INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **28**: 401–420 (2008)
Published online 18 June 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.1529

RMetS
ROYAL METEOROLOGICAL SOCIETY

Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine

Aavudai Anandhi,[a] V. V. Srinivas,[a] Ravi S. Nanjundiah[b] and D. Nagesh Kumar[a]*
[a] *Department of Civil Engineering, Indian Institute of Science, Bangalore, India*
[b] *Center for Atmospheric and Oceanic Sciences, Indian Institute of Science, Bangalore, India*

ABSTRACT: This paper presents a methodology to downscale monthly precipitation to river basin scale in Indian context for special report of emission scenarios (SRES) using Support Vector Machine (SVM). In the methodology presented, probable predictor variables are extracted from (1) the National Center for Environmental Prediction (NCEP) reanalysis data set for the period 1971–2000 and (2) the simulations from the third generation Canadian general circulation model (CGCM3) for SRES emission scenarios A1B, A2, B1 and COMMIT for the period 1971–2100. These variables include both the thermodynamic and dynamic parameters and those which have a physically meaningful relationship with the precipitation. The NCEP variables which are realistically simulated by CGCM3 are chosen as potential predictors for seasonal stratification. The seasonal stratification involves identification of (1) the past wet and dry seasons through classification of the NCEP data on potential predictors into two clusters by the use of K-means clustering algorithm and (2) the future wet and dry seasons through classification of the CGCM3 data on potential predictors into two clusters by the use of nearest neighbour rule. Subsequently, a separate downscaling model is developed for each season to capture the relationship between the predictor variables and the predictand. For downscaling precipitation, the predictand is chosen as monthly Thiessen weighted precipitation for the river basin, whereas potential predictors are chosen as the NCEP variables which are correlated to the precipitation and are also realistically simulated by CGCM3. Implementation of the methodology presented is demonstrated by application to Malaprabha reservoir catchment in India which is considered to be a climatically sensitive region. The CGCM3 simulations are run through the calibrated and validated SVM downscaling model to obtain future projections of predictand for each of the four emission scenarios considered. The results show that the precipitation is projected to increase in future for almost all the scenarios considered. The projected increase in precipitation is high for A2 scenario, whereas it is least for COMMIT scenario. Copyright © 2007 Royal Meteorological Society

KEY WORDS climate change; downscaling; hydroclimatology; precipitation; support vector machine (SVM); IPCC SRES scenarios; clustering

Received 30 November 2006; Revised 19 February 2007; Accepted 24 February 2007

1. Introduction

It is necessary to access the past and assess the future precipitation and its variability at different timescales to study the impact of climate change on the following: (1) hydrology (Cohen *et al*., 2000; Gosain *et al*., 2006; Menzel *et al*., 2006; Yaning *et al*., 2006); (2) water resources management (Krasovskaia, 1995; Risby and Entekhabi, 1996; Kabat and van Schaik, 2003); (3) agriculture (Darwin *et al*., 1995; Adams *et al*., 1998; Selvaraju, 2003); (4) forestry (Kirilenko and Solomon, 1998); (5), floods (Mirza *et al*., 1998; Reynard *et al*., 1998; Miller *et al*., 2004); (6) droughts (Vicente–Serrano and Lopez–Moreno, 2006); (7) soil erosion (Valentin, 1996; Gregory *et al*., 1999); (8) land use change (IPCC, 2001); (9) groundwater (Sandstrom, 1995; Allen and Scibek, 2006); (10) environment (Jones and Jongen, 1996; Coughenour and Chen, 1997); (11) tourism (More, 1988; Richardson, 2003; Wamukonya, 2003); (12) human and animal population (Brotton and Wall, 1997; Ferguson, 1999) and their health (Tyler, 1994; Hamilton, 1995; Pounds *et al*., 1999).

A proper assessment of probable future precipitation and its variability is to be made for various climate scenarios. These scenarios refer to plausible future climates which have been constructed for explicit use in investigating the potential consequences of anthropogenic climate change and natural climate variability. Since climate scenarios envisage assessment of future developments in complex systems, they are often inherently unpredictable, insufficiently assessed, and have high scientific uncertainties (Carter *et al*., 2001). So, it is preferable to consider a range of scenarios in climate impact studies as such an approach better reflects the uncertainties of the possible future climate change (Houghton *et al*., 2001). The scenarios considered in this study are relevant to the fourth assessment report of the Intergovernmental Panel on Climate Change (IPCC), which is to be released in 2007.

* Correspondence to: D. Nagesh Kumar, Department of Civil Engineering, Indian Institute of Science, Bangalore, India.
E-mail: nagesh@civil.iisc.ernet.in

Copyright © 2007 Royal Meteorological Society

General circulation models (GCMs) are among the most advanced tools, which use transient climate simulations to simulate climatic conditions on earth hundreds of years into the future. In a transient simulation, anthropogenic forcings, which are mostly decided on the basis of IPCC climate scenarios, are changed gradually in a realistic fashion. The GCMs are usually run at coarse grid resolution and as a result they are inherently unable to represent sub-grid scale features like orography and land use, and dynamics of mesoscale processes. Consequently, outputs from these models cannot be used directly for climate impact assessment on a local scale. Hence in the past decade, several downscaling methodologies have been developed to transfer the GCM simulated information to local scale. In general, local scale is defined on the basis of geographical, political or physiographic considerations.

This study is motivated by a desire to develop effective downscaling model using a novel machine learning technique called Support Vector Machine (SVM), to assess implications of climate change on precipitation in Malaprabha river basin of India, which is considered to be a climatically sensitive region. Herein a river basin refers to the portion of land drained by many streams and creeks that flow downhill to form tributaries to the main river. It is necessary to assess the impact of climate change on river basin scale, as such a scale integrates some of the important systems like ecological and socioeconomic systems and is closely linked with the availability of drinking water which is one of the most vulnerable resources of the world.

The remainder of this paper is structured as follows: Section 2 presents an overview of downscaling models and the underlying principle of SVM which is used in this study to translate information from GCMs to local scale. Section 3 provides a description of the study region and motivation for its selection. Section 4 provides details of data used in the study. Section 5 describes probable predictor variables considered in the study and explains the reasons behind their selection. Section 6 presents the procedure proposed for seasonal stratification. Section 7 describes the methodology proposed for development of SVM model for downscaling precipitation to a river basin. Section 8 presents the results and discussion. Finally, Section 9 provides a summary of the work presented in this paper and conclusions drawn from the study.

2. Methods of downscaling

This section briefly outlines the various downscaling methods available in literature. The two downscaling approaches that are being pursued currently are dynamic downscaling and statistical downscaling. In the dynamic downscaling approach, a Regional Climate Model (RCM) is nested into GCM. The RCM is essentially a numerical model in which GCMs are used to fix boundary conditions. The major drawback of RCM, which restricts its use in climate impact studies, is its complicated design and high computational cost. Moreover, RCM is inflexible in the sense that expanding the region or moving to a slightly different region requires redoing the entire experiment (Crane and Hewitson, 1998). Dynamic downscaling can be further subdivided into one-way nesting and two-way nesting (Wang *et al*., 2004).

The Statistical downscaling involves developing quantitative relationships between large-scale atmospheric variables (predictors) and local surface variables (predictands). There are three types of statistical downscaling namely, weather classification methods, weather generators and transfer functions. The most common statistical downscaling approaches are based on transfer functions, which model direct relationships between predictors and predictands (Schoof *et al*., 2007). The transfer functions are conceptually simple means of representing linear or nonlinear relationships between predictors and predictands. Examples of transfer function based statistical downscaling methods include those which use linear and nonlinear regression, artificial neural networks, canonical correlation and principal component analysis (PCA).

In recent times, SVM approach is recognized for its ability to capture nonlinear regression relationships between variables (Vapnik, 1995; Vapnik, 1998; Haykin, 2003). The SVMs implement the structural risk minimization principle which attempts to minimize an upper bound on the generalization error, by striking a right balance between the training error and the capacity of machine (i.e. the ability of machine to learn any training set without error). Global optimum solution is guaranteed with SVM (Haykin, 2003). Further, for the SVMs the learning algorithm automatically decides the model architecture (number of hidden units). The flexibility of the SVM is provided by the use of kernel functions that implicitly map the data to a higher, possibly infinite, dimensional space. A linear solution in the higher dimensional feature space corresponds to a nonlinear solution in the original lower dimensional input space. This makes SVM a feasible choice for downscaling problems which are nonlinear in nature. Recently, Tripathi *et al*. (2006) proposed the SVM approach for downscaling precipitation to meteorological subdivisions in India.

The review of latest literature on downscaling of precipitation using transfer functions is presented in Table I. For details pertaining to predictors and techniques that have been used for downscaling precipitation in the past century, readers are referred to Wilby and Wigley (2000).

There are also statistical approaches to space-time downscaling of precipitation based on scale invariance also referred in literature as dynamic scaling (Perica and Foufoula-Georgiou, 1996; Venugopal *et al*., 1999a, b; Nykanen *et al*., 2001). These studies have considered precipitation from several storm events for downscaling to smaller time and space scales. The time scale considered in the approach is of the order of a few minutes to several hours, while the space scale is of the order of 2–64 km^2. This approach is yet to be extended for downscaling precipitation to river basin scale at longer time scales.

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

Table I. Literature review of predictor selection in statistical downscaling using transfer functions.

SN	Predictor	Timescale	Data	Technique	Region	Reference
1	mslp and pr	Monthly	ECMWF, NCEP reanalysis data sets ECHAM5 GCM	Regression model	Norway	Benestad *et al.* (2007)
2	slp	Monthly	ECHAM4.5	Two regression based models	Southern Africa	Shongwe *et al.* (2006)
3	slp, rh_7, hus_7, ta_7, ps, ta_s, hus_s, rh_s, ws, wd, Z, Di, hus, rh, zg		HadAM3P SRES A2 and B2	Direct method – CCA Indirect method – MLP, single site MLP, RBF, SDSM.	United Kingdom	Haylock *et al.* (2006)
4	mslp, afs_s, afs_5, afs_8, ua_s, ua_5, ua_8, va_s, va_5, va_8, Z_s, Z_5, Z_8, Di_s, Di_8, zg_5, zg_8, wd_5, wd_8, rh_ns, hus_ns, hus_5, hus_8, ta_2m	Daily	CGCM1 using the IPCC 'IS92a', NCEP reanalysis data sets	TNN and SDSM; Correlation and scatter plot to select potential predictors	Canada (single station representing river basin)	Dibike and Coulibaly (2006)
5	ua_7, va_7, hus, rh, ta_s	Daily	NCEP reanalysis data sets, HadAM3, ECHAM4.5, CSIRO k2.	SOM defines the atmospheric state. Each unique atmospheric state is associated with an observed precipitation probability density function (PDF).	South Africa (0.1° and 0.25° grid scale)	Hewitson and Crane (2006)
6	zg_5	Seasonal	NCEP reanalysis data sets, HadAM3	CCA, PCA	Greece (point station)	Tolika *et al.* (2006).
7	pr, zg_0	Daily	ECMWF, NCEP reanalysis data sets	Statistical downscaling by Dynamical scaling, LS, Dynamical intensity scaling, Local intensity scaling	European Alps and adjacent regions	Schmidli *et al.* (2006)
8	ta, rh, hus, zg, ua, va, wa, at various pressure levels and slp.	Monthly	NCEP reanalysis data sets, CGCM 2	SVM, Artificial Neural Network	India (metereological regions)	Tripathi *et al.* (2006)
9	zg_7	Daily	NCAR reanalysis data sets	Fuzzy rule-based methodology	Germany (one rain gauging station)	Bardossy *et al.* (2005)
10	ta_ns, slp, hus_ns	Monthly	CSIRO/Mk2, CCC/CGCM2, UKMO/HadCM3, DOE-NCAR/ PCM, and MPI-DMI/ECHAM4-OPYC3. SRES A2 and B2 emission scenarios	The multi-way partial least squares regression	Central and Western Europe (climate region)	Bergant and Kajfez–Bogataj (2005)
11	zg_5, ta_8, hus_7	Daily	ECHAM4/OPYC3 IPCC IS95a, NCEP reanalysis data sets	Regression method where variability is added by randomization model, inflation model and Expansion Model	Germany (rain gauging stations in river basins)	Burger and Chen (2005)

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

Table I. (*Continued*).

