INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **29**: 583–603 (2009)
Published online 3 June 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.1719

Role of predictors in downscaling surface temperature to river basin in India for IPCC SRES scenarios using support vector machine

Aavudai Anandhi,[a] V. V. Srinivas,[a] D. Nagesh Kumar[a]* and Ravi S. Nanjundiah[b]
[a] *Department of Civil Engineering, Indian Institute of Science, Bengalooru-560 012, India*
[b] *Centre for Atmospheric and Oceanic Sciences, Indian Institute of Science, Bengalooru-560012, India*

ABSTRACT: In this paper, downscaling models are developed using a support vector machine (SVM) for obtaining projections of monthly mean maximum and minimum temperatures ($T_{\max}$ and $T_{\min}$) to river-basin scale. The effectiveness of the model is demonstrated through application to downscale the predictands for the catchment of the Malaprabha reservoir in India, which is considered to be a climatically sensitive region. The probable predictor variables are extracted from (1) the National Centers for Environmental Prediction (NCEP) reanalysis dataset for the period 1978–2000, and (2) the simulations from the third-generation Canadian Coupled Global Climate Model (CGCM3) for emission scenarios A1B, A2, B1 and COMMIT for the period 1978–2100. The predictor variables are classified into three groups, namely A, B and C. Large-scale atmospheric variables such as air temperature, zonal and meridional wind velocities at 925 mb which are often used for downscaling temperature are considered as predictors in Group A. Surface flux variables such as latent heat (LH), sensible heat, shortwave radiation and longwave radiation fluxes, which control temperature of the Earth's surface are tried as plausible predictors in Group B. Group C comprises of all the predictor variables in both the Groups A and B. The scatter plots and cross-correlations are used for verifying the reliability of the simulation of the predictor variables by the CGCM3 and to study the predictor-predictand relationships. The impact of trend in predictor variables on downscaled temperature was studied. The predictor, air temperature at 925 mb showed an increasing trend, while the rest of the predictors showed no trend. The performance of the SVM models that are developed, one for each combination of predictor group, predictand, calibration period and location-based stratification (land, land and ocean) of climate variables, was evaluated. In general, the models which use predictor variables pertaining to land surface improved the performance of SVM models for downscaling $T_{\max}$ and $T_{\min}$. Copyright © 2008 Royal Meteorological Society

KEY WORDS climate change; downscaling; hydroclimatology; maximum and minimum temperature; support vector machine; IPCC SRES scenarios

Received 18 April 2007; Revised 6 December 2007; Accepted 7 April 2008

1. Introduction

Information concerning spatio-temporal patterns of temperature and their variability is necessary to model various surface processes at global and local scales in disciplines like hydrology, anthropology, agriculture, forestry, environmental engineering and climatology. Temperature influences biological events like diseases (Collinson and Sparks, 2004), phenological events (e.g. the timing of natural events such as flowering, breeding) and agronomy (Croxton *et al.*, 2006), and is as an indicator of climate change. Hence, there is a need to access the past and assess the future temperature and its variability at different time scales to study the impact of climate change at both global and local scales. In general, local scale is defined based on geographical, political or physiographic considerations and is of the order of hundreds of square kilometers.

A proper assessment of probable future temperature and its variability is to be made for various climate scenarios. These scenarios refer to plausible future climates, which have been considered 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). Therefore it is preferable to consider a range of scenarios in climate impact studies, as such an approach better reflects the uncertainties of possible future climate change (Houghton *et al.*, 2001). The scenarios which are studied in this paper are relevant to Intergovernmental Panel on Climate Change's (IPCC's) fourth assessment report (AR4) which was released in 2007.

Global climate models (GCMs) are among the most advanced tools which use transient climate simulations to

* Correspondence to: D. Nagesh Kumar, Department of Civil Engineering, Indian Institute of Science, Bengalooru-560 012, India.
E-mail: nagesh@civil.iisc.ernet.in

simulate climatic conditions on earth, hundreds of years into the future. In a transient simulation, anthropogenic forcings, which are mostly decided based on IPCC climate scenarios, are changed gradually in a realistic pattern. 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.

The present study is motivated to develop effective models for downscaling temperature using a novel machine-learning technique called support vector machine (SVM). The role of predictors on the downscaled temperature and the implications of climate change on temperature in the Malaprabha river basin of India are studied. The river basin is considered to be a climatically sensitive region. In general, 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.

The remainder of this paper is structured as follows: Section 2 presents an overview of the study. 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 how the various predictor variables behave for the different scenarios and the reasons for selection of the probable predictor variables for downscaling. Section 6 explains the proposed methodology for development of the SVM model for downscaling $T_{\max}$ and $T_{\min}$ to the river basin. Section 7 presents the results and discussion. Finally, Section 8 provides a summary of the work presented in the paper and the conclusions drawn from the study.

2. Overview of the study

This section briefly outlines the objectives of the study. The various downscaling methods available in literature, the advantages of SVM for downscaling, the fundamental principle of SVM and its formulation are discussed in detail in Anandhi *et al*. (2008) and Tripathi *et al*. (2006).

