A major drawback of the method described above is that, if the future GCM streamflow is out of the range of historical GCM streamflow, the methodology of bias correction with Weibull's plotting position will fail. If such cases appear, then different parametric probability distribution (with upper bound of random variable as ∞) can be fitted to the observed GCM streamflow and the best pdf can be selected among them with Akaike Information Criteria (AIC) or χ^2-test. As the new range is now extended to ∞, it is possible to perform the quantile transformation even if the future GCM streamflow is out of the range of historical GCM streamflow. In the present case, the future GCM streamflows are all within the range of observed GCM streamflow and therefore the bias is corrected with Weibull's plotting position and quantile transformation. The long term mean and standard deviation of observed streamflow are 7332.0 Mm^3 and 5995.6 Mm^3 and those of GCM projected streamflow before bias correction were 7194.2 Mm^3 and 5607.2 Mm^3. After bias correction mean and standard deviation of GCM projected streamflow are 7331.7 Mm^3 and 6009.4 Mm^3, respectively, which shows bias has been significantly reduced. The CDFs projected future streamflow is plotted for standard 30 year time slices 2020 s, 2050 s and 2080 s in Fig. 8b–d, which clearly shows a decrease in the high flows of the monsoon season in Mahanadi. The occurrence of extreme high flow events will reduce significantly and therefore there is a decreasing trend in the monthly peak flow. The projection of CCSR/NIES GCM with B2 scenario presents a favorable condition for Hirakud dam in future for flood control operation. Earlier study [34] on Mahanadi river also revealed decrease in monsoon streamflow for the historic period with an increasing trend in surface temperature. It is concluded in that study, that due to increase in temperature, the water yields in the river is adversely affected. Following the study, it can be inferred that one of the probable reason of such decreasing trend in streamflow may be significant increase in temperature due to climate warming.

Analysis of instrumental climate data revealed that the mean surface temperature over India has warmed at a rate of about 0.4 °C. per century [20,39,40], which is statistically significant. The increasing trend of temperature in Mahanadi river basin due to climate change is more severe. Rao [33] found that the surface air temperature over this basin is increasing at a rate of 1.1 °C per century, which is more than double of that of entire India. Fig. 9a presents box plots of the temperature projected by CCSR/NIES with B2 scenario for historic base period (1961–1990), 2020s, 2050s and 2080s. The box plot presents the median, upper and lower quartiles and the outliers. The middle line of the box gives the median whereas the upper and lower edge give the 75 percentile and 25 percentile of the data set, respectively. A significant increasing trend is observed in the surface air temperature. The corresponding box plots for the monsoon streamflow are presented in Fig. 9b. The result shows that although there is no significant change in the median of the monsoon flow, the occurrence of high flows will reduce significantly because of high surface warming and therefore there is a decreasing trend in the monthly peak flow. The reason may be the insensitivity of climatic variables towards low flow because of significant ground water component and therefore only the effect on high flow, which is of interest, is reflected in the results. It is worth mentioning that the projected streamflow presented, is due to a single GCM using a single scenario and it is widely acknowledged that disagreements between different GCMs over regional climate changes represents a significant source of uncertainty [55,14]. Therefore overreliance on a single GCM could lead to inappropriate planning and adaptation responses. Thus future decision making should incorporate all the GCMs with scenarios to model the underlying GCM and scenario uncertainty.

6. Concluding remarks

Downscaling of GCM output to monsoon streamflow is performed using SVM and RVM in the present analysis. Standardisation is performed to remove the biases present in the mean and the variances of the predictor variables. PCA and fuzzy clustering are performed prior to training to improve the model performance. As the model is a