SN	Predictor	Timescale	Data	Technique	Region	Reference
12	mslp, afs_s, afs_5, afs_8, ua_s, ua_5, ua_8, va_s, va_5, va_8, Z_s, Z_5, Z_8, Di_s, Di_8, zg_5, zg_8, wd_5, wd_8, rh_ns, hus_ns, hus_5, hus_8, ta_2m	Daily	CGCM1 using the IPCC 'IS92a', NCEP reanalysis data sets	SDSM, weather generator	Canada (single station representing river basin)	Dibike and Coulibaly, (2005)
13	pr	Daily	ECHAM4, HadCM3, and NCAR–PCM SRES A2 and B2	LS	USA (0.125° grid scale)	Salathe (2005)
14	slp	Daily, Monthly	NCEP reanalysis data sets	AM	Central Sweden Europe (rain gauging stations in a region)	Wetterhall *et al.* (2005)
15	pr	Daily Monthly	HadCM3 for IPCC 'IS95a'	Transfer function bet GCM and local for spatial downscaling for month, CLIGEN for temporal downscaling to daily time step	USA (one rain gauging station for one region)	Zhang (2005)
16	slp	Seasonal	HadCM2, CGCM1 CSIRO–Mk2, CCSR	SVDA with the EOF, truncation to correct the systematic bias in the dynamic models.	Korea (rain gauging stations in a region)	Kim *et al.* (2004)
17	zg_5	Seasonal	NCEP reanalysis data sets	Linear regression and multiple linear regression. Correlation by Spearman's rank correlation.	Greece (metereological regions)	Maheras *et al.* (2004)
18	slp, zg_8, zg_7, zg_5, zg_3, pr	Seasonal	NCEP reanalysis data sets	EOF and CCA	Mediterranean Region (rain gauging stations in a region)	Xoplaki *et al.* (2004)
19	pr, zg_5, zg_7, slp, w_5, zgt_5-0.	Monthly	NCEP – NCAR reanalysis data sets	The Sampson correlation ratio to predictors, preprocessing by independent component analysis. Downscaling by RNN model.	Turkey (point downscaling)	Tatli *et al.* (2004)
20	slp, ta_8, pwr, zg_0, zg_5, zgt_0–5	Daily	NCEP reanalysis data sets, BMRC, CSIRO, LMD	AM	France (17 point station downscaling)	Timbal *et al.* (2003)
21	pr, zg_0. AM–zg_0.	Daily	NCEP reanalysis data sets	LS of the simulated large-scale precipitation;	USA (river basin 50 km grid)	Salathe (2003)

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

Table I. (*Continued*).

SN	Predictor	Timescale	Data	Technique	Region	Reference
22	Lag-1 predictand, Wet, *T*mean, hus_ns, RH_ns, mslp, ua, va, F, *Z*, zg_5,		NCEP reanalysis data sets, CGCM1-greenhouse-gas-plus-sulphate-aerosols experiment	SDSM	Canada (region Toronto)	Wilby *et al*. (2002)
23	zg_7, zg_5, pr	Daily	Observed stations	Classification pattern based on fuzzy rule. Multivariate stochastic downscaling for each pattern.	Germany (8 point stations), Greece (21 point stations)	Stehlik and Bardossy (2002)
24	Indices such as CAPE, CI, BRN, WS, SWTI	Minutes, hours	47 selected radar scans, MM5	Spatial-temporal downscaling. Relation between indices and scale independent parameter.	Central USA (single storms of 2 days, 4–64 km)	Nykanen *et al*. (2001)

Note: Abbreviations are explained in Appendix.

In the current study, the objective is to effectively downscale precipitation from tens of thousands of km^2 (i.e. spatial domain of predictor variables) to a river basin with an area of 2500 km^2 at monthly timescale. To the knowledge of the authors, no studies have been carried out in India on downscaling precipitation to a river basin scale nor are we aware of any prior work aimed at downscaling third generation Canadian general circulation model (CGCM3) simulations to river basin scale for various IPCC scenarios using SVM approach.

2.1. Least square support vector machine

The Least Square Support Vector Machine (LS-SVM) has been used in this study to downscale precipitation. The LS-SVM provides a computational advantage over standard SVM (Suykens, 2001). This section presents an underlying principle of the LS-SVM

Consider a finite training sample of N patterns $\{(\mathbf{x}_i, y_i), i = 1, \ldots, N\}$, where $\mathbf{x}_i$ denoting the 'i-th' pattern in n-dimensional space (i.e. $\mathbf{x}_i = [x_{1i}, \ldots, x_{ni}] \in \Re^n$) constitutes input to LS-SVM and $y_i \in \Re$ is the corresponding value of the desired model output. Further, let the learning machine be defined by a set of possible mappings $\mathbf{x} \mapsto f(\mathbf{x}, \mathbf{w})$, where $f(\cdot)$ is a deterministic function which, for a given input pattern $\mathbf{x}$ and adjustable parameters $\mathbf{w}$ ($\mathbf{w} \in \Re^n$), always gives the same output. Training phase of the learning machine involves adjusting the parameters $\mathbf{w}$. The parameters are estimated by minimizing the cost function $\Psi_{\mathrm{L}}(\mathbf{w}, e)$. The LS-SVM optimization problem for function estimation is formulated by minimizing the cost function.

$$\Psi_{\mathrm{L}}(\mathbf{w}, e) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\,\mathbf{w} + \frac{1}{2}C\sum_{i=1}^{N} e_i^2$$

subject to the equality constraint

$$y_i - \hat{y}_i = e_i \quad i = 1, \ldots, N \tag{1}$$

where C is a positive real constant and $\hat{y}$ is the actual model output. The first term of the cost function represents weight decay or model complexity-penalty function. It is used to regularize weight sizes and to penalize large weights. This helps in improving generalization performance (Hush and Horne, 1993). The second term of the cost function represents penalty function.

The solution of the optimization problem is obtained by considering the Lagrangian as

$$L(\mathbf{w}, b, \mathbf{e}, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \frac{1}{2}C\sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N}\alpha_i\{\hat{y}_i + e_i - y_i\} \tag{2}$$

where α_i are Lagrange multipliers and b is the bias term. The conditions for optimality are given by

$$\begin{cases} \dfrac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{N}\alpha_i\phi(\mathbf{x}_i) = 0 \\ \dfrac{\partial L}{\partial b} = \sum_{i=1}^{N}\alpha_i = 0 \\ \dfrac{\partial L}{\partial e_i} = \alpha_i - Ce_i = 0 \qquad i = 1, \ldots, N \\ \dfrac{\partial L}{\partial \alpha_i} = \hat{y}_i + e_i - y_i = 0 \quad i = 1, \ldots, N \end{cases} \tag{3}$$

The above conditions of optimality can be expressed as the solution to the following set of linear equations

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

after elimination of $\mathbf{w}$ and e_i.

$$\left[\begin{array}{c|c} 0 & \vec{\mathbf{1}}^{\mathrm{T}} \\ \hline \vec{\mathbf{1}} & \Omega + \mathrm{C}^{-1}\mathbf{I} \end{array}\right]\left[\begin{array}{c} b \\ \hline \boldsymbol{\alpha} \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline \mathbf{y} \end{array}\right]$$

$$\text{where } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}; \vec{\mathbf{1}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{N\times 1};$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix}; \mathbf{I} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}_{N\times N} \tag{4}$$

In Equation (4), $\boldsymbol{\Omega}$ is obtained from the application of Mercer's theorem.

$$\Omega_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_j)^T \phi(\mathbf{x}_j) \qquad \forall i, j \tag{5}$$

where $\phi(\cdot)$ represents nonlinear transformation function defined to convert a non linear problem in the original lower dimensional input space to linear problem in a higher dimensional feature space.

The resulting LS-SVM model for function estimation is:

$$f(\mathbf{x}) = \sum \alpha_i{}^{*}\, K(\mathbf{x}_i, \mathbf{x}) + b^* \tag{6}$$

where α_i^* and b^* are the solutions to Equation (4) and $K(\mathbf{x}_i, \mathbf{x})$ is the inner product kernel function defined in accordance with Mercer's theorem (Mercer, 1909; Courant and Hilbert, 1970) and b^* is the bias. There are several possibilities for the choice of kernel function, including linear, polynomial, sigmoid, splines and radial basis function (RBF). The linear kernel is a special case of RBF (Keerthi and Lin, 2003). Further, the sigmoid kernel behaves like RBF for certain parameters (Lin and Lin, 2003). In this study RBF is used to map the input data into higher dimensional feature space, which is given by:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}\right) \tag{7}$$

where, σ is the width of RBF kernel, which can be adjusted to control the expressivity of RBF. The RBF kernels have localized and finite responses across the entire range of predictors.

The advantage with RBF kernel is that it nonlinearly maps the training data into a possibly infinite-dimensional space and thus it can effectively handle the situations when the relationship between predictors and predictand is nonlinear. Moreover, the RBF is computationally simpler than polynomial kernel, which has more parameters. It is worth mentioning that developing LS-SVM with RBF kernel involves selection of RBF kernel width σ and parameter C.

In the present study the LS-SVM model, which has been proven to be effective for downscaling precipitation to Indian meteorological subdivisions in an earlier work (Tripathi *et al*., 2006), is developed for implementation at river basin scale. The scale of meteorological subdivision used by Tripathi *et al*. (2006) is much larger than that of the river basin used in the present study. The effectiveness of the developed model is demonstrated through application to downscale precipitation in catchment of Malaprabha reservoir from simulations of the third generation CGCM3 for latest IPCC scenarios namely, A1B, A2, B1 and COMMIT. In the earlier work (Tripathi *et al*., 2006), simulations of second generation Canadian General Circulation Model for IS92a scenario are used for downscaling precipitation. The IS92a scenario, which is also known as Business-as-usual scenario, considers emissions to grow at the present rate. The scenarios A1B, A2, B1 and COMMIT which are considered in the present study provide a more meaningful basis for impact estimates because they are based on different viewpoints on possible future development pathways and include the major driving forces behind human development such as economic, demographic, social and technological change other than anthropogenic emissions. Each of the scenarios is explained briefly in Table II. Further, issues associated with seasonal stratification and downscaling using SVM model are also discussed in the present study.

3. Study region

The study region is the catchment of Malaprabha river, upstream of Malaprabha reservoir in Karnataka state of India. It has an area of 2564 km^2 situated between 15°30′N and 15°56′N latitudes and 74°12′E and 75°15′E longitudes. It receives an average annual precipitation of 1051 mm. It has a tropical monsoon climate where most of the precipitation is confined to a few months of the monsoon season (Figure 1). The south–west (summer) monsoon has warm winds blowing from the Indian Ocean causing copious amount of precipitation during June–September months. The Malaprabha basin is one of the major lifelines for the arid regions of north Karnataka

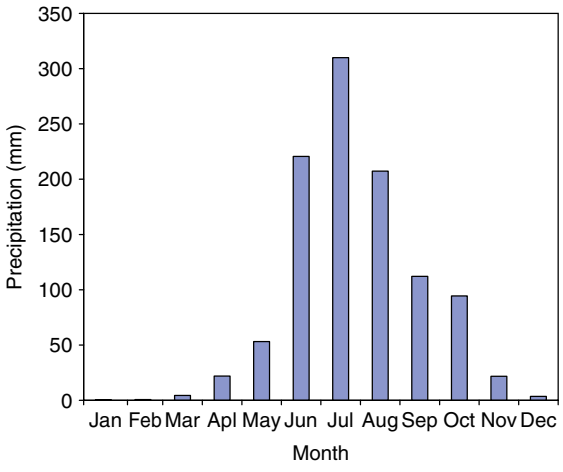

Figure 1. Thiessen Weighted Precipitation (TWP) for the study region. This figure is available in colour online at www.interscience.wiley.com/ijoc

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

Table II. A brief explanation of the scenarios considered in the study.

Name of IPCC scenarios	Data set	Description	Duration
20C3M	Climate of the 20th century (20c3m)	Atmospheric CO_2 concentrations and other input data are based on historical records or estimates beginning around the time of the Industrial Revolution.	1870–2000
COMMIT	Year 2000 CO_2 maximum (COMMIT)	Atmospheric CO_2 concentrations are held at year 2000 levels. This experiment is based on conditions that already exist (e.g. 'committed' climate change).	2001–2100
SRES B1	550 ppm CO_2 maximum (SRES B1)	Atmospheric CO_2 concentrations reached 550 ppm in the year 2100 in a world characterized by low population growth, high GDP growth, low energy use, high land-use changes, low resource availability and medium introduction of new and efficient technologies.	2001–2100
SRES A1B	720 ppm CO_2 maximum (SRES A1B)	Atmospheric CO_2 concentrations reach 720 ppm in the year 2100 in a world characterized by low population growth, very high GDP growth, very high energy use, low land-use changes, medium resource availability and rapid introduction of new and efficient technologies.	2001–2100
SRES A2	850 ppm CO_2 maximum (SRES A2)	Atmospheric CO_2 concentrations reach 850 ppm in the year 2100 in a world characterized by high population growth, medium GDP growth, high energy use, medium/high land-use changes, low resource availability and slow introduction of new and efficient technologies.	2001–2100

(possibly the largest arid region in India outside the Thar desert). Malaprabha reservoir supplies water for irrigation to the districts of northern Karnataka with an irrigable area of 218 191 hectares and the mean annual precipitation in the reservoir command area is 576 mm (Suresh and Mujumdar, 2004). The location map of the study region is shown in Figure 2. The fact that the command area has a much lower mean rainfall than the catchment area indicates that though its origin is in a region of high rainfall it feeds arid and semi-arid regions downstream.