A review of the latest literature on downscaling of temperature by using transfer functions is presented in Table I. Details pertaining to selection of predictors for downscaling temperature are given in Schoof *et al*. (2007) from articles catalogued up to 2004. To the knowledge of the authors, no studies have so far been carried out in India for downscaling temperature to a river-basin scale, nor was there any prior work aimed at downscaling third-generation Canadian Coupled Global Climate Model (CGCM3) simulations to temperature at river-basin scale for various IPCC emission scenarios. Further,

Table I. Literature Review on predictors used for statistical downscaling of temperature by using transfer functions.

Sl. no	Predictor	Predictand	Data	Technique	Region	Author
1	ua_5, va_5, zg_5, ua_7, va_7, zg_7	Daily near surface lapse rates	NCEP–NCAR reanalysis datasets	Extrapolation	Canada	Marshall *et al*. (2007)
2	Mgeos, Mz_5, Mz_8 Mrh850/Mhus850, mslp, Mzgt_8_5 for downscaling $T_{\min}$	Daily $T_{\max}$ and $T_{\min}$	HadCM3 and CGCM2 simulations for SRES A2 scenario, and NCEP–NCAR and ECMWF reanalysis data sets	MLR	USA (26 stations)	Schoof *et al*. (2007)
	Mgeow, Mz_5, Mz_8 mslp, Mzgt_8_5 for downscaling $T_{\max}$					
3	mslp, afs_s, afs_5, afs_8, ua_s, ua_5, ua_8, va_s, va_5, va_8, zg_5, zg_8, di_s, di_8, wd_5, wd_8, rh_ns, hus_ns, hus_5, hus_8, ta_2m, Z_s, Z_5, Z_8	Daily $T_{\max}$ and $T_{\min}$	CGCM1 simulations for IS92a scenario, NCEP data for grid point closest to watershed	TNN and multiple regression based SDSM	Canada (river basin)	Dibike and Coulibaly (2006)
4	ta_ns, mslp	Daily temperature	CSIRO/Mk2, HadCM3, PCM, and ECHAM4 datasets for SRES A2 and B2 scenarios	Regression models	Slovenia	Bergant *et al*. (2006)

Note: Abbreviations are explained in Appendix.

Table I. (*Continued*).

Sl. no	Predictor	Predictand	Data	Technique	Region	Author
5	Monthly T_{max} for downscaling T_{max}	Daily and monthly T_{max} and T_{min}	HadCM3 projections for GGa emissions scenario	Transfer function for spatial downscaling from GCM grid box to station, CLIGEN for temporal downscaling from monthly to daily scale	USA (one station in Oklahoma)	Zhang (2005)
	Monthly T_{min} for downscaling T_{min}					
6a	va_ns, hus_ns, hus_8, zg_5, ta_m	Daily T_{max}	CGCM1 datasets for IS92a scenario, and NCEP–NCAR reanalysis datasets	SDSM	Canada (river basin)	Dibike and Coulibaly (2005)
6b	va_ns, zg_ns, hus_ns, hus_8, zg_5, ta_m	Daily T_{min}				
7	T_{mean}	Monthly temperature	HadCM3, ECHAM4 datasets for SRES A2 and B2 scenarios	LS	Sri Lanka	Droogers and Aerts (2005)
8	zg_5, zgt_0_5	Monthly T_{mean}, T_{min} and T_{max}	NCEP–NCAR reanalysis data sets	SSA, PCA, CCA	Turkey (62 stations)	Tatli *et al.* (2005)
9	Mslp, ta_8, prw, zg_0, zg_5, zgt_0_5	Daily T_{min} and T_{max}	NCEP–NCAR reanalysis data sets, simulations from three AOGCMs - BMRC, CSIRO, LMD	AM	France (17 stations)	Timbal *et al.* (2003)
10	zg_5	Winter monthly temperature	NCEP–NCAR reanalysis data sets	CCA	China (147 stations)	Chen and Chen (2003)
11	T_{max} and T_{min} value for previous day, T_{mean}_2m, hus_ns, rh_ns, mslp, ua, va, F, Z, zg_5	Daily T_{max} and T_{min}	NCEP–NCAR reanalysis data sets, CGCM1 dataset for greenhouse-gas-plus-sulphate-aerosols experiment	SDSM	Canada (region Toronto)	Wilby *et al.* (2002)
12	ta_2m, slp	Monthly T_{mean}	ECHAM4	EOF	Norway (gridded region)	Benestad (2001)

Note: Abbreviations are explained in Appendix.

it is noted that latent heat (LH), sensible heat (SH), shortwave and longwave radiation fluxes, which control the temperature at the surface, have not been considered as plausible predictor variables for downscaling temperature.

In the present study, the least square-support vector machine (LS-SVM) model is introduced to downscale T_{max} and T_{min} to a river-basin scale. The effectiveness of the SVM is demonstrated through application to downscale T_{max} and T_{min} in catchment of Malaprabha reservoir from simulations of CGCM3 for latest IPCC scenarios given in Special report of Emission scenarios (SRES), namely, A1B, A2, B1 and COMMIT. Each of the scenarios is explained briefly in Table II.

The effectiveness of the LS-SVM in downscaling precipitation to the river-basin scale has been brought out in Anandhi *et al*. (2008). Therein, the climate of the study region is stratified into two seasons (wet/monsoon season and dry season) based on precipitation to effectively capture the relationship between precipitation and its predictor variables in each season.