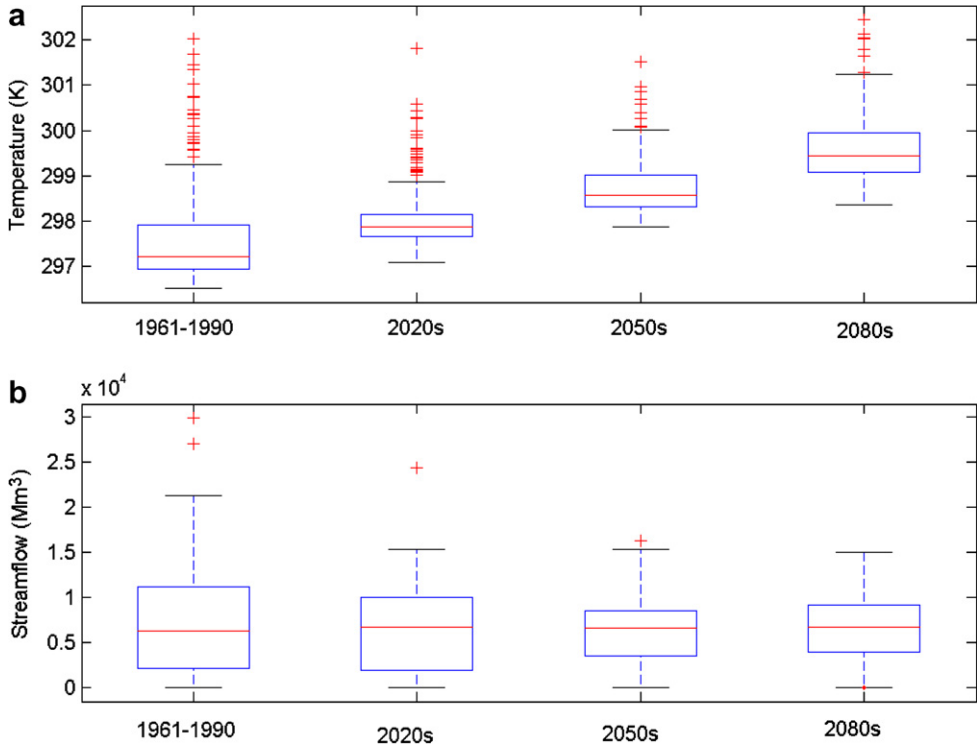

Fig. 9. (a) Box plot of projected temperature of the case-study area by CCSR/NIES GCM with B2 Scenario and (b) box plot of downscaled streamflow from CCSR/NIES GCM output with B2 Scenario.

combination of classification and regression, it can be categorized into a hybrid model of weather typing and transfer function. It has been observed that RVM not only involves probabilistic reasoning but also outperforms SVM for regression based statistical downscaling in terms of goodness of fit. RVM involves fewer number of relevant vectors and the chance of overfitting is less than that of SVM. The model developed in the present study is capable of producing a satisfactory value of goodness of fit in terms of R value and Nash–Sutcliffe coefficient. However, from Fig. 7 it is found that even RVM is not able to mimic the extreme streamflow observed in the record. Possibly this could be because regression based statistical downscaling models often cannot explain entire variance of the downscaled variable [54]. Bias resulting from the drawback is corrected at the end of downscaling. The GCM CCSR/NIES with B2 scenario projects a decreasing trend in future monsoon streamflow of Mahanadi. In Rao [34], a decreasing trend in the streamflow of Mahandi River with an increasing trend in surface temperature is observed, and it is concluded in that study that due to increase in temperature, the water yields in the river is adversely affected. Following the study, it can be inferred that one of the possible reasons for such a decrease in Mahanadi River streamflow may be increase in surface temperature. Such a decrease in streamflow may cause a critical situation for Hirakud dam in meeting the future irrigation and power demand. The methodology developed can be used to project the streamflow for other GCMs and scenarios also and there is a possibility of mismatch in the projections resulting GCM and scenario uncertainty. Modeling of such uncertainty is necessary for future decision making. The methodology presented, does not limit its usefulness only for modeling streamflow. It is adaptable and can be used to model any other hydrologic variable, viz. precipitation, evaporation, etc. to assess the impact of climate change on hydrology.

Appendix 1. Support vector regression

The basic concept of SV regression is discussed in the present section first with a linear model and then it is extended to a nonlinear model using Kernels. Given a training data $\{(x_1, y_1), \ldots, (x_l, y_l), X \in \Re^n, Y \in \Re\}$, the SV regression equation can be given by [37]

$$f(x) = \langle w, x \rangle + b, \quad w \in X, \quad b \in \Re \tag{20}$$

where, $\langle \cdot, \cdot \rangle$ denoted the dot product in X. The objective of SVM regression is to find the function $f(x)$ with minimum value of loss function and at the same time is as flat as possible [38]. Flatness mathematically denotes the smaller value of w and one way to ensure this is to minimize the norm, i.e. $\|w\|^2 = \langle w, w \rangle$. Thus the model can be expressed as the following convex optimization problem.