Regions with an arid and semi-arid climate could be sensitive to even insignificant changes in climatic characteristics (Linz *et al*., 1990). In the past decade, Lal *et al*. (1995) predicted a decrease in South Asian monsoon precipitation owing to radiative cooling induced by sulphate aerosols. In contrast, investigations of IPCC (2001) indicate that the mean monsoon precipitation in the region will intensify in future. The motivation of the present study is the need to assess plausible impact of climate change on precipitation in the study region, which has implications on inflows into the Malaprabha reservoir and water availability for irrigation in its command area which is frequently prone to water shortage and is considered to be a climatically sensitive region. To the best of our knowledge, hardly any attempt has been made in the past to downscale precipitation to river basin scale in this part of the world.

4. Data extraction

For the study region, reanalysis data of the monthly mean atmospheric variables prepared by National Center for Environmental Prediction (NCEP; Kalnay *et al*., 1996), is extracted for the period from January 1971 to December 2000 for nine grid points whose latitude ranges from 12.5 °N to 17.5 °N and longitude ranges from 72.5 °E to 77.5 °E at a spatial resolution of 2.5°.

Thiessen weighted precipitation (TWP) data is estimated at a monthly timescale for the period from January 1971 to December 2003, using records available at 11 rain gauge stations which are well distributed in the study region. Primary source of the data is from Department of Economics and Statistics, Government of Karnataka, India. The Thiessen polygons for the study were prepared using ArcGIS 9.0.

The GCM data used in the study are simulations obtained from CGCM3 of the Canadian Center for Climate Modeling and Analysis (CCCma), through website http://www.cccma.bc.ec.gc.ca/. The data consisted of present-day (20C3M) and future simulations forced by four emission scenarios namely, A1B, A2, B1 and COMMIT. A brief description of these scenarios is provided in Table II. The climate data is extracted at a monthly timescale for the period from January 1971 to December 2100, for nine grid points whose latitude ranges from 12.99 °N to 20.41 °N and longitude ranges from 71.25 °E to 78.75 °E. The CGCM3 grid is uniform along the longitude with grid box size of 3.75° and nearly uniform along the latitude (approximately 3.75°). Herein, it is to be mentioned that the spatial domain of climate variables is chosen following the suggestions in Wilby and Wigley (2000).

The development of downscaling model begins with selection of probable predictors, followed by seasonal stratification, and finally ends with training and validation

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

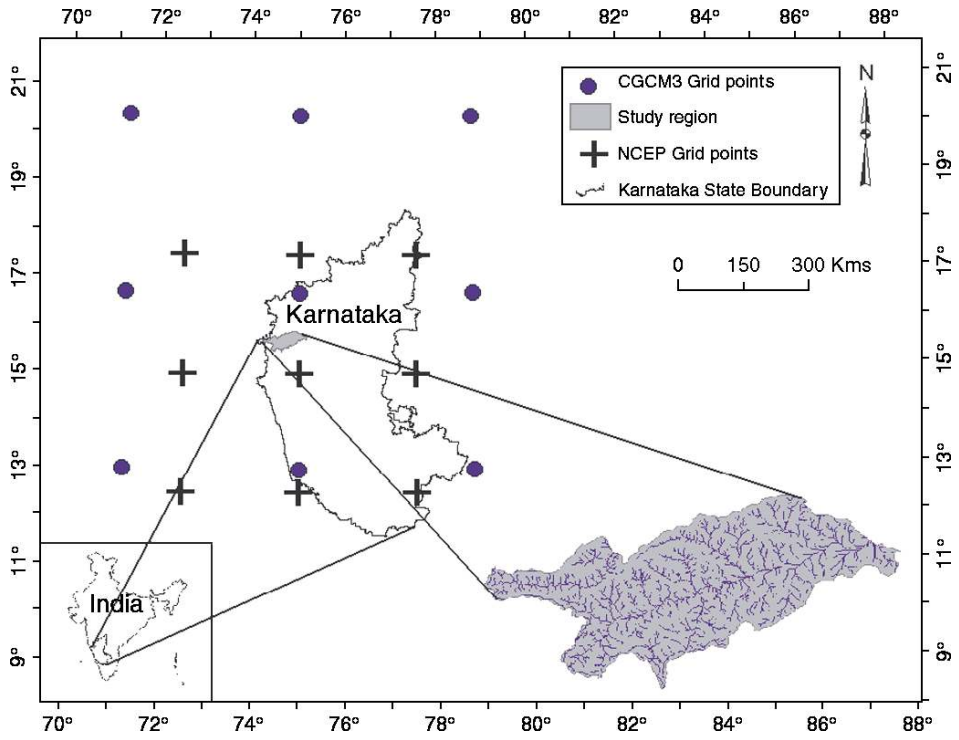

Figure 2. Location map of the study region in Karnataka state of India. The latitude, longitude and scale of the map refer to Karnataka state. This figure is available in colour online at www.interscience.wiley.com/ijoc

of the SVM model. The developed model is subsequently used to obtain projections of precipitation for simulations of CGCM3.

5. Probable predictor selection

The selection of appropriate predictors is one of the most important steps in a downscaling exercise (Hewitson and Crane, 1996; Cavazos and Hewitson, 2005). The choice of predictors could vary from region to region depending on the characteristics of the large-scale atmospheric circulation and the predictand to be downscaled. Any type of variable can be used as predictor as long as it is reasonable to expect that there exists a relationship between the predictor and the predictand (Wetterhall *et al*., 2005). Often, in climate impact studies predictors are chosen as variables that are as follows: (1) reliably simulated by GCMs and are readily available from archives of GCM output and reanalysis data sets and (2) strongly correlated with the predictand.

Other examples of predictors include indices such as 'Convective Available Potential Energy' (CAPE, Perica and Foufoula-Georgiou, 1996; Nykanen *et al*., 2001; Nishiyama *et al*., 2002), 'convergence of air' (CONV) and 'precipitable water' (PW, Nishiyama *et al*., 2002). The CAPE is effective as an index to downscale precipitation at timescale of the order of storm duration. However, the relationship between CAPE and precipitation may not be correctly represented at monthly timescale.

In this study, predictor variables are screened on the basis that monsoon rain is dependent on dynamics through advection of water from the surrounding seas and thermodynamics through effects of moisture and temperature which can modify the local vertical static stability. In a changed climate scenario, both the thermodynamic and dynamic parameters may undergo changes. Therefore in the present study, probable predictor variables (m_1) which incorporate both the effects are identified. The variables are chosen to have a physically meaningful relationship with the predictand. For example, winds during the southwest monsoon season advect moisture into the region while temperature and humidity are associated with local thermodynamic stability and hence are useful as predictors. The probable predictors extracted from the NCEP reanalysis and CGCM3 data sets include air temperature (at 925, 700, 500 and 200 mb pressure levels), geopotential height (at 925, 500 and 200 mb pressure levels), specific humidity (at 925 and 850 mb pressure levels), zonal and meridional wind velocities (at 925 and 200 mb pressure levels), precipitable water and surface pressure. The GCM data is re-gridded to NCEP grid using Grid Analysis and Display System (GrADS; Doty and Kinter, 1993).

6. Seasonal stratification

The climate of the study region can be broadly classified into two seasons: a wet season (the monsoon season) and a dry season. This section outlines the procedure for seasonal stratification performed on the potential predictors chosen from correlated NCEP and GCM probable predictor variables.

The relationship between the predictor variables and the predictand varies seasonally because of the seasonal variation of the atmospheric circulation (Karl *et al*.,

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

1990). Hence seasonal stratification has to be performed to select the appropriate predictor variables for each season to facilitate development of separate downscaling model for each of the seasons. The seasonal stratification can be carried out by defining the seasons as either conventional (fixed) seasons or as 'floating' seasons. In a fixed season, the starting dates and lengths of seasons remain the same for every year. In contrast, in a 'floating' season, the date of onset and duration of each season is allowed to change from year to year. Studies have shown that floating seasons reflect 'natural' seasons contained in the climate data better than fixed seasons, especially under altered climate conditions (Winkler *et al*., 1997). Therefore identification of the floating seasons under altered climate conditions helps to effectively capture the relationships between predictor variables and predictands for each season, thereby enhancing the performance of downscaling model. Hence floating method of seasonal stratification is opted for in this study to identify dry and wet seasons in a calendar year for both NCEP and GCM data sets. In the floating method of seasonal stratification, the NCEP data set is partitioned into two clusters depicting wet and dry seasons by using K-means clustering (MacQueen, 1967), while the GCM data set is partitioned into two clusters by using nearest neighbour rule (Fix and Hodges, 1951) and the results obtained with K-means clustering.

The m_2 climate variables which are realistically simulated by the GCM are selected from among the m_1 probable predictors (listed in Section 5), by specifying a threshold value (T_{ng1}) for correlation between the probable predictor variables in NCEP and GCM data sets. For this purpose, product moment correlation (Pearson, 1896), Spearman's rank correlation (Spearman, 1904a,b) and Kendall's tau (Kendall, 1951) are used as the three measures of dependence. From NCEP data on the m_2 variables, n principal components (PCs) which preserve more than 98% of the variance are extracted to form feature vectors. The PCs, which are extracted along principal directions obtained for the NCEP data, are used to form feature vectors for GCM data. Subsequently, the feature vectors of the NCEP data are partitioned into two clusters (depicting wet and dry seasons) using the K-means cluster analysis. In this analysis, each feature vector (representing a month) of the NCEP data is treated as an object having a location in space. The feature vectors are partitioned into two clusters such that the feature vectors within each cluster are as close to each other as possible in space, and are as far as possible in space from the feature vectors in other clusters. The distance between feature vectors in space is estimated using Euclidian measure. Subsequently, each feature vector (representing a month) of the NCEP data is assigned a label that denotes the cluster (season) to which it belongs. Following this, the feature vectors prepared from GCM simulations (past and future) are labeled using the nearest neighbour rule. As per this rule, each feature vector formed using the GCM data is assigned the label of its nearest neighbour from among the feature vectors formed using the NCEP data. To determine the neighbours for this purpose, the distance is computed between NCEP and GCM feature vectors using Euclidean measure.

Optimal T_{ng1} is identified as a value for which the wet and dry seasons formed for the study region using NCEP data correlate well with the true seasons for the region. The seasons projected for future using GCM data on potential predictors corresponding to the optimal T_{ng1} are deemed acceptable. The plausible true wet and dry seasons are identified in the study region for the period from January 1971 to December 2000 using a method based on truncation level (TL). In this method, the dry season is viewed as consisting of months for which estimated TWP value lies below the specified TL, whereas the wet season is viewed as consisting of months for which estimated TWP value lies above the TL. Herein, two options have been used to specify the TL. In the first option, the TL has been chosen as percentage of the observed mean monthly precipitation (MMP) (70 to 100% MMP at intervals of 5% MMP). In the second option, the TL has been chosen as the mean monthly value of the actual evapotranspiration in the river basin.

Let the probable predictor and predictand for month t be denoted as X_t and Y_t respectively. Then the product moment correlation which measures the linear relationship between probable predictor and predictand is given by

$$P = \frac{\sum_{t=1}^{N}(X_t - \overline{X})(Y_t - \overline{Y})}{N\sigma_X\sigma_Y} \tag{8}$$

where N refers to the number of months in the data sets; $\overline{X}$ and $\overline{Y}$ represent the means of predictor and predictand respectively, while σ_X and σ_Y represent the standard deviations of the same.