Though conceptually the work carried out in this study is similar to Anandhi *et al*. (2008), there are certain differences in the actual procedure of implementation and validation. The surface temperature in a region is dominated by localized effects such as evaporation, SH flux and vegetation in the region. Therefore, the predictor variables influencing surface temperature in the study

DOI: 10.1002/joc

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

Dataset	Description	IPCC name	Dates
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.	20C3M	1870–2000
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).	COMMIT	2001–2100
550 ppm CO_2 maximum (SRES B1)	Atmospheric CO_2 concentrations reach 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.	SRES B1	2001–2100
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.	SRES A1B	2001–2100
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.	SRES A2	2001–2100

region are stratified based on location (i.e. whether the surface is land or ocean) to assess the impact of using predictor variables pertaining to (1) only land grid points, and (2) both ocean and land grid points on downscaled temperature. As there are no distinct seasons based on temperature, seasonal stratification as in the case of precipitation is not relevant. Further, in this study, (1) in addition to the predictors generally used for downscaling temperature, a new set of predictors namely the LH, SH, shortwave and longwave radiation fluxes which control the temperature at the surface, have been additionally considered as plausible predictor variables; (2) effect of length of the calibration period on the downscaled results is examined; (3) relationship between the trend of the predictors and predictand is analysed; (4) sensitivity of the projections obtained for temperature to the predictor group is studied.

3. Study region

The study region is the catchment of Malaprabha reservoir in the Karnataka state of India. It covers 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. The mean monthly T_{max} in the catchment varies from 25 to 34 °C and mean annual T_{max} is 28 °C. The mean monthly T_{min} ranges from 17 to 21 °C (Figure 1). The day temperatures rarely fall below 25 °C. The hottest months are April and May with mean maximum temperature of around 34 °C. December and January are the coldest months with mean minimum temperature of around 17 °C. On annual basis, the diurnal difference between the maximum and the minimum temperatures is in the range of 8–13 °C. The Malaprabha basin is one of the major lifelines for the arid regions of north Karnataka (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. The location map of the study region is shown in Figure 2.

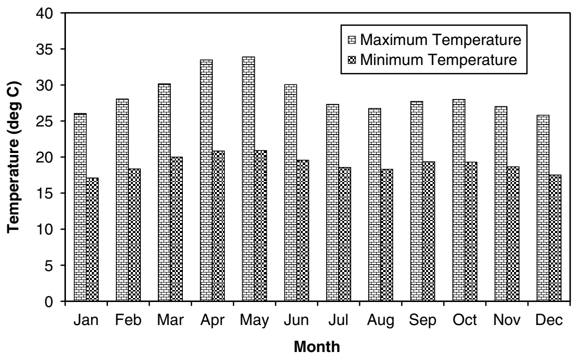

Figure 1. Maximum and minimum temperature in the study region.

Regions with arid and semi-arid climates could be sensitive even to insignificant changes in climatic

DOI: 10.1002/joc

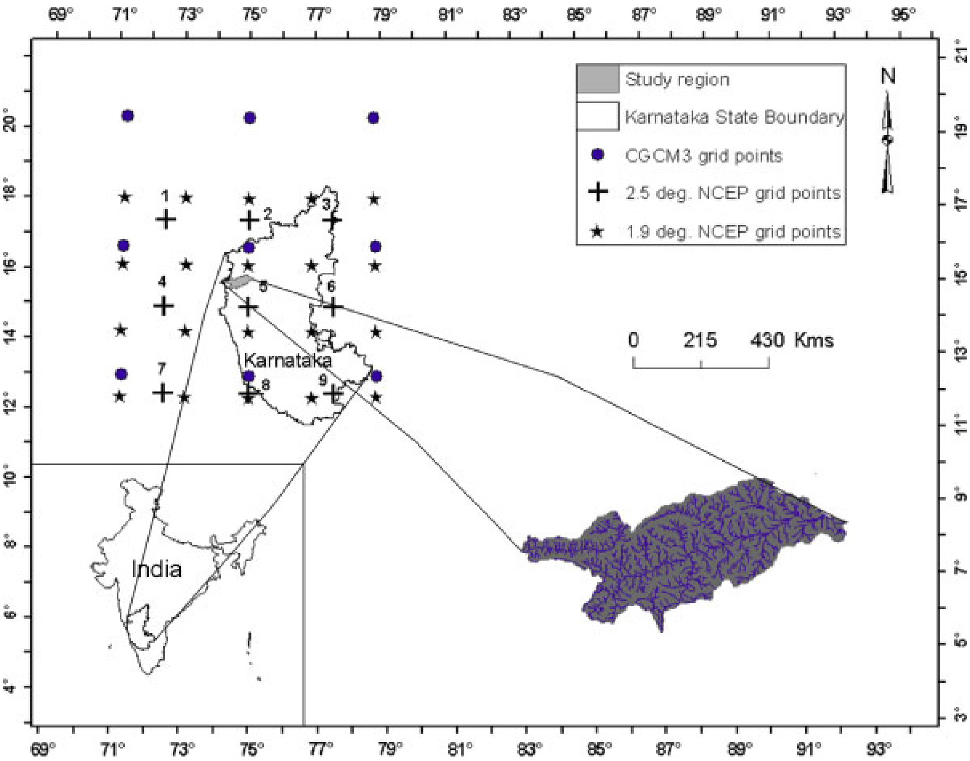

Figure 2. Location map of the study region in Karnataka State of India. The latitude, longitude and scale of the map refer to Karnataka State. The data extracted at CGCM3 and 1.9° grid points are re-gridded to the nine 2.5° NCEP grid points. This figure is available in colour online at www.interscience.wiley.com/ijoc