$$\text{Minimize } \frac{1}{2}\|w\|^2 + C\left(\sum_{i}^{l}\xi_i^* + \sum_{i=1}^{l}\xi_i\right) \tag{21}$$

$$\text{subject to } y_i - \langle w, x\rangle - b \leqslant \varepsilon + \xi_i \tag{22}$$

$$\langle w, x\rangle + b - y_i \leqslant \varepsilon + \xi_i^* \tag{23}$$

$$\xi_i, \xi_i^* \geqslant 0 \tag{24}$$

where C is a pre-specified value which determines the trade-off between the flatness of $f(x)$ and the amount up to which deviations larger than ε is tolerated (ξ_i and ξ_i^*), which correspond to ε-insensitive loss function as presented in Eq. (6). The optimization model presented in Eqs. (21)–(24) can be solved using Lagrange multipliers. A dual set of variables are introduced to construct the Lagrange function, which is given below:

$$\begin{aligned} L = {} & \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{l}\xi_i^* + \sum_{i=1}^{l}\xi_i\right) - \sum_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ & - \sum_{i=1}^{l}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x\rangle + b) \\ & - \sum_{i=1}^{l}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x\rangle - b) \end{aligned} \tag{25}$$

where L is the Lagrangian and $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ are Lagrangian multipliers satisfying the positivity constraints.

$$\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geqslant 0 \tag{26}$$

From the saddle point condition, the partial derivatives of L with respect to the primal variables (w, b, ξ_i, ξ_i^*) have to vanish for optimality:

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0 \tag{27}$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i = 0 \tag{28}$$

$$\frac{\partial L}{\partial \xi_i^{(*)}} = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \tag{29}$$

where $\xi_i^{(*)}, \alpha_i^{(*)}, \eta_i^{(*)}$ refer to ξ_i and ξ_i^*; α_i and α_i^*; η_i and η_i^* respectively.

Substituting Eqs. (27)–(29) in Eq. (25) the following dual optimization problem is formulated.

$$\begin{aligned} \text{Maximize } & -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle - \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) \\ & + \sum_{i=1}^{l}y_i(\alpha_i - \alpha_i^*) \end{aligned} \tag{30}$$

$$\text{subject to } \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \tag{31}$$

$$\alpha_i, \alpha_i^* \in [0, C]. \tag{32}$$

Eq. (28) can be re-written as

$$w = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i \tag{33}$$

and, thus from Eq. (20):

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b \tag{34}$$

This is called the Support Vector Expansion for linear model which is used in SV regression. b can be computed by using Karush Kuhn Tucker (KKT) condition [38].

For most of the hydrologic analysis linear regression is not appropriate and thus a nonlinear mapping using kernel K is used to map the data into a higher dimensional feature space, where, with the kernel, linear analysis is performed. Using the kernel, the regression equation (Eq. (34)) can be modified to (Eq. (5))

Appendix 2. Kernel functions

Kernel functions are used in SVM for nonlinear mapping of the original data or input into a high dimensional feature space. Kernel function used in a SVM should follow Mercer's theorem, according to which it can be written that:

$$\int_{X\times X} K(x, x')f(x)f(x')\mathrm{d}x\,\mathrm{d}x' \geqslant 0 \quad \forall f \in L_2(X) \tag{35}$$

Some of the valid kernel functions satisfying the above mentioned condition are given below.

A. *Linear kernel*: The linear kernels are the simplest kernels used in SVM for linear regression. They can be given by

Homogeneous kernel:

$$K(x, x') = \langle x, x'\rangle \tag{36}$$

Non-homogeneous kernel:

$$K(x, x') = (\langle x, x'\rangle + 1) \tag{37}$$

The performance of SVM with linear kernel function, being similar to that of linear regression, is not capable of modeling complicated and nonlinear relationship between climatological variables and streamflow and therefore such kernels are not used in the present study.

B. *Gaussian Radial Basis Function*: Radial Basis Functions (RBFs) have received significant attention, most commonly with Gaussian form,

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \tag{38}$$

where σ is the width of Gaussian RBF kernel, giving an idea about the smoothness of the derived function. A large kernel width acts as a low-pass filter in frequency domain, attenuating higher order frequencies and thus resulting in a smooth function. Alternatively, RBF kernel with small kernel width retains most of the higher order frequencies leading to an approximation of a complex function by learning machine [38].

C. *Laplacian or Exponential Radial Basis Function*: Laplacian or Exponential RBF of the form,

$$K(x,x') = \exp\left(-\frac{\|x-x'\|}{2\sigma^2}\right) \quad (39)$$

produces a piecewise linear solution which can be attractive when discontinuities are acceptable.

D. *Heavy tailed or Sublinear Radial Basis Function*: Heavy tailed RBFs or Sublinear RBFs are introduced by Chapelle et al. [10] which sometimes outperform traditional Gaussian or RBFs [10]. They can be given by:

$$K(x,x') = \exp\left(-\frac{\|x-x'\|^{0.5}}{2\sigma^2}\right) \quad (40)$$

It is worth mentioning that a generalized RBF can be given by

$$K(x,x') = \exp\left(-\frac{\|x^a-x'^a\|^b}{2\sigma^2}\right) \quad (41)$$

and it will satisfy Mercer's condition if and only if $0 \leqslant b \leqslant 2$. The choice of a has no impact on Mercer's condition.