Spearman's rank correlation and Kendall's tau are the two nonparametric correlations used in this study which are estimated based on ranks assigned to data points in predictor and predictand data sets. The advantage of these rank correlations over the linear correlation stems from the use of ranks rather than numerical values of the predictor and the predictand variables for estimation of the correlations (Press *et al*., 1992). The ranks are assigned to the N data points in each data set after arranging them in increasing order of magnitude, such that the least value in the data has the first rank. Spearman's rank correlation (ρ) is computed using the difference between the ranks of contemporaneous values of predictor and predictand (D_i).

$$\rho = 1 - \frac{6\sum_{i=1}^{N} D_i^2}{N(N^2-1)} \tag{9}$$

Estimation of the Kendall's tau (τ) for a pair of predictor and predictand data sets involves preparation of N pairs of data ranks $\{(u_i, v_i), i = 1, \ldots, N\}$, where u_i and

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

v_i denote ranks of contemporaneous data points in the predictor and predictand data sets at *ith* time step respectively. Let two pairs of ranks be (u_j, v_j) and (u_k, v_k). The two pairs are concordant if $u_j > u_k$ and $v_j > v_k$, or if $u_j < u_k$ and $v_j < v_k$, for which $(u_j - u_k)(v_j - v_k) > 0$. The two pairs are discordant, if $u_j > u_k$ and $v_j < v_k$, or if $u_j < u_k$ and $v_j > v_k$, for which $(u_j - u_k)(v_j - v_k) < 0$. A tied pair is neither concordant nor discordant, i.e. $(u_j - u_k)(v_j - v_k) = 0$. The Kendall's τ is calculated using the formula given below.

$$\tau = \frac{4\lambda}{\frac{1}{2}N(N-1)} - 1 \tag{10}$$

where λ is the difference between the number of concordant pairs and the number of discordant pairs. So, a high value of λ means that most pairs are concordant, indicating that the two rankings are consistent. Further, $N(N-1)/2$ - total number of possible pairs of ranks. If there are a large number of tied pairs it should be adjusted accordingly. A positive value of τ indicates that the ranks of both the variables increase together, whilst a negative correlation indicates that as the rank of one variable increases the rank of the other decreases. The Kendall coefficient has advantages over the Spearman coefficient (Leach, 1979). The first advantage of Kendall's Tau is that it is appropriate when a large number of ties are present within ranks. The second advantage of the dependence measure is its direct and simple interpretation in terms of probabilities of observing concordant and discordant pairs. The Spearman's coefficient can be considered as the regular Pearson's correlation coefficient in terms of the proportion of variability accounted for, whereas Kendall's coefficient represents a probability, i.e. the difference between the probabilities that the observed data are in the same order and the observed data are not in the same order. The advantages of Kendall coefficient makes it useful to effectively interpret the relationship between (1) the predictors in NCEP and GCM data sets and (2) predictors in NCEP and the predictand.

7. Development of SVM downscaling model

Herein for downscaling precipitation, the m_1 probable predictor variables that have been selected for seasonal stratification are considered at each of the nine grid points surrounding and within the study region (shown in Figure 2). Following this, cross-correlations are computed between the probable predictor variables in NCEP and GCM data sets and the probable predictor variables in NCEP data set and the predictand. Subsequently, a pool of potential predictors is identified for each season by specifying threshold values for the computed cross-correlations. In the discussion to follow, the threshold value for cross-correlation between NCEP and GCM data sets is denoted by (T_{ng2}), whereas the same between NCEP and predictand is depicted as (T_{np}). The T_{np} should be reasonably high to ensure choice of appropriate predictors for downscaling precipitation. Similarly, T_{ng2} should also be reasonably high to ensure that the predictor variables used in downscaling are realistically simulated by the GCM in the past, so that the future projections of the predictand that are obtained from GCM are acceptable.

The downscaling model is calibrated to capture the relationship between NCEP data on potential predictors and the estimated TWP for the Malaprabha catchment. The data on potential predictors is first standardized for each season separately for a baseline period extending from 1971 to 2000. Standardization is widely used prior to statistical downscaling to reduce systemic bias (if any) in the mean and variance of GCM predictors relative to NCEP reanalysis data (Wilby *et al*., 2004). The procedure typically involves subtraction of mean and division by the standard deviation of the predictor for the baseline period. The standardized NCEP predictor variables are then processed using PCA to extract PCs which are orthogonal and which preserve more than 98% of the variance originally present in them. A feature vector is formed for each month of the record using the PCs. The feature vector forms the input to the SVM model, whereas TWP (predictand) constitutes the output of the model. The PCs account for most of the variance in the input and also remove the correlations among the input data. Hence, the use of PCs as input to a downscaling model helps in making the model more stable and at the same time reduces the computational burden.

To develop the SVM downscaling model, the available feature vectors are partitioned into a training set and a test set. The partitioning was initially carried out using a multifold cross-validation procedure, which was adopted from Haykin (2003) for use in an earlier work (Tripathi *et al*., 2006). In this procedure, about 70% of the available feature vectors are randomly selected for training the model and the remaining 30% are used for validation. However, in this study the multifold cross-validation procedure is found to be ineffective because the time span considered for analysis is small and there are more extreme precipitation events in the period 1970–1990 than in the next 10 years. Therefore, feature vectors prepared from the first 22 years (approximately 70% of record) of data are chosen for calibrating the model and the remaining feature vectors are used for validation. The 'normalized mean squared error' (NMSE) is used as an index to assess the performance of the model.

The training of SVM involves selection of the model parameters σ and C. The width of RBF kernel σ can give an idea about the smoothness of the derived function. Smola *et al*. (1998), in their attempt to explain the regularization capability of RBF kernel, have shown that 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 with small kernel width retains most of the higher-order frequencies leading to an approximation of a complex function by the learning machine. In this study, grid search procedure (Gestel *et al*., 2004) is used to find

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

the optimum range for the parameters. Subsequently, the optimum values of parameters are obtained from the selected range using stochastic search technique of genetic algorithm (Haupt and Haupt, 2004).

The feature vectors prepared from GCM simulations are run through the calibrated and validated SVM downscaling model to obtain future projections of predictand for each of the four emission scenarios considered (i.e. SRES A1B, A2, B1 and COMMIT). Subsequently, for each scenario, the projected values of predictand are divided into five parts (2001–2020, 2021–2040, 2041–2060, 2061–2080 and 2081–2100) to determine the trend in the projected values of precipitation.

8. Results and discussion

In the context of seasonal stratification, the highly correlated predictor variables between NCEP and GCM data sets are selected from among the 15 probable predictors listed in Section 5 using the three measures of dependence described in Section 6. In this context, cross-correlation is computed between data points of each probable predictor in NCEP and GCM records using each of the three dependence measures. Subsequently, the cross-correlations computed using each dependence measure are arranged in descending order of magnitude and ranks are assigned to the predictors, such that the predictor having the highest cross-correlation between NCEP and GCM data sets has rank one. Results from the foregoing analysis showed similar (or nearly equal) rank for any chosen predictor by all the three dependence measures considered, indicating that the ranking of predictors is robust. Following this, the highly correlated probable predictor variables between NCEP and GCM data sets are identified as potential predictors. To aid in this task, a threshold (T_{ng1}) is chosen for product moment correlation to segregate high and low correlations. In general, the choice of threshold is subjective. A higher threshold results in selecting a few of the probable predictors as potential predictors. In contrast, a very low threshold results in selecting almost all the probable predictors as potential predictors. In this study, the effect of threshold value on seasonal stratification is analyzed by varying its value in the range −0.2 to 1.0. The set of predictors selected for a few typical threshold values are shown in Table III, for brevity. Results of the seasonal stratification performed using three of the predictor sets, following the procedure described in Section 6, are presented in Figure 3. It can be seen from the figure that seasonal stratification is sensitive to the choice of potential predictors, which in turn depends on the choice of threshold T_{ng1}. In general, selection of a very few predictor variables may be insufficient to represent the thermodynamics and dynamics of the circulation in the region, whereas inclusion of too many predictor variables creates noise as seen in Figure 3(c), where the choice of too many predictors resulted in classifying the entire year as wet season for special report of emission scenarios (SRES) A2 for the period 2081–2100. Hence, the task of selecting the optimum number of predictor variables is crucial for proper seasonal stratification.

Optimal value of T_{ng1} is identified as 0.8 following the procedure described in Section 6. Table IV presents some of the typical results from this analysis. The results of seasonal stratification obtained for the optimum T_{ng1} for the four scenarios namely, SRES A1B, SRES A2, SRES B1 and COMMIT, indicate that the date of onset and duration of seasons are likely to vary in the future (Figure 3(b)). Selection of a higher threshold value would have resulted in elimination of some of the important predictors that influence the thermodynamic and dynamic parameters in the study region. As a consequence, the relationship between predictors and predictand could not have been captured. For example, bias in precipitation simulated by SVM is more for T_{ng1} equal to 0.9, than for T_{ng1} equal to 0.8 as shown in Figures 4(i) and 5(i). The ill-effect of using a low threshold value on seasonal stratification has already been discussed.

In these figures, the precipitation simulated using NCEP and GCM data sets are compared with the

Table III. Potential predictors selected for seasonal stratification using different values of threshold T_{ng1} for product moment correlation between probable predictors in NCEP and CGCM3 data sets.

SN	T_{ng1}	Potential predictors selected for seasonal stratification
1	1.0–0.93	–
2	0.83–0.92	ua_9
3	0.81–0.82	ua_9, ua_2, prw
4	0.77–0.8	ua_9, ua_2, prw, zg_9
5	0.76	ua_9, ua_2, prw, zg_9, hus_8
6	0.75	ua_9, ua_2, prw, zg_9, hus_8, ta_2
7	0.71–0.74	ua_9, ua_2, prw, zg_9, hus_9, hus_8, ta_2
8	0.68–0.7	ua_9, ua_2, prw, zg_9, zg_2, hus_9, hus_8, ta_2
9	0.63–0.67	ua_9, ua_2, prw, zg_9, zg_2, hus_9, hus_8, ta_7, ta_2
10	0.55–0.62	ua_9, ua_2, prw, zg_9, zg_2, hus_9, hus_8, ta_7, ta_5, ta_2
11	0.35–0.54	ua_9, ua_2, prw, zg_9, zg_2, hus_9, hus_8, ta_9, ta_7, ta_5, ta_2

Note: Abbreviations are explained in Appendix.

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

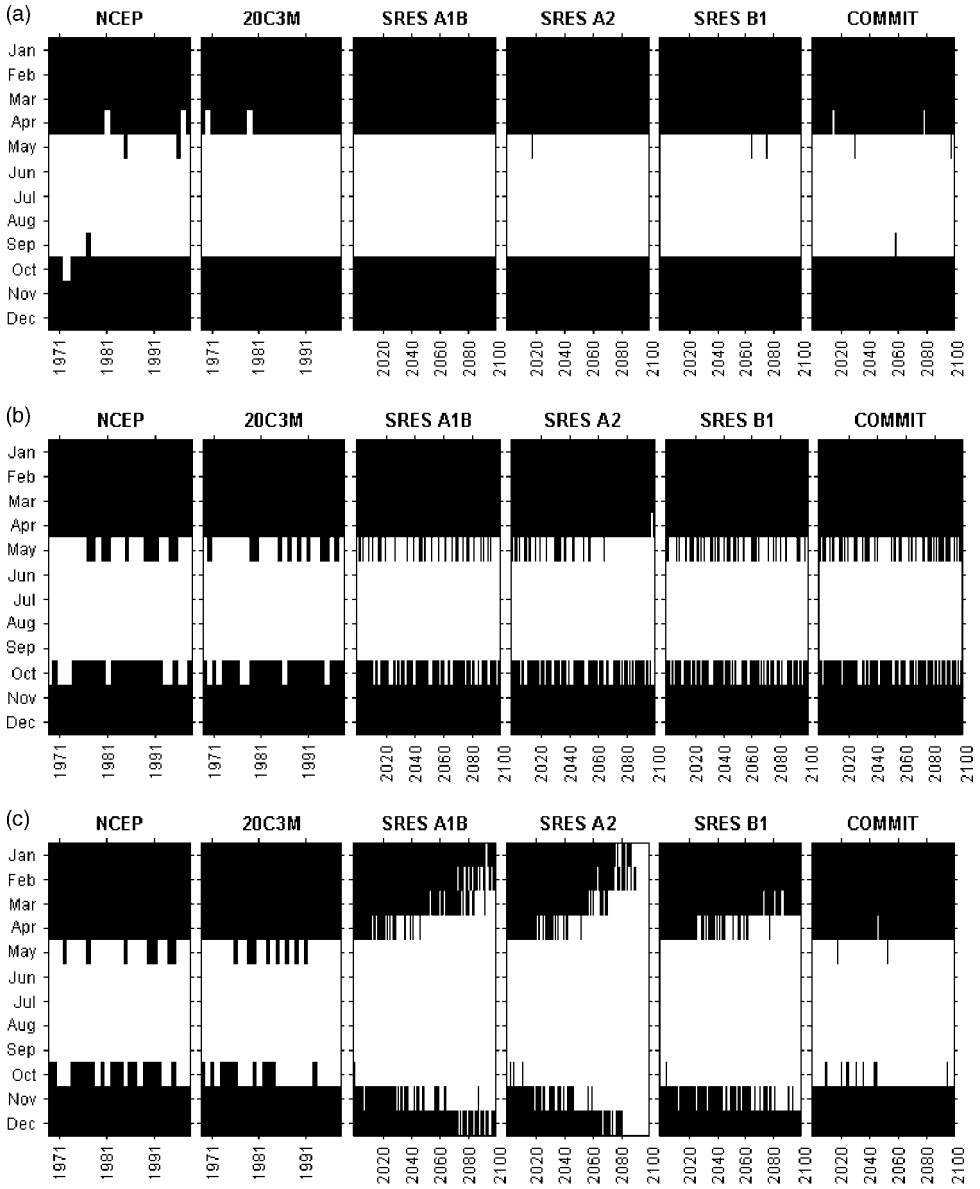

Figure 3. Typical results of seasonal stratification performed using cluster analysis. Dry season is shown in black color, whereas the wet season is shown in white color. The T_{ng1} values chosen are 0.9, 0.8 and 0.45 in Figure 3 (a)–(c) respectively.