characteristics (Linz *et al.*, 1990). Temperature affects the evapotranspiration (ET, Jessie *et al.*, 1996), evaporation and desertification processes and is also considered as an indicator of environmental degradation and climate change. Changes in variables such as ET and soil evaporation affect soil moisture content (Pitman, 2003). Increase in temperature would result in increase in ET which is a major cause of water depletion from riverine systems in arid and semi-arid climates (Dahm *et al.*, 2002). Interestingly, investigations of Roderick and Farquhar (2005) indicate a decline in potential evaporation in India for the period 1961–1992, despite increase in near-surface air temperature. This is because temperature is only one of the factors that determines the evaporative demand of the atmosphere, the others being vapour-pressure deficit, wind speed and net radiation. The change in evaporative demand depends on how those factors change, as well as on the change in temperature (Rosenberg *et al.*, 1989).

The motivation for the present study is to assess plausible impact of climate change on T_{max} and T_{min} in the study region, which indirectly have implications on inflows into the Malaprabha reservoir, water availability for irrigation and the ET in the command area.

4. Data extraction

The reanalysis data of the monthly mean atmospheric variables and fluxes of the study region prepared by National Centers for Environmental Prediction (NCEP; Kalnay *et al.*, 1996), are extracted for the period January 1978 to December 2000. The atmospheric variables are extracted for nine grid points whose latitude ranges from 12.5 to 17.5 °N, and longitude ranges from 72.5 to 77.5 °E at a spatial resolution of 2.5°. The atmospheric fluxes are extracted for 20 grid points whose latitude ranges from 12.3 to 20.0 °N and longitude ranges from 73.5 to 77.5 °E at a spatial resolution of approximately 1.9°.

The T_{max} and T_{min} are estimated at monthly time scale using records available from two temperature gauging stations. One of them is located in Santhebastwadi at 15°46′N latitude and 74°27′E longitude and the other is situated in Gadag at 15°25′N latitude and 75°38′E longitude. The gauging station at Santhebastwadi lies within the study region and data is available for the period January 1992 to December 2000. The station at Gadag, lies in the Malaprabha command area and data is available for the period January 1978 to December 2000. Primary source of the data is Water Resources Development Organization, Government of Karnataka, Bengalooru, India.

The GCM data used in the study are simulations obtained from CGCM3 of the Canadian Center for Climate Modeling and Analysis (CCCma), through its website http://www.cccma.bc.ec.gc.ca/. The data comprise 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 are extracted at monthly time scale for the period January 1978 to December

DOI: 10.1002/joc

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 grid spatial resolution of CGCM3 is uniform along the longitude with grid box size of 3.75° and nearly uniform along the latitude (approximately 3.75°). The spatial domain of climate variables is chosen as nine grid points. In general, the explanatory power of a given predictor will vary both spatially and temporally for a given predictand. The use of predictors directly overlying the target grid box fails to capture the strongest correlations (between predictor and predictand), as this domain may be geographically smaller in extent than the circulation domains of the predictors. Hence the comparison of different predictors with a larger spatial domain is found useful in downscaling as they may be critical factors affecting the realism and stationarity of the downscaled predictand (Wilby and Wigley, 2000). However, the correlation between the predictors and a given predictand vary both seasonally and geographically. The spatial domain selected is subjective to the predictor, predictand, season and geographical location and for this purpose no fixed rules are available. The nine grid points surrounding the study region are selected as the spatial domain of the predictors to adequately cover the various circulation domains of the predictors considered in this study. However, while working on location-based stratification, the spatial domain could be reduced to only land grid points as the predictand in the region is dominated by land effects. The GCM data and the information extracted on atmospheric fluxes is re-gridded to a common 2.5° using grid analysis and display system (GrADS; Doty and Kinter, 1993).

The development of downscaling models for each of the predictand variables $T_{\max}$ and $T_{\min}$, begins with selection of potential predictors, followed by training and validation of the SVM downscaling model. The developed model is then used to obtain projections of $T_{\max}$ and $T_{\min}$ from simulations of CGCM3.

5. Selection of the probable predictors

The selection of appropriate predictors for downscaling predictands 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 or index 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, such predictors are chosen as variables that are: (1) reliably simulated by GCMs and are readily available from archives of GCM output and reanalysis datasets, (2) strongly correlated with the predictand and (3) based on previous studies.

For this study, predictor variables which have a physically meaningful relationship with each of the two predictands ($T_{\max}$ and $T_{\min}$) are classified into three groups A, B and C. Large-scale atmospheric variables, namely air temperature, zonal and meridional wind velocities at 925 mb, which are often used for downscaling temperature, are considered as predictors in Group A. Surface flux variables namely LH, SH, shortwave radiation and longwave radiation fluxes fall in Group B. Group C comprises of all the predictor variables in both the Groups A and B. To the best of our knowledge, the predictors in Group B have not been considered for downscaling temperature in the past. In this study, these variables have been tried as they control the temperature of the earth's surface. The incoming solar radiation is the source of heating the surface, while LH flux, SH flux and longwave radiation will cool the surface.

Scatter plots and cross-correlations are in use to select predictors (Dibike and Coulibaly, 2006). In this study, scatter plots are prepared and cross-correlations are computed to investigate the presence of nonlinearity/linearity in dependence structure (1) between the predictor variables in NCEP and GCM datasets (Figures 3 and 4) and (2) between the predictor variables in NCEP dataset and each of the predictands (Figure 5). The cross-correlations are estimated using three measures of dependence namely, product moment correlation (Pearson, 1896), Spearman's rank correlation (Spearman, 1904a and b) and Kendall's tau (Kendall, 1951). Scatter plots and cross-correlations between each of the predictor variables in NCEP and GCM datasets are useful to verify if the predictor variables are realistically simulated by the GCM. The same between the predictor variables in NCEP dataset and each of the predictands are useful to verify if the predictor and predictand are well correlated.

6. Development of SVM downscaling model

This section outlines the procedure to develop a SVM model for downscaling temperature. A separate SVM model was developed for downscaling each predictand ($T_{\max}$ and $T_{\min}$). Further, each group of predictors (A, B and C) from each of the two domains (land, land and ocean) is considered as input to the model for downscaling each predictand. Furthermore, for downscaling $T_{\max}$, each model is calibrated using shorter and longer records to examine the sensitivity of performance of the model to the length of the record. Thus, 18 SVM models are developed, one for each combination of predictor group, predictand, calibration period and spatial domain of climate variables (Table III). The methodology used for developing all the 18 SVM downscaling models is unique as explained below.

The procedure for downscaling the predictands starts with the selection of seven predictors that are divided into Groups A, B and C. m_1 indicates the number of probable predictors in each group. For Groups A, B and C, the values of m_1 are 3, 4 and 7 respectively. Scatter plots and cross-correlation bar plots are used to study the predictors and their relationship with $T_{\max}$ and $T_{\min}$.

On an annual basis, the surface temperature difference between the hottest and coolest months is about 3 °C on

DOI: 10.1002/joc

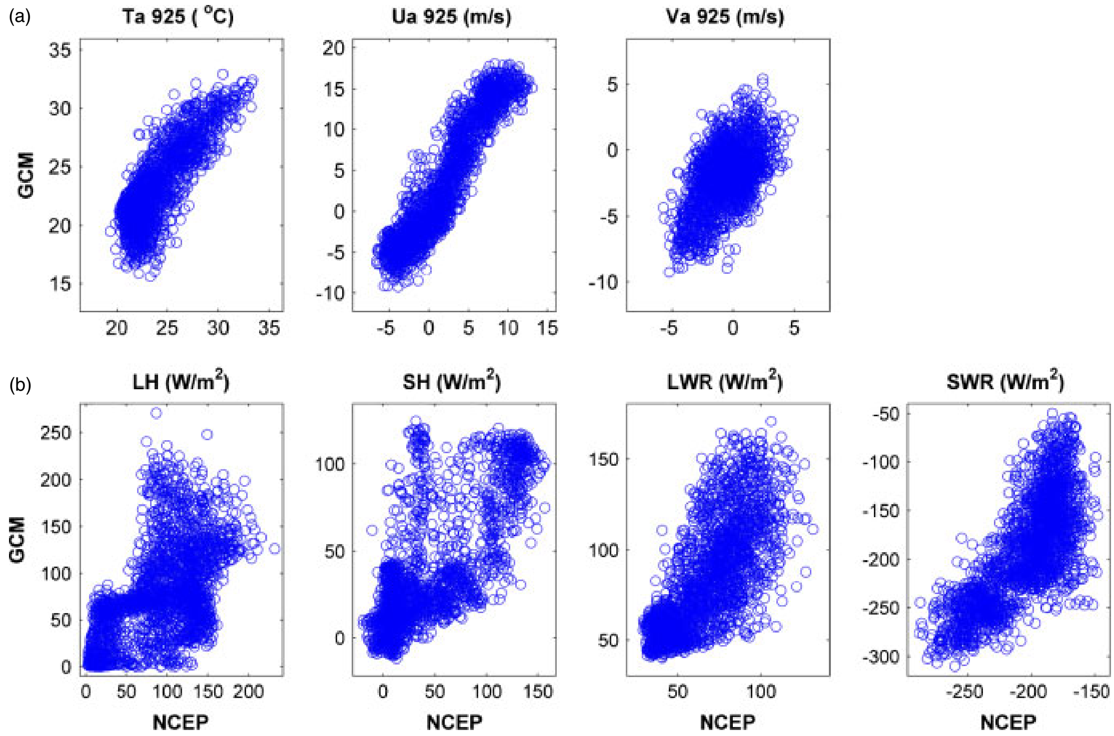

Figure 3. Scatter plots prepared to investigate dependence structure between probable predictor variables in NCEP and GCM datasets. (a) and (b) denote plots based on Group A and Group B predictors, respectively. In each plot, ordinate denotes GCM value of predictor variable, whereas abscissa represents NCEP value of the predictor variable. This figure is available in colour online at www.interscience.wiley.com/ijoc

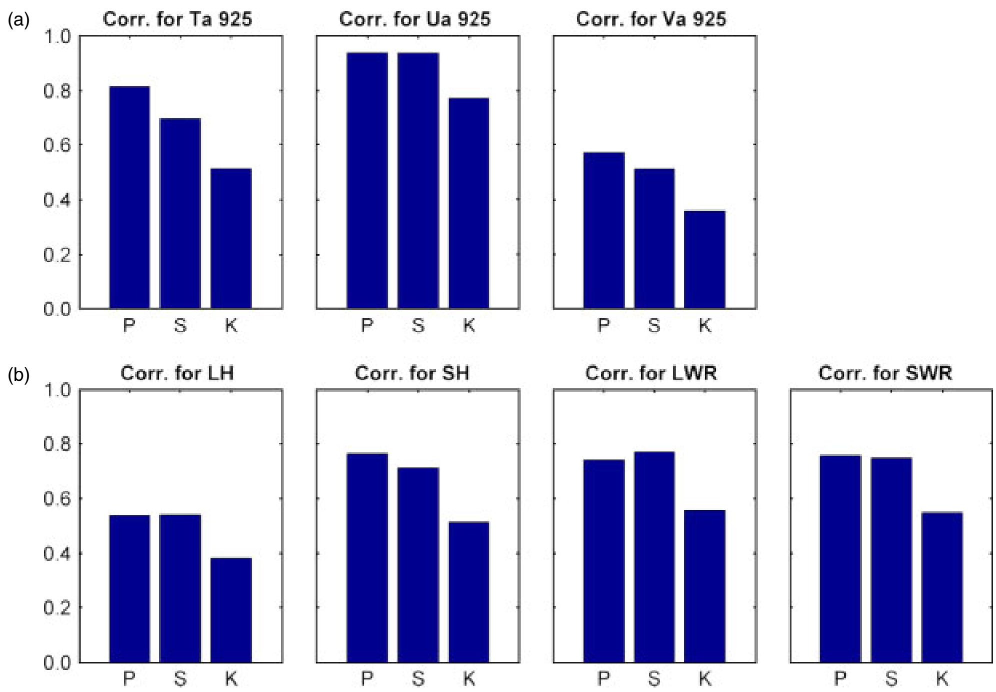

Figure 4. Bar plots for cross-correlation computed between probable predictors in NCEP and GCM datasets. (a) and (b) denote plots based on Group A and Group B predictors respectively. P, S and K represent product moment correlation, Spearman's rank correlation and Kendall's tau respectively. This figure is available in colour online at www.interscience.wiley.com/ijoc

the oceans and about 8 °C on land. On the other hand, the same at 925 mb is about 6 °C on oceans and about 8 °C on land. Therefore, in the second step, to assess the effect of variation of temperature patterns on land and sea, location-based stratification was carried out to form two domains, one comprising of predictor variables pertaining to only land grid points (number of grid points = 6) and the other containing those pertaining to both ocean and land (number of grid points = 9). From the m_1 probable predictors, m_2 potential predictors for downscaling are

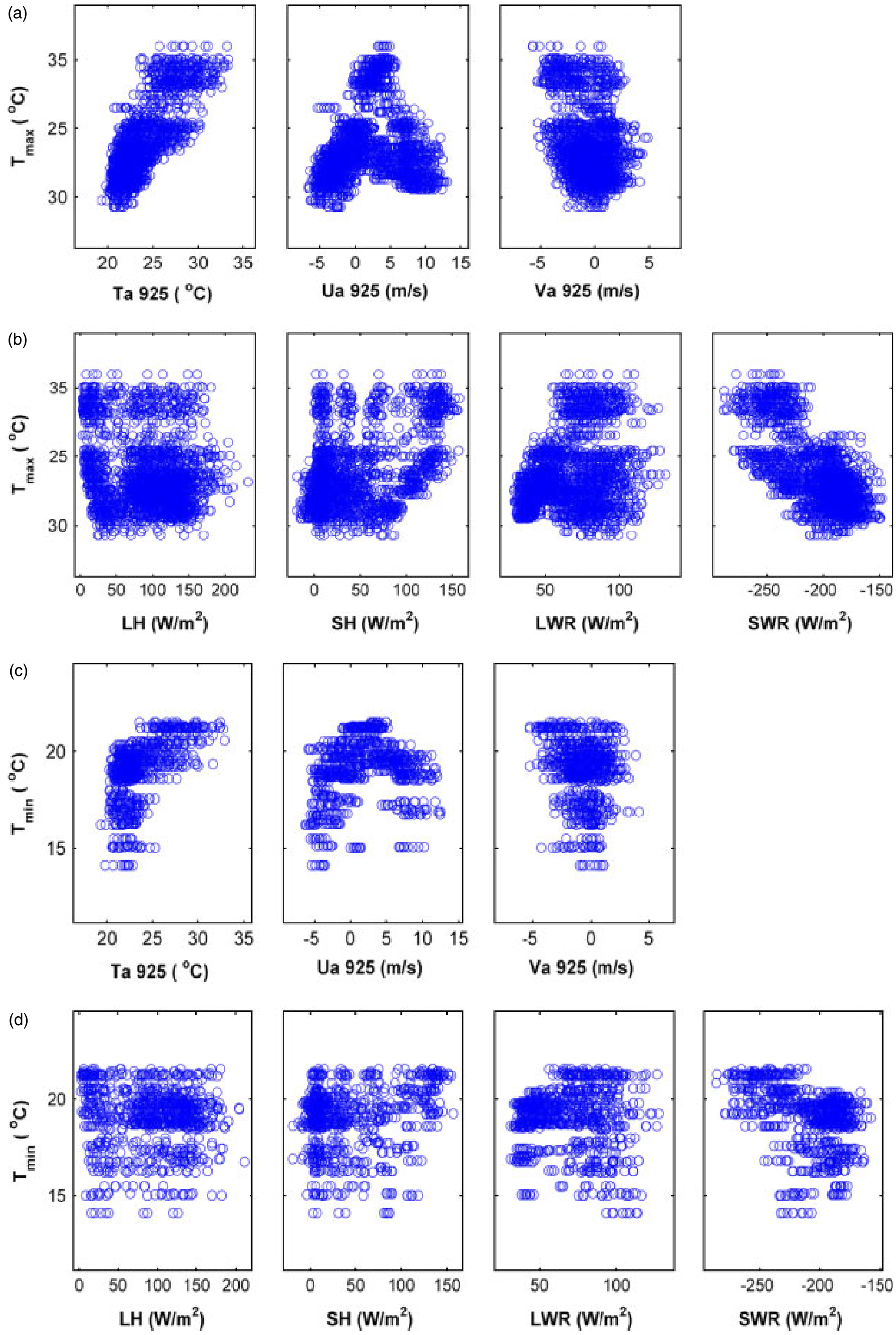

Figure 5. Scatter plots prepared to investigate dependence structure between probable predictor variables in NCEP data and the observed T_{max} and T_{min}. (a) and (b) denote plots based on Group A and Group B predictors, respectively, for the predictand T_{max}, while (c) and (d) denote plots based on Group A and Group B predictors, respectively, for the predictand T_{min}. This figure is available in colour online at www.interscience.wiley.com/ijoc

selected by specifying two threshold values (T_{ng1} and T_{np}). For example, for Group A $m_2 = 3 \times$ number of grid points, and for Group B $m_2 = 4 \times$ number of grid points. The T_{ng1} is for cross-correlation between NCEP and GCM datasets, whereas the same between NCEP and predictand datasets is T_{np}. The three dependence measures (product moment correlation, Spearman's rank correlation and Kendall's tau) were considered for computation of cross-correlation. The m_2 predictors with correlations above the threshold values are selected as the potential predictors. The data of potential predictors is first standardized. Standardization is widely used prior to

DOI: 10.1002/joc

Table III. Different SVM downscaling model variants used in the study for obtaining projections of predictands T_{max} and T_{min}.

Predictand	Predictor	Spatial domain	Time period of downscaling	Calibration period	Model number
T_{max}	Group A	Land (small domain)	1992–2100	1992–1997	Model 1
			1978–2100	1978–1993	Model 2
		Land + sea (large domain)	1992–2100	1992–1997	Model 3
			1978–2100	1978–1993	Model 4
	Group B	Land (small domain)	1992–2100	1992–1997	Model 5
			1978–2100	1978–1993	Model 6
		Land + sea (large domain)	1992–2100	1992–1997	Model 7
			1978–2100	1978–1993	Model 8
	Group C	Land (small domain)	1992–2100	1992–1997	Model 9
			1978–2100	1978–1993	Model 10
		Land + sea (large domain)	1992–2100	1992–1997	Model 11
			1978–2100	1978–1993	Model 12
T_{min}	Group A	Land (small domain)	1992–2100	1992–1997	Model 13
		Land + sea (large domain)	1992–2100	1992–1997	Model 14
	Group B	Land (small domain)	1992–2100	1992–1997	Model 15
		Land + sea (large domain)	1992–2100	1992–1997	Model 16
	Group C	Land (small domain)	1992–2100	1992–1997	Model 17
		Land + sea (large domain)	1992–2100	1992–1997	Model 18

statistical downscaling to reduce bias (if any) in the mean and the variance of GCM predictors with respect to that of NCEP-reanalysis data (Wilby *et al*., 2004). The procedure typically involves subtraction of mean and division by the standard deviation of the predictor. The data of standardized NCEP predictor variables is then processed using principal component analysis to extract principal components (PCs) which are orthogonal and which preserve more than 98% of the variance originally present in it. A feature vector is formed for each month of the record using the PCs. The feature vector is the input to the SVM model, and the contemporaneous value of predictand is the output. The PCs account for most of the variance in the input data and are also independent of each other. Hence, the use of PCs as input to a downscaling model helps in making the model more stable and also reduces the computational burden.

To develop the SVM downscaling model, the feature vectors which are prepared from NCEP record are partitioned into a training set and a test set. The training set comprises approximately the first 75% of the feature vectors, and the remaining form the test set. Feature vectors in the training set are used for calibrating the model, and those in the test set are used for validation. The normalized mean squared error (NMSE) under validation 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 radial basis function (RBF) kernel σ provides an idea of the smoothness of the derived function. Smola *et al*. (1998), while explaining the regularization capability of the RBF kernel, have shown that a large kernel width acts as a low-pass filter in frequency domain. It attenuates higher-order frequencies, resulting in a smooth function. On the other hand, 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 the optimum ranges for the parameters. Subsequently, the optimum values of parameters are obtained from the selected ranges using stochastic search technique of genetic algorithm (Haupt and Haupt, 2004).

The feature vectors that are 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 (i.e. A1B, A2, B1 and COMMIT). Subsequently, for each scenario, the projected values of predictand are segregated into five parts (2001–2020, 2021–2040, 2041–2060, 2061–2080 and 2081–2100) to determine the future trend in projections.

The performance of the developed SVM models is evaluated using the following statistical measures and product moment correlation coefficient (CC).

1. Sum of squares of errors (SSE), defined as

$$\mathrm{SSE} = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{1}$$

2. Mean square error (MSE), given as

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2}$$

3. Root mean square error (RMSE), defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{3}$$

DOI: 10.1002/joc

4. NMSE, (Zhang and Govindaraju, 2000), given as

$$\text{NMSE} = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{(S_{\text{obs}})^2} \tag{4}$$

5. Nash-Sutcliffe error estimate (E_f, Nash and Sutcliffe, 1970), defined as

$$E_f = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y_i})^2} \tag{5}$$

6. Mean absolute error (MAE, Johnson *et al*., 2003), given as

$$\text{MAE} = 1 - \frac{\sum_{i=1}^{N}|(y_i - \hat{y}_i)|}{\sum_{i=1}^{N}|(y_i - \overline{y_i})|} \tag{6}$$

7. Mean cumulative error (MCE, Johnson *et al*., 2003), defined as

$$\text{MCE} = 1 - \left| \sqrt{\frac{\sum_{i=1}^{N}\hat{y}_i}{\sum_{i=1}^{N}y_i}} - \sqrt{\frac{\sum_{i=1}^{N}y_i}{\sum_{i=1}^{N}\hat{y}_i}} \right| \tag{7}$$

where N represents the number of feature vectors prepared from the NCEP record, y_i and $\hat{y}_i$ denote the observed and the simulated values of predictand respectively, $\overline{y_i}$ and S_{obs} are the mean and the standard deviation of the observed predictand.

7. Results and discussion

Downscaling models are developed following the methodology described in Sections 5 and 6. The results and discussion are presented in this section.

7.1. Probable predictor selection

The most relevant probable predictor variables necessary for developing the SVM downscaling model are identified by using scatter plots and the three measures of dependence following the procedure described in Section 5. The scatter plots and cross-correlations enable verifying the reliability of the simulations of the predictor variables by the GCM and to study the predictor–predictand relationships. For Groups A and B, the scatter plots between the probable predictor variables in NCEP and GCM datasets are shown in Figure 3, while the cross-correlations computed between the same are shown in Figure 4. In general, the predictor variables in Groups A and B are realistically simulated by the GCM. From the scatter plots shown in Figure 3, it can be inferred that predictors in Group A are simulated better than those in Group B by the GCM. Further, it is noted that zonal wind velocity at 925 mb (Ua 925) is the most realistically simulated variable with a CC greater than 0.9, while LH flux is the least correlated variable between NCEP and GCM datasets (CC = 0.56; Figures 3 and 4). It is to be noted that these figures represent how well the predictors simulated by NCEP and GCM are correlated. Generally, the correlations are not very high due to the differences in the simulations of GCM (e.g. for different runs) and possible errors in NCEP-reanalysis. In addition, the inherent errors due to re-gridding from GCM scale to NCEP scale also contribute to low correlation.

To investigate the relationship between the probable predictors and predictands, scatter plots and cross-correlation bar plots between the probable predictor variables in NCEP data and each of the predictands ($T_{\max}$ and $T_{\min}$) are presented in Figures 5 and 6 respectively. From a perusal of the scatter plots, it appears that the linear dependence structure between predictor variables and predictands is weaker for $T_{\min}$ when compared to $T_{\max}$. From the two figures, it can be observed that Ta 925 and shortwave radiation (SWR) flux have high correlation with both the predictands, while Ua 925, Va 925, LH and longwave radiation (LWR) fluxes have less correlation with the same. Ta 925, Ua 925, SH, and LWR have a positive correlation with both $T_{\max}$ and $T_{\min}$. LH, Va 925 and SWR have a negative correlation with both the predictands. Among the two predictands, the $T_{\max}$ is more correlated with the predictors.

The predictors can be ranked based on the relative magnitude of cross-correlations estimated by each measure of dependence. Results show similar (or nearly equal) rank for any chosen predictor by all the three dependence measures considered, indicating that the results are reliable. The results of this analysis indicate that Ta 925 is a better predictor in the Group A, while SWR and SH are better predictors in the Group B, while all these three (Ta 925, SWR and SH) are better predictors in the Group C, since Group C is a combination of predictors in Groups A and B. These results give an overall picture of relationships between predictors and predictands over all the nine grid points considered.

7.2. Analysis of selected GCM and NCEP probable predictors

At each of the NCEP grid points, the trend in the GCM data and bias in the mean and variance of the same relative to that of the NCEP data are assessed using box plots for the period 1992–2100. The span of the box represents the interquartile range of the predictor variable. The whiskers extend from the box to 5 and 95% quantiles on the lower and the upper side of the box, respectively. In Figure 7, typical results of the box

DOI: 10.1002/joc