References

[1] ASCE Task Committee on definition of Criteria for Evaluation of Watershed Models of the Watershed Management Committee, Irrigation and Drainage Division. Criteria for Evaluation of Watershed Models. J Irrig Drain Eng 1993;119(3):429–442.

[2] Bardossy A, Plate EJ. Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence. J Hydrol 1991;122:33–47.

[3] Bardossy A, Duckstein L, Bogardi I. Fuzzy rule-based classification of atmospheric circulation patterns. Int J Climatol 1995;15:1087–97.

[5] Bates BC, Charles SP, Hughes JP. Stochastic downscaling of numerical climate model simulations. Environ Model Software 1998;13:325–31.

[6] Berger JO. Statistical decision theory and Bayesian analysis. 2nd ed. New York: Springer; 1985.

[7] Brown BG, Katz RW. Regional analysis of temperature extremes: spatial analog for climate change? J Clim 1995;8:108119.

[8] Burges CJC, Schölkopf B. Improving the accuracy and speed of support vector machines. In: Mozer MC, Jordan MI, Petsche T, editors. Advances in neural information processing systems, vol. 9. MIT Press; 1997. p. 375–81.

[9] Cannon AJ, Whitfield PH. Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models. J Hydrol 2002;259:136–51.

[10] Chapelle O, Haffner P, Vapnik VN. Support vector machines for histogram-based image classification. IEEE Trans Neural Networks 1999;10(5):1055–64.

[11] Crane RG, Hewitson BC. Doubled CO_2 precipitation changes for the Susquehanna Basin: down-scaling from the genesis general circulation model. Int J Climatol 1998;18:6576.

[12] Dibike YB, Velickov S, Solomatine D, Abbott MB. Model induction with support vector machines: introduction and applications. J Comput Civil Eng, ASCE 2001;15(3):208–16.

[13] Ghosh S, Mujumdar PP. Future rainfall scenario over Orissa with GCM projections by statistical downscaling. Curr Sci 2006;90(3):396–404.

[14] Ghosh S, Mujumdar PP. Nonparametric methods for modeling GCM and scenario uncertainty in drought assessment. Water Resour Res [in print MS No. 2006WR005351].

[15] Govindaraju RS. Bayesian learning and relevance vector machines for hydrologic applications. In: Proceedings of the second Indian international conference on artificial intelligence (IICAI-05), Pune, India; 2005.

[16] Güler C, Thyne GD. Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy c-means clustering. Water Resour Res 2004;40:W12503. doi:10.1029/2004WR003299.

[17] Gunn SR, Brown M, Bossley KM. Network performance assessment for neuro fuzzy data modelling. In: Liu X, Cohen P, Berthold M, editors. Intelligent data analysis. Lecture notes in computer science, vol. 1208. Berlin Heidelberg: Springer-Verlag; 1997. p. 313–23.

[18] Hewitson BC, Crane RG. Large-scale atmospheric controls on local precipitation in tropical Mexico. Geophys Res Lett 1992;19(18):18351838.

[19] Hewitson BC, Crane RG. Climate downscaling: techniques and application. Climate Res 1996;7:85–95.

[20] Hingane LS, Rupakumar K, Ramanamurthy BV. Long-term trends of surface air temperature in india. J Climatol 1985;5:521–8.

[21] Hughes JP, Lettenmaier DP, Guttorp P. A stochastic approach for assessing the effect of changes in synoptic circulation patterns on Gauge precipitation. Water Resour Res 1993;29(10):3303–15.

[22] Hughes JP, Guttorp P. A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. Water Resour Res 1994;30(5):1535–46.

[23] Hughes JP, Guttorp P, Charles SP. A non-homogeneous hidden Markov model for precipitation occurrence. Appl Statist 1999;48(1):15–30.

[24] Jones PD, Murphy JM, Noguer M. Simulation of climate change over europe using a nested regional-climate model, I: assessment of control climate, including sensitivity to location of lateral boundaries. QJR Meteorol Soc 1995;121:1413–49.

[25] Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, et al. The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorolog Soc 1996;77(3):437471.

[26] Katz RW, Parlange MB. Mixtures of stochastic processes: applications to statistical downscaling. Clim Res 1996;7:185–93.