Table IV. Selection of optimal value for threshold (T_{ng1}) between predictors in NCEP and GCM data sets. TL refers to truncation level expressed as percentage of TWP for the study region or as mean monthly Actual Evapotranspiration (MME).

Variable	TL	T_{ng1}	
		0.9	0.8
TWP	0.70	0.70	0.75
	0.75	0.70	0.76
	0.80	0.70	0.76
	0.85	0.70	0.78
	0.90	0.71	0.78
	0.95	0.71	0.78
	1.00	0.72	0.77
MME	–	0.64	0.67

estimated TWP for the study region at annual and seasonal scales using boxplots. The span of the box represents the interquartile range of the simulated (or observed) precipitation. The whiskers extend from the box to 5 and 95% quantiles on the lower and the upper side of the box, respectively. Results for the past (1971–2000) are shown in (i). Whereas, the projected precipitation for 2001–2020, 2021–2040, 2041–2060, 2061–2080 and 2081–2100, for the four scenarios SRES A1B, A2, B1 and COMMIT are shown in (ii), (iii), (iv) and (v) respectively.

For downscaling precipitation, the potential predictor variables were identified for each season following the procedure described in Section 7. The selected potential predictors are then used to develop the SVM downscaling model. The optimal values of SVM parameters C and σ are found to be 550 and 50 for wet season and 850

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

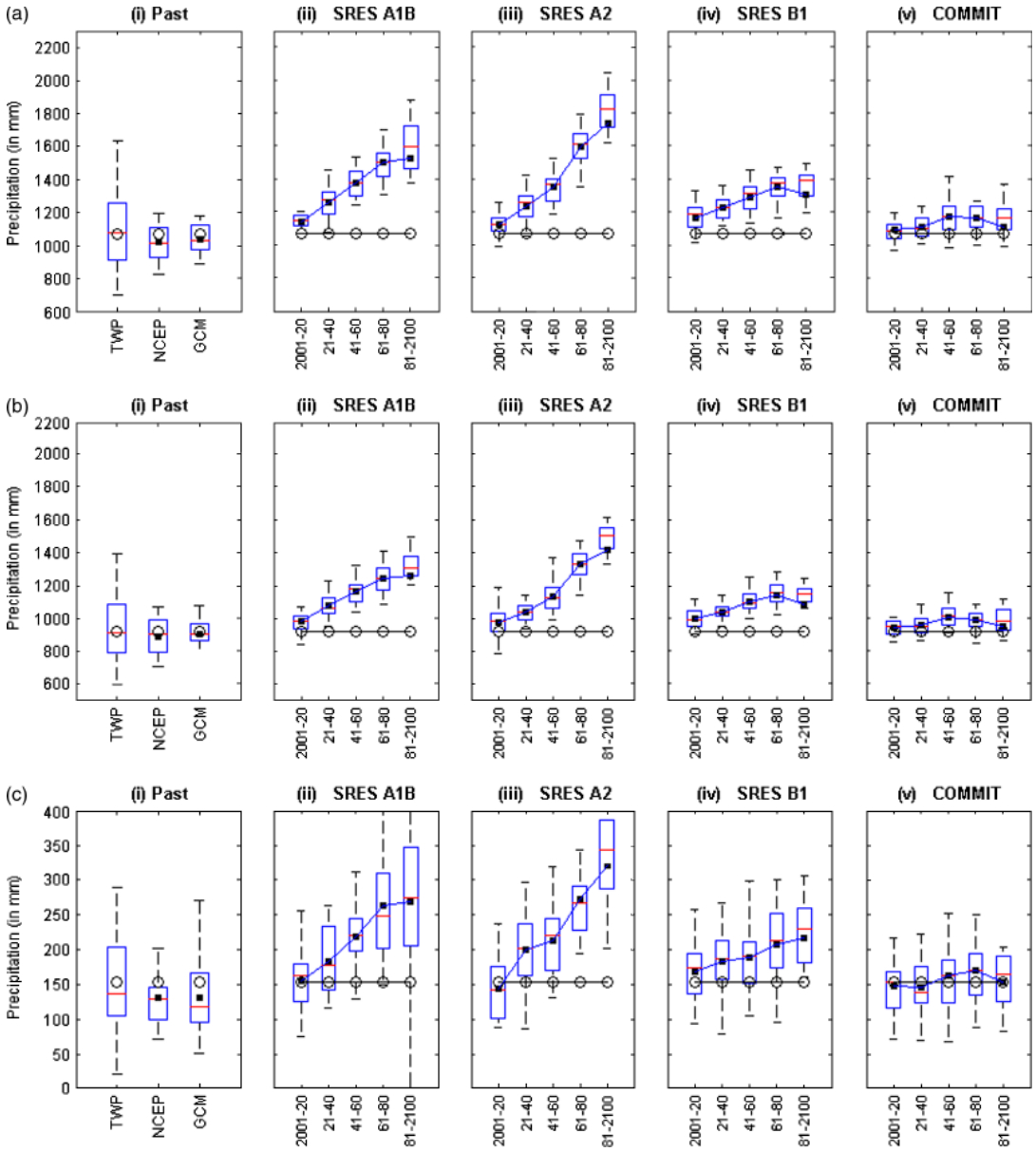

Figure 4. Typical results from the SVM-based downscaling model for T_{ng1} equal to 0.8 graphed using box plots. Figure 4(a) denotes annual scale, while Figures 4(b) and (c) refer to wet and dry seasons respectively. The horizontal line in the middle of the box represents median. The circle denotes the mean value of annual TWP and the darkened square represents the mean value of simulated annual precipitation. In the figure, the gap between darkened square and circle denotes bias in the precipitation simulated by the downscaling model for NCEP and GCM data sets. In (ii), (iii), (iv) and (v) the solid line that joins the circles indicates the historical trend of TWP, while the line connecting the solid squares depicts the mean trend of precipitation projected by GCM. This figure is available in colour online at www.interscience.wiley.com/ijoc

and 50 for dry season respectively, using the grid search procedure described therein (Figure 6).

The robustness of the SVM downscaling model was tested for various threshold values. Typical results of this analysis are presented in Figures 7 and 8. Interestingly, the projections of precipitation obtained from downscaling are largely insensitive to the potential predictor variables considered for downscaling. Hence, for better confidence in the results, we can consider the combination of results as an ensemble.

In the Figures 7 and 8, the span of boxplots (from 5% to 95% quantiles) prepared for precipitation downscaled using from NCEP as well as GCM data sets is found to be less than that of the observed data. This indicates that the SVM model is not able to mimic a few extreme observed precipitation events, which is further confirmed from the results of the model calibration and validation (Figure 9). The inability of the model to simulate high rainfall could be because the regression-based statistical downscaling models cannot often explain the entire variance of the downscaled variable (Wilby *et al*., 2004; Tripathi *et al*., 2006). Exploration of a longer record for calibration of the SVM model could possibly provide more insight into this problem. However, in the present study, investigation in this direction is constrained by the paucity of data. The results presented in Figures 7 and 8 show that precipitation is projected to increase in future for A1B, A2, and B1 scenarios. The projected

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

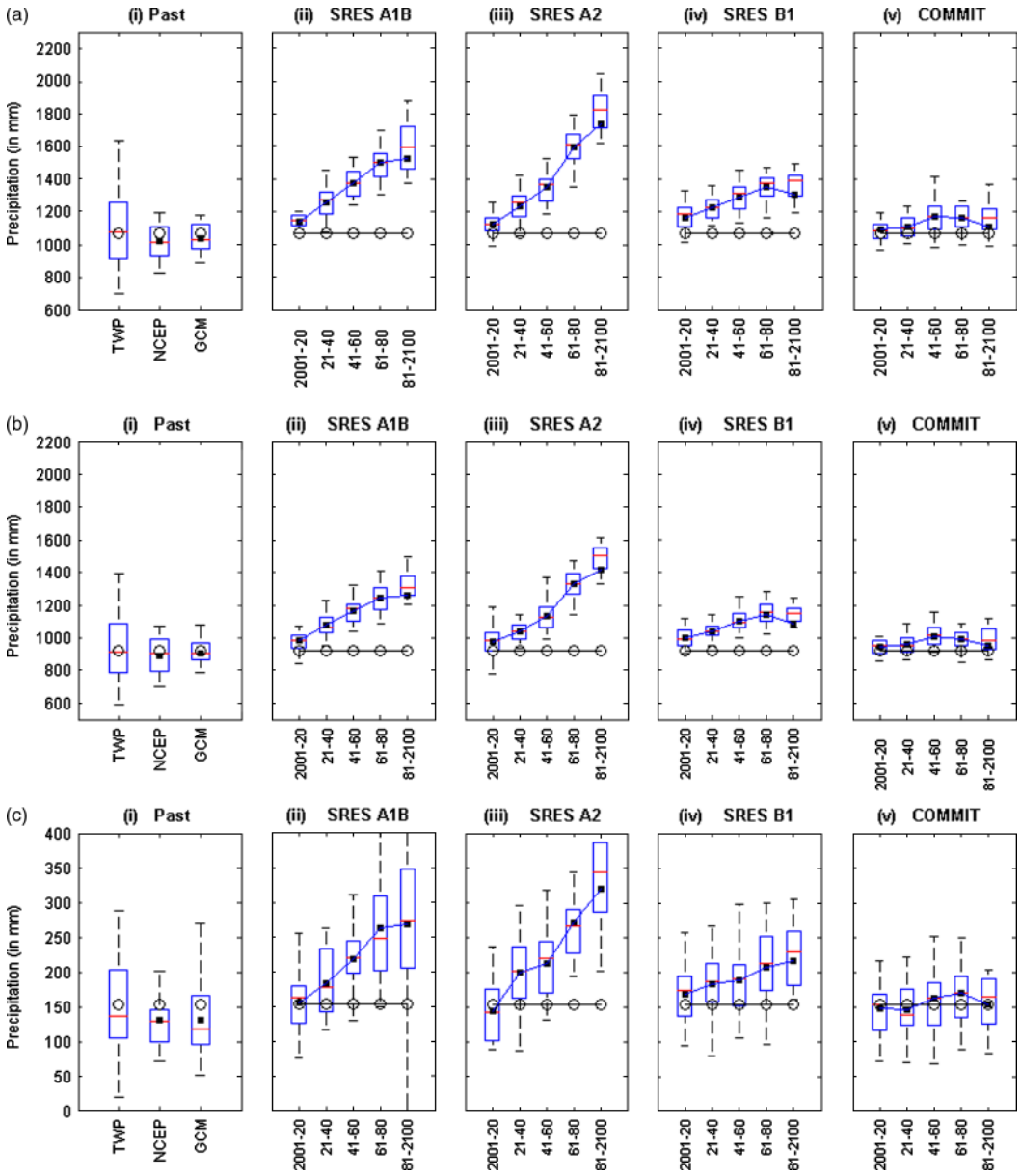

Figure 5. Typical results from the SVM-based downscaling model for T_{ng1} equal to 0.9 graphed using box plots. Figure 5(a) denotes annual scale, while Figures 5(b) and (c) refer to wet and dry seasons, respectively. The horizontal line in the middle of the box represents median. The circle denotes the mean value of annual TWP and the darkened square represents the mean value of simulated annual precipitation. In the figure, the gap between darkened square and circle denotes bias in the precipitation simulated by the downscaling model for NCEP and GCM data sets. In (ii), (iii), (iv) and (v) the solid line that joins the circles indicates the historical trend of TWP, while the line connecting the solid squares depicts the mean trend of precipitation projected by GCM. This figure is available in colour online at www.interscience.wiley.com/ijoc

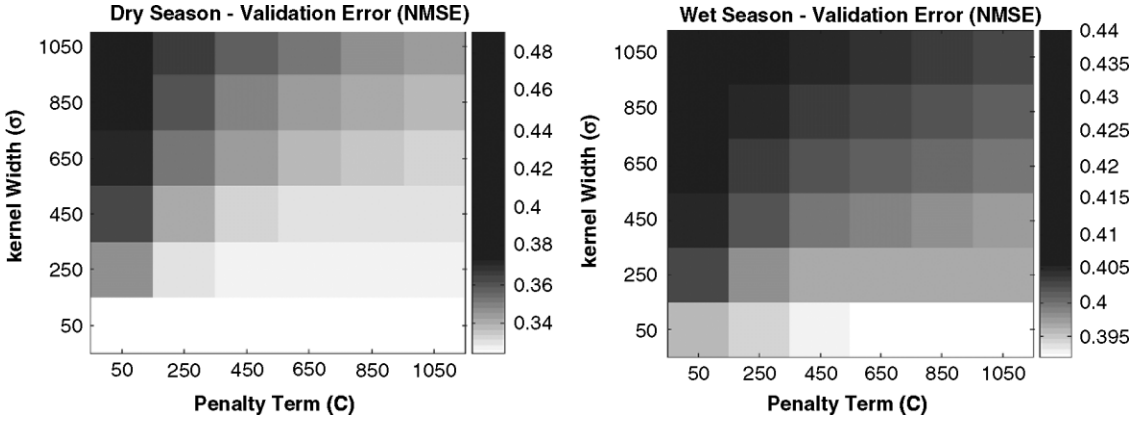

Figure 6. Illustration of the domain search performed to estimate optimal values of kernel width (σ) and penalty (C) for the SVM for dry and wet seasons.

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

increase in precipitation is high for A2 scenario, whereas it is least for B1 scenario. This is because among the scenarios considered, the scenario A2 has the highest concentration of carbon dioxide (CO_2) equal to 850 ppm, while the same for A1B, B2 and COMMIT scenarios are 720 ppm, 550 ppm and ≈370 ppm respectively. Rise in concentration of CO_2 in the atmosphere causes the earth's average temperature to increase, which in turn causes increase in evaporation especially at lower latitudes. The evaporated water would eventually precipitate. In the COMMIT scenario, where the emissions are held the same as in the year 2000, no significant trend in the pattern of projected future precipitation could be discerned. The overall results show that the projections obtained for precipitation are indeed robust.

The results presented in this section demonstrate the usefulness of SVM downscaling model as a feasible choice for obtaining projections of future precipitation at river basin scale in climate change scenarios.

9. Summary and conclusions

The SVM model which has been proven to be effective for downscaling precipitation at regional scale in

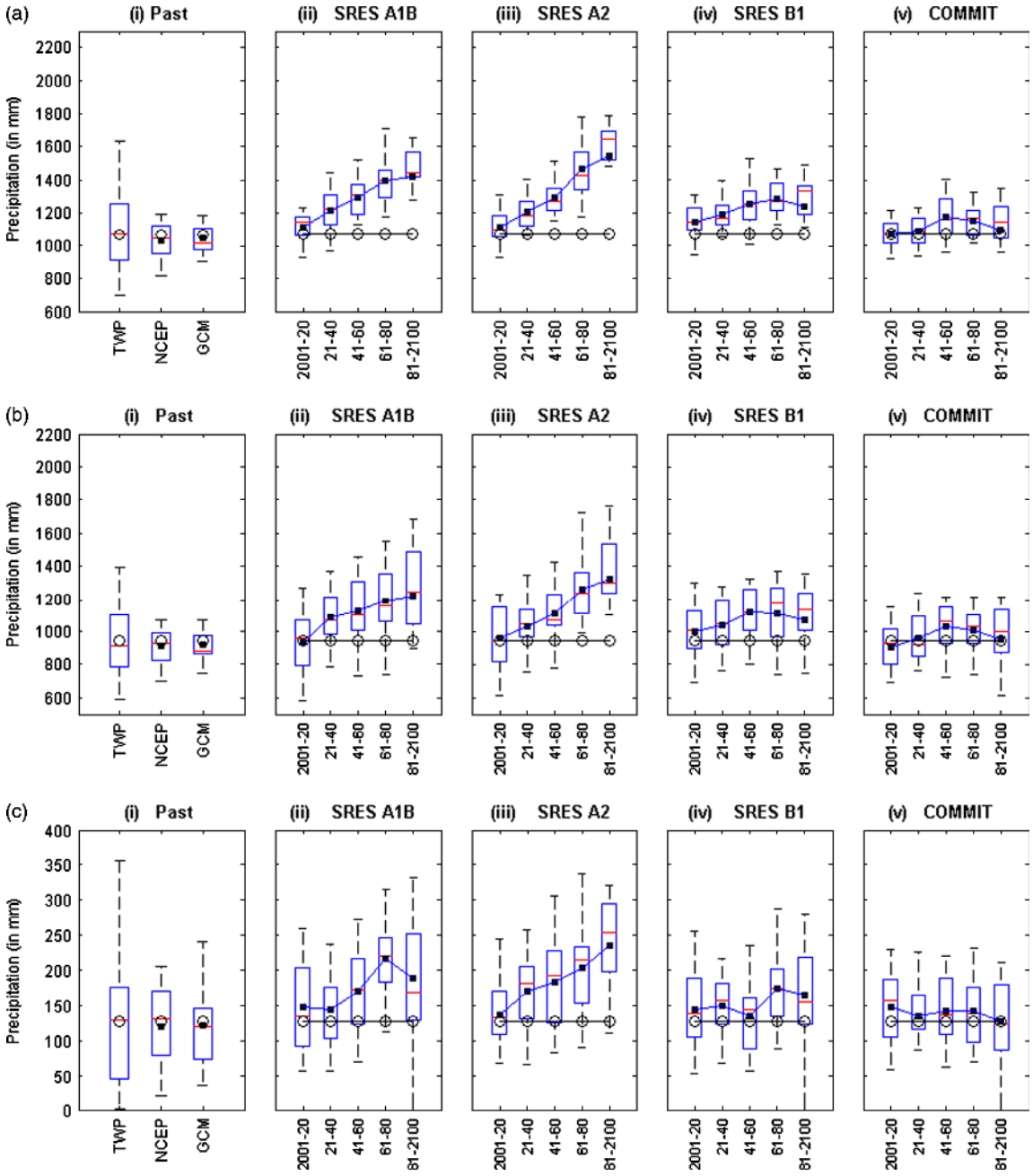

Figure 7. Typical results from the SVM-based downscaling model for T_{ng1} equal to 0.8. The values of T_{ng2} for the wet and the dry seasons are chosen as 0.6 and 0.5, respectively. Whereas, the values of T_{np} for the wet and the dry seasons are selected as 0.7 and 0.65, respectively. Figure 7(a) denotes annual scale, while Figure 7(b) and (c) refer to wet and dry seasons respectively. This figure is available in colour online at www.interscience.wiley.com/ijoc

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

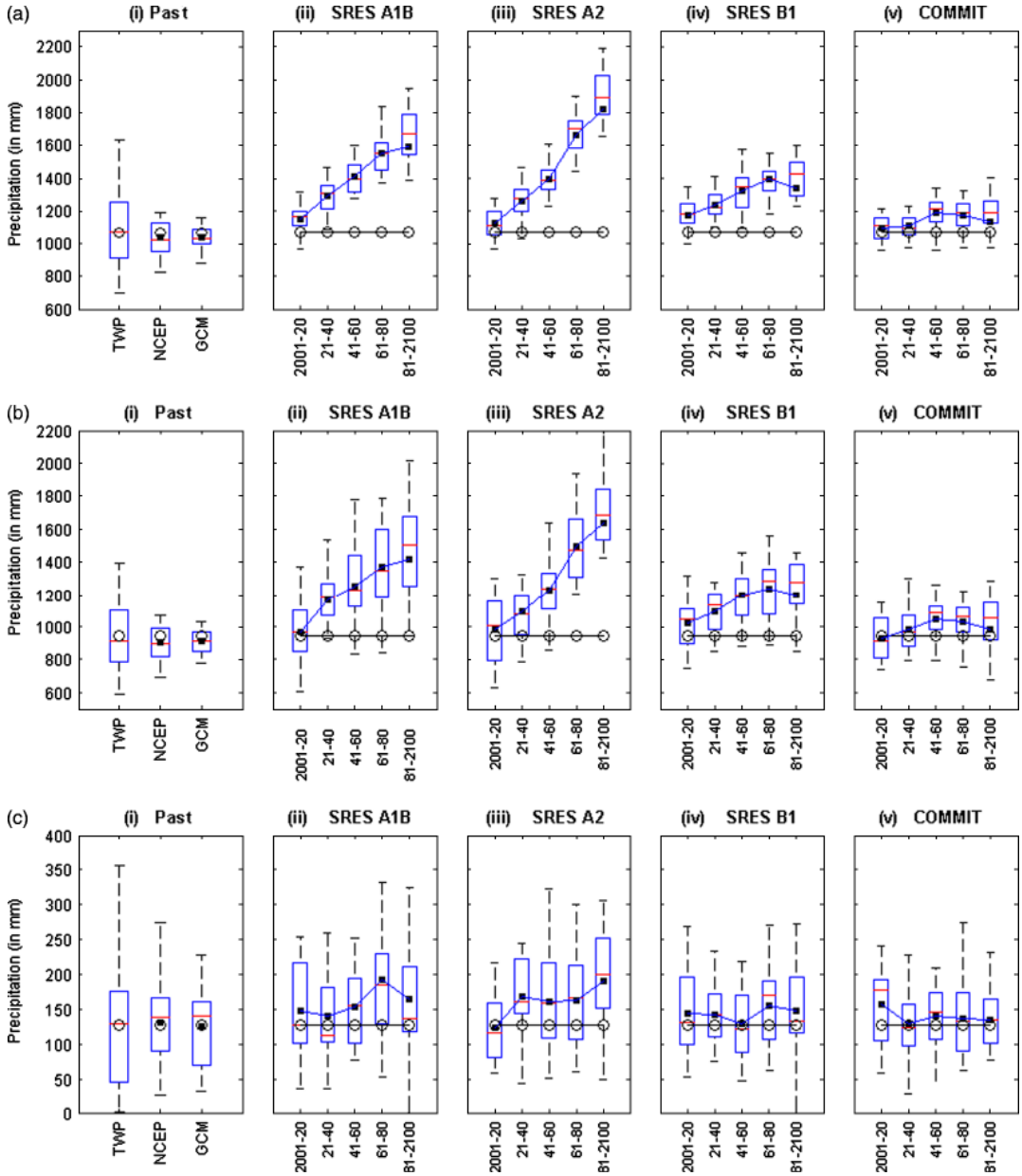

Figure 8. Typical results from the SVM-based downscaling model for T_{ng1} equal to 0.80. The values of T_{ng2} for the wet and the dry seasons are chosen as 0.75 and 0.70 respectively. Whereas, the values of T_{np} for the wet and the dry seasons are selected as 0.55 and 0.50 respectively. Figure 8(a) denotes annual scale, while Figure 8(b) and (c) refer to wet and dry seasons respectively. This figure is available in colour online at www.interscience.wiley.com/ijoc

an earlier work (Tripathi *et al*., 2006) is extended for implementation at river basin scale. The effectiveness of the model is demonstrated through application to downscale precipitation in the catchment of Malaprabha reservoir from simulations of the third generation CGCM3 for four IPCC scenarios namely SRES A1B, A2, B1 and COMMIT. The results of validation clearly indicate that the model is a feasible choice for downscaling precipitation to river basin scale.

The variables which include both the thermodynamic and dynamic parameters and which have a physically meaningful relationship with the precipitation are chosen as the probable predictors. Seasonal stratification was performed to facilitate the development of separate downscaling model for capturing relationship between predictors and predictand for each season.

Seasonal stratification is found to be sensitive to the choice of predictors and the projections obtained from the SVM downscaling model are found to be sensitive to the seasonal stratification. Hence plausible 'true' dry and wet seasons obtained by analyzing records of past precipitation and actual evapotranspiration were used to identify the appropriate potential predictors and the seasons projected for future using CGCM3 data on these predictors are deemed acceptable.

The methodology presented for downscaling is found to be robust in the selection of potential predictors for downscaling. The results show that precipitation is

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

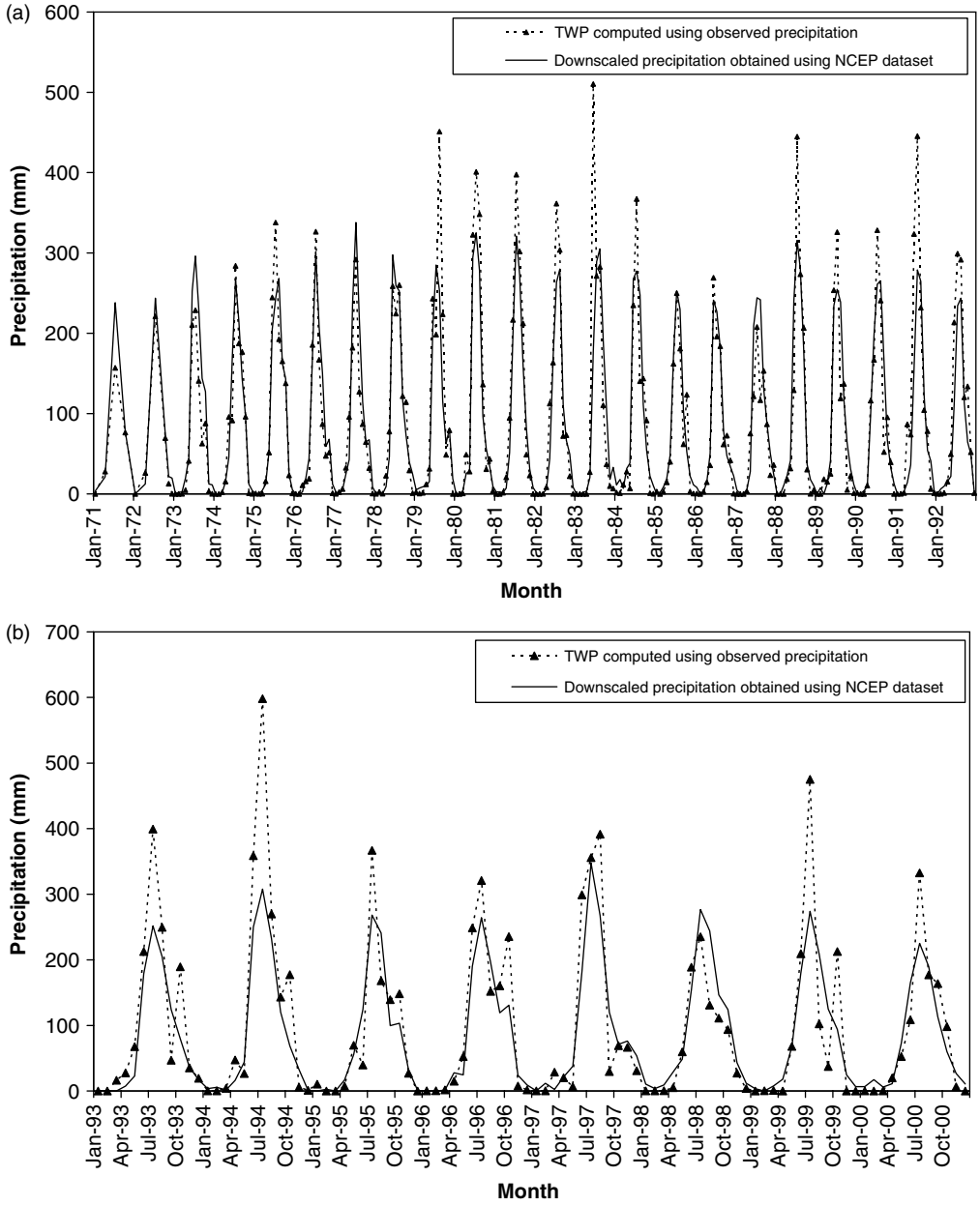

Figure 9. Comparison of the monthly TWP with precipitation simulated using SVM downscaling model for NCEP in Figures 9(a) calibration period, and (b) validation period.

projected to increase in future for almost all the scenarios considered. The projected increase in precipitation is high for A2 scenario, whereas it is least for COMMIT scenario. These results appear to be robust with respect to predictor variables, since many of the variables which undergo changes in a changed climate scenario are considered in predictor selection.

10. Scope of future research

In the present study, simulations from CGCM3 are considered to obtain projections of precipitation for the study region. The future projections of hydrologic variables provided by a downscaling model for a given climate change scenario depend on the capability of GCM to simulate future climate. A realistic simulation by GCM could yield a pragmatic projection of the predictand, while an inconsistent simulation could result in absurd values of the predictand. It may be preferable to use several GCMs and emission scenarios to cover the inherent uncertainties in the scenarios and the possible biases in the GCM simulations. Hence we propose to look for more models in the IPCC Fourth Assessment report 4 (AR4) and check for inter-model robustness of the downscaled results. Besides this, there are uncertainties associated with the assumption that the empirical relationships which are developed on the basis of the current state of atmosphere remain valid for the future. In spite of this assumption, statistical downscaling still remains the most popular tool for hydrologists to assess the impact of climate change on hydrological processes over a smaller

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

region because its computational overheads are practically insignificant compared to dynamic downscaling.

Several avenues can be explored to further refine this attempt to statistically downscale GCM simulations. This proposed approach to statistical downscaling of precipitation is planned to be extended to five other cardinal variables namely, maximum and minimum temperature, wind speed, relative humidity and solar radiation. Extended research work in this direction is underway.

Acknowledgement

This work is partially supported by INCOH, Ministry of Water Resources, Govt. of India, through project no. 23/52/2006-R&D. The authors express their gratitude to the two anonymous reviewers for their constructive comments and suggestions on the earlier draft of the paper. Special thanks are also due to our alumnus Mr Shivam and Ms.Vidyunmala, Indian Institute of Science, Bangalore, for their valuable inputs.

Appendix: Abbreviations

Abbreviations used in text

CCCma	Canadian Center for Climate Modeling and Analysis
CGCM3	Third generation Canadian General Circulation Model
GCM	General Circulation Model
IPCC	Intergovernmental Panel on Climate Change
LS-SVM	Least Square-Support Vector Machine
MMP	mean monthly precipitation
NMSE	Normalized mean square error
PCA	Principal component analysis
PC	Principal component
RBF	Radial basis function
SRES	Special report of Emission scenarios
SVM	Support Vector Machine
TL	truncation level
TWP	Thiessen Weighted precipitation

Abbreviations used in Tables I and III

Predictor Names

afs	Airflow strength
BRN	Bulk Richardson number
CAPE	Convective Available Potential Energy
CI	Convective inhibition
F	Geostropic airflow
hus	specific humidity
pr	precipitation
prw	precipitable water content
ps	pressure
rh	relative humidity
SWTI	Severe weather threat index
ta	air temperature
ua	zonal wind
va	meridional wind
wa	vertical wind
wd	wind direction
WS	wind shear
W	wind speed
zg	geopotential height
zgt	geopotential height thickness
Z	vorticity

Measurement Height of Predictors

_0	pressure height at 1000mb
_2	pressure heights at 200mb
_2m	2m from surface
_5	pressure height at 500mb
_7	pressure height at 700mb
_8	pressure height at 850mb
_9	pressure height at 925mb
_ns	near surface
_s	surface

Techniques

AM	Analogue Method
CCA	Canonical Correlation Analysis
EOF	Empirical Orthogonal Function
LS	Local scaling
MLP	Multilayer perceptron
PCA	Principal Component Analysis
RNN	Recurrent Neural Network
SDSM	Statisitical downscaling model
SOM	Self Organizing Maps
SVDA	Singular value decomposition analysis
TNN	Temporal Neural Network

Data Source

BMRC	Bureau of Meteorology Research Center
CGCM	Canadian General Circulation Model
CSIRO	Commonwealth Scientific and Industrial Research Organization
ECMWF	European Center for Medium-Range Weather Forecasts
HadAM	Hadley center Atmosphere Model
LMD	Laboratoire de Météorologie Dynamique du.

References

Adams RM, Hurd BH, Lenhart S, Leary N. 1998. Effects of global climate change on agriculture: an interpretive review. *Climate Research* **11**: 19–30.

Allen DM, Scibek J. 2006. Comparing modelled responses of two high-permeability unconfined aquifers to predicted climate change. *Global and Planetary Change (Netherlands)* **50**(1–2): 50–62.

Bardossy A, Bogardi I, Matyasovszky I. 2005. Fuzzy rule-based downscaling of precipitation. *Theoretical and Applied Climatology* **82**: 116–119.

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc

Benestad RE, Hanssen-Bauer I, Førland EJ. 2007. An evaluation of statistical models for downscaling precipitation and their ability to capture long-term trends. *International Journal of Climatology* **27**(5): 649–655.

Bergant K, Kajfez-Bogataj L. 2005. N-PLS regression as empirical downscaling tool in climate change studies. *Theoretical and Applied Climatology* **81**: 11–23.

Brotton J, Wall G. 1997. Climate change and the Bathurst caribou herd in the Northwest Territories Canada. *Climatic Change* **35**: 35–52.

Burger G, Chen Y. 2005. Regression-based downscaling of spatial variability for hydrologic applications. *Journal of Hydrology* **96**: 299–317.

Carter TR, La Rovere EL, Jones RN, Leemans R, Mearns LO, Nakicenovic N, Pittock AB, Semenov SM, Skea J. 2001. Developing and applying scenarios. *Climate Change 2001: Impacts Adaptation and Vulnerability*. Cambridge University Press: Cambridge.

Cavazos T, Hewitson BC. 2005. Performance of NCEP variables in statistical downscaling of daily precipitation. *Climate Research* **28**: 95–107.

Cohen SJ, Miller KA, Hamlet AF, Avis W. 2000. Climate change and resource management in the Columbia River basin. *Water International* **25**(2): 253–272.

Coughenour MB, Chen DX. 1997. Assessment of grassland ecosystem responses to atmospheric change using plant-soil process models. *Ecological Applications* **7**: 802–827.

Courant R, Hilbert D. 1970. *Methods of Mathematical Physics*, Vol. I and II. Wiley Interscience: New York.

Crane RG, Hewitson BC. 1998. Doubled CO_2 precipitation changes for the Susquehanna basin: down-scaling from the genesis general circulation model. *International Journal of Climatology* **18**: 65–76.

Darwin RF, Tsigas M, Lewandrowski J, Raneses A. 1995. World agriculture and climate change. Economic Adaptations Agricultural Economic Report Number 703. US Department of Agriculture Economic Research Service: Washington, DC.

Dibike YB, Coulibaly P. 2005. Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. *Journal of Hydrology* **307**: 145–163.

Dibike YB, Coulibaly P. 2006. Temporal neural networks for downscaling climate variability and extremes. *Neural Networks* **19**(2): 135–144.

Doty B, Kinter JL III. 1993. The Grid Analysis and Display System (GrADS): a desktop tool for earth science visualization. In *American Geophysical Union 1993 Fall Meeting*, San Fransico, CA, 6–10 December.

Ferguson MAD. 1999. Arctic tundra caribou and climatic change: questions of temporal and spatial scales. *Geoscience Canada* **23**: 245–252.

Fix E, Hodges JL. 1951. Discriminatory analysis: nonparametric discrimination: consistency properties. Report. 4, Project 21-49-004, USAF School of Aviation Medicine.

Gestel TV, Suykens JAK, Baesens B, Viaene S, Vanthienen J, Dedene G, Moor BD, Vandewalle J. 2004. Benchmarking least squares support vector machine classifiers. *Machine Learning* **54**(1): 5–32.

Gosain AK, Sandhya R, Debajit B. 2006. Climate change impact assessment on hydrology of Indian river basins Special Section: Climate Change and India. *Current Science* **90**(3): 346–353.

Gregory P, Ingram J, Campbell B, Goudriaan J, Hunt T, Landsberg J, Linder S, Stafford-Smith M, Sutherst B, Valentin C. 1999. Managed production systems. *The Terrestrial Biosphere and Global Change Implications for Natural and Managed Ecosystems*. Cambridge University Press: Cambridge.

Hamilton LS. 1995. Mountain cloud forest conservation and research: a synopsis. *Mountain Research and Development* **15**: 259–266.

Haupt RL, Haupt SE. 2004. *Practical Genetic Algorithm*. John Wiley and Sons: New Jersey.

Haylock MR, Cawley GC, Harpham C, Wilby RL, Goodess C. 2006. Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology* **26**: 1397–1415, DOI: 10.1002/joc.1318.

Haykin S. 2003. *Neural Networks: A Comprehensive Foundation*, Fourth Indian Reprint. Pearson Education: Singapore.

Hewitson BC, Crane RG. 1996. Climate downscaling: techniques and application. *Climate Research* **7**: 85–95.

Hewitson BC, Crane RG. 2006. Consensus between GCM climate change projections with empirical downscaling: precipitation downscaling over South Africa. *International Journal of Climatology* **26**: 1315–1337, DOI: 10.1002/joc.1314.

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA. 2001. *Climate Change 2001: The Scientific Basis*. Cambridge University Press: Cambridge (UK) and New York (USA).

Hush DR, Horne BG. 1993. Progress in supervised neural networks: what's new since Lippmann? *IEEE Signal Processing Magazine* **10**: 8–39.

IPCC. 2001. *Climate Change 2001: The Scientific Basis Contributions of Working Group I to the Third Assessment Report of the International Panel on Climate Change*. Cambridge University Press: Cambridge.

Jones MB, Jongen M. 1996. Sensitivity of temperate grassland species to elevated atmospheric CO2 and the interaction with temperature and water stress. *Agricultural and Food Science in Finland* **5**: 271–283.

Kabat P, van Schaik SH. 2003. *Climate changes the water rules: how water managers can cope with today's climate variability and tomorrow's climate change Dialogue on Water and Climate Netherlands*, Available from: ⟨www.waterandclimate.org⟩.

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* **77**(3): 437–471.

Karl TR, Wang WC, Schlesinger ME, Knight RW, Portman D. 1990. A method of relating general circulation model simulated climate to the observed local climate Part I: Seasonal statistics. *Journal of Climate* **3**: 1053–1079.

Kendall MG. 1951. Regression structure and functional relationship part I. *Biometrika* **38**: 11–25.

Keerthi SS, Lin CJ. 2003. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation* **15**(7): 1667–1689.

Kim MK, Kang IS, Park CK, Kim KM. 2004. Super ensemble prediction of regional precipitation over Korea. *International Journal of Climatology* **24**: 777–790.

Kirilenko AP, Solomon AM. 1998. Modeling dynamic vegetation response to rapid climate change using bioclimatic classification. *Climatic Change* **38**: 15–49.

Krasovskaia I. 1995. Quantification of the stability of river flow regimes. *Hydrological Sciences Journal* **40**: 587–598.

Lal M, Cubasch U, Voss R, Waszkewitz J. 1995. Effect of transient increase in greenhouse gases and sulfate aerosols on monsoon climate. *Current Science* **69**: 752–763.

Leach C. 1979. *Introduction to Statistics: A Nonparametric Approach for the Social Sciences*. Wiley: New York.

Lin HT, Lin CJ. 2003. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Technical report. Department of Computer Science and Information Engineering, National Taiwan University.

Linz H, Shiklomanov I, Mostefakara K. 1990. Chapter 4 Hydrology and water *Likely impact of climate change*, IPCC WGII report WMO/UNEP, Geneva.

MacQueen J. 1967. Some methods for classification and analysis of multivariate observation. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Le Cam LM, Neyman J (eds). University of California Press: Berkeley; 281–297.

Maheras P, Tolika K, Anagnostopoulou C, Vafiadis M, Patrikas I, Flocas H. 2004. On the relationships between circulation types and changes in rainfall variability in Greece. *International Journal of Climatology* **24**: 1695–1712.

Menzel L, Thieken A, Schwandt D, Bürger G. 2006. Impact of climate change on the regional hydrology – scenario-based modeling studies in the German Rhine catchment. *Natural Hazards* **38**(1–2): 45–61.

Mercer J. 1909. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society, London* **209**: 415–446.

Miller JR, Dixon MD, Turner MG. 2004. Response of avian communities in large-river floodplains to environmental variation at multiple scales. *Ecological Applications* **14**: 1394–1410.

Mirza MQ, Warrick RA, Ericksen NJ, Kenny GJ. 1998. Trends and persistence in precipitation in the Ganges Brahmaputra and Meghna Basins in South Asia. *Hydrological Sciences Journal* **43**: 845–858.

More G. 1988. Impact of climate change and variability on recreation in the prairie provinces. In *The Impact of ClimateVariability*

Copyright © 2007 Royal Meteorological Society

and Change on the Canadian Prairies: Proceedings of the Symposium/Workshop Edmonton Alberta Alberta, Canada, September 9–11 1987.

Nishiyama K, Ishikawa I, Jinno K, Kawamura A, Wakimizu K. 2002. Radar and GPV-based consideration on significant indices for detecting heavy rainfall. In *Proceedings of ERAD Copernicus GmbH*; Delft, The Netherlands, 272–276.

Nykanen DK, Foufoula-Georgiou E, Lapenta WM. 2001. Impact of small-scale rainfall variability on larger-scale spatial organization of land–atmosphere fluxes. *Journal of Hydrometeorology* **2**: 105–120, DOI: 10.1175/1525-7541(2001)002.

Pearson K. 1896. Mathematical contributions to the theory of evolution III regression heredity and panmixia. *Philosophical Transactions of the Royal Society of London Series* **187**: 253–318.

Perica S, Foufoula-Georgiou E. 1996. Linkage of scaling and thermodynamic parameters of rainfall: results from midlatitude mesoscale convective systems. *Journal of Geophysical Research* **101**(D3): 7431–7448, DOI: 10.1029/95JD02372.

Pounds JA, Fogden MPL, Campbell JH. 1999. Biological response to climate change on a tropical mountain. *Nature* **398**: 611–615.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 1992. *Numerical Recipes in Fortran 77: The Art of Scientific Computing*. Cambridge University Press: New York.

Reynard NS, Prudhomme C, Crooks SM. 1998. The potential impacts of climate change on the flood characteristics of a large catchment in the UK. *Proceedings of the Second International Conference on Climate and Water Espoo Finland August 1998*. Helsinki University of Technology, Helsinki, Finland.

Risby JS, Entekhabi D. 1996. Observed Sacremento Basin streamflow response to precipitation and temperature changes and its relevance to climate impact studies. *Journal of Hydrology* **184**: 209–223.

Richardson R. 2003. The effects of climate change on mountain tourism: a contingent behavior methodology. In *First International Conference on Climate Change and Tourism Djerba Tunisia 9-11* Djerba, Tunisia.

Salathe EP. 2003. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. *International Journal of Climatology* **23**(8): 887–901, DOI: 10.1002/joc.922.

Salathe EP. 2005. Downscaling simulations of future global climate with application to hydrologic modeling. *International Journal of Climatology* **25**: 419–436.

Sandstrom K. 1995. Modeling the effects of rainfall variability on ground water recharge in semi-arid Tanzania. *Nordic Hydrology* **26**: 313–330.

Schmidli J, Christoph F, Pier LV. 2006. Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology* **26**(5): 679–689.

Schoof JT, Pryor SC, Robeson SM. 2007. Downscaling daily maximum and minimum temperatures in the midwestern USA: a hybrid empirical approach. *International Journal of Climatology* **27**(4): 439–454.

Selvaraju R. 2003. Impact of El-Nino–southern oscillation on Indian foodgrain production. *International Journal of Climatology* **23**: 187–206.

Shongwe EM, Landman WA, Mason SJ. 2006. Performance of recalibration systems for GCM forecasts for Southern Africa. *International Journal of Climatology* **26**: 1567–1585, DOI: 10.1002/joc.1319.

Smola AJ, Scholkopf B, Muller KR. 1998. The connection between regularization operators and support vector kernels. *Neural Networks* **11**(4): 637–649.

Spearman CE. 1904a. General intelligence objectively determined and measured. *American Journal of Psychology* **5**: 201–293.

Spearman CE. 1904b. Proof and measurement of association between two things. *American Journal of Psychology* **15**: 72–101.

Stehlik J, Bardossy A. 2002. Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation. *Journal of Hydrology* **256**: 120–141.

Suresh KR, Mujumdar PP. 2004. A fuzzy risk approach for performance evaluation of an irrigation reservoir system. *Agricultural Water Management* **69**: 159–177.

Suykens JAK. 2001. Nonlinear modeling and support vector machines. In *Proceeding of IEEE Instrumentation and Measurement Technology Conference*, Budapest, Hungary; 287–294.

Tatli H, Dalfes HN, Mentes S. 2004. A statistical downscaling method for monthly total precipitation over Turkey. *International Journal of Climatology* **24**(2): 161–180.

Timbal B, Dufour A, McAvaney A. 2003. An estimate of future climate change for western France using a statistical downscaling technique. *Climate Dynamics* **20**: 807–823, DOI 10.1007/s00382-002-0298-9.

Tolika K, Maheras P, Flocas HA, Papadimitriou AA. 2006. An evaluation of a general circulation model (GCM) and the NCEP–NCAR reanalysis data for winter precipitation in Greece. *International Journal of Climatology* **26**: 1376–1385.

Tripathi S, Srinivas VV, Nanjundiah RS. 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2006.04.030.

Tyler MJ. 1994. Climatic change and its implications for the amphibian fauna. *Transactions of the Royal Society of South Australia* **118**: 53–57.

Valentin C. 1996. Soil erosion under global change. *Global Change and Terrestrial Ecosystems*. Cambridge University Press: Cambridge.

Vapnik VN. 1995. *The Nature of Statistical Learning Theory*. Springer: New York.

Vapnik VN. 1998. *Statistical Learning Theory*. Wiley: New York.

Venugopal V, Foufoula-Georgiou E, Sapozhnikov V. 1999a. Evidence of dynamic scaling in space-time rainfall. *Journal of Geophysical Research* **104**(D24): 31599–31610, DOI: 10.1029/1999JD900437.

Venugopal V, Foufoula-Georgiou E, Sapozhnikov V. 1999b. A space-time downscaling model for rainfall. *Journal of Geophysical Research* **104**(D16): 19705–19722, DOI: 10.1029/1999JD900338.

Vicente-Serrano SM, Lopez-Moreno JI. 2006. The influence of atmospheric circulation at different spatial scales on winter drought variability through a Semi-Arid Climatic Gradient in Northeast Spain. *International Journal of Climatology* **26**: 1427–1453.

Wamukonya N. 2003. Managing climate change: strategies for Governments Private sector and consumers. In *International Conference on Climate Change and Tourism Djerba Tunisia 9–11* Djerba, Tunisia.

Wang YQ, Leung LR, McGregor JL, Lee DK, Wang WC, Ding YH, Kimura F. 2004. Regional climate modeling: progress, challenges, and prospects. *Journal of the Meteorological society of Japan* **82**: 1599–1628.

Wetterhall F, Halldin S, Xu CY. 2005. Statistical precipitation downscaling in central Sweden with the analogue method. *Journal of Hydrology* **306**: 136–174.

Wilby RL, Wigley TML. 2000. Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology* **20**(6): 641–661.

Wilby RL, Dawson CW, Barrow EM. 2002. SDSM – a decision support tool for the assessment of climate change impacts. *Environmental Modeling & Software* **17**: 147–159.

Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO. 2004. The guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change (IPCC), prepared on behalf of Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGICA), (http://ipcc-ddc.cru.uea.ac.uk/guidelines/StatDown_Guide.pdf).

Winkler JA, Palutikof JP, Andresen JA, Goodess CM. 1997. The simulation of daily temperature time series from GCM output Part II: Sensitivity analysis of an empirical transfer function methodology. *Journal of Climate* **10**(10): 2514–2532.

Xoplaki E, Gonzalez-Rouco JF, Luterbacher J, Wanner H. 2004. Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends. *Climate Dynamics* **23**: 63–78.

Yaning C, Kuniyoshi T, Changchun X, Yapeng C, Zongxue X. 2006. Regional climate change and its effects on river runoff in the Tarim basin China. *Hydrological Processes* DOI: 101002/hyp6200.

Zhang XC. 2005. Spatial downscaling of global climate model output for site-specifitc assessment of crop production and soil erosion. *Agricultural and Forest Meteorology* **135**: 215–229.

Copyright © 2007 Royal Meteorological Society