[27] Khalil A, McKee M, Kemblowski M, Asefa T. Sparse Bayesian learning machine for real-time management of reservoir releases. Water Resour Res 2005;41:W11401. doi:10.1029/2004WR003891.

[28] Khalil AF, McKee M, Kemblowski M, Asefa T, Bastidas L. Multiobjective analysis of chaotic dynamic systems with sparse learning machines. Adv Water Resour 2006;29(1):72–88.

[29] MacKay D. Information theory, inference, and learning algorithms. Cambridge University Press; 2003.

[30] Mehrotra R, Sharma A. A nonparametric nonhomogeneous hidden Markov model for downscaling of multi-site daily rainfall occurrences. J Geophys Res – Atmos 2005;110(D16):16108.

[31] Murphy JM. An evaluation of statistical and dynamical techniques for downscaling local climate. J Clim 1999;12:2256–84.

[32] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models, Part-1: a discussion of principles. J Hydrol 1970;10:282290.

[33] Rao PG, Kumar KK. Climatic shifts over Mahanadi River Basin. Curr Sci 1992;63:192–6.

[34] Rao PG. Effect of climate change on streamflows in the Mahanadi River Basin, India. Water Int 1995;20:205–12.

[35] Ross TJ. Fuzzy logic with engineering applications. McGrawhill International Edition; 1997. p. 379–96.

[36] Roubens M. Fuzzy clustering algorithms and their cluster validity. Eur J Oper Res 1982;10:294–301.

[37] Smola AJ. Regression estimation with support vector learning machines. Munchen, Germany: Technische Universitat Munchen; 1996.

[38] Smola AJ, Schoelkopf B. A tutorial on support vector regression, NeuroCOLT2. Technical Report NC2-TR-1998-030, Royal Holloway College, University of London, UK; 1998.

[39] Srivastava HN, Devan BN, Dixit SK, Prakasarao GS, Singh SS, Rao RK. Decadal trends in climate over India. Mausam 1992;43:7–20.

[40] Thapliyal V, Kulshrestha UC. Climate changes and trends over India. Mausam 1991;42:333–8.

[41] Tripathi S, Srinivas VV. Downscaling of general circulation models to assess the impact of climate change on rainfall of India. In: Proceedings of international conference on hydrological perspectives for sustainable development (HYPESD – 2005), 23–25 February, IIT Roorkee, India; 2005, p. 509–17.
[42] Tripathi S, Srinivas VV, Nanjundiah RS. Downscaling of precipitation for climate change scenarios: a support vector machine approach. J Hydrol 2006;330:621–40.
[43] Suykens JAK. Nonlinear modelling and support vector machines. In: Proceedings of IEEE instrumentation and measurement technology conference, Budapest, Hungary; 2001. p. 287–94.
[44] Tipping ME. Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res 2001;1:211–44.
[45] Trigo RM, Palutikof JP. Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach. Clim Res 1999;13:45–59.
[46] Vapnik VN. The nature of statistical learning theory. New York: Springer Verlag; 1995.
[47] Vapnik VN. Statistical learning theory. New York: Wiley; 1998.
[48] Wahba G. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline-smoothing problem. Ann Stat 1985;4:1378402.
[49] Wetterhall F, Halldin S, Xu C. Statistical precipitation downscaling in Central Sweden with the analogue method. J Hydrol 2005;306:174–90.
[50] Wilby RL, Dawson CW. Using SDSM version 3.1 A decision support tool for the assessment of regional climate change impacts, User Manual; 2004.
[51] Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, et al. Statistical downscaling of general circulation model output: a comparison of methods. Water Resour Res 1998;34:2995–3008.
[52] Wilby RL, Hay LE, Leavesly GH. A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River Basin, Colorado. J Hydrol 1999;225:67–91.
[53] Wilby RL, Dawson CW, Barrow EM. SDSM – a decision support tool for the assessment of regional climate change impacts. Environ Modell Softw 2002;17(2):147–59.
[54] Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO. 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). <http://ipccddc.cru.uea.ac.uk/guidelines/StatDown_Guide.pdf>; 2004.
[55] Wilby RL, Harris I. A framework for assessing uncertainties in climate change impacts: low-flow scenarios for the River Thames, UK. Water Resour Res 2006;42:W02419. doi:10.1029/2005WR004065.
[56] Wilks DS. Multisite downscaling of daily precipitation with a stochastic weather generator. Clim Res 1999;11:125–36.
[57] Willmott CJ, Rowe CM, Philpot WD. Small-scale climate map: a sensitivity analysis of some common assumptions associated with the grid-point interpolation and contouring. Am Cartographer 1985;12:5–16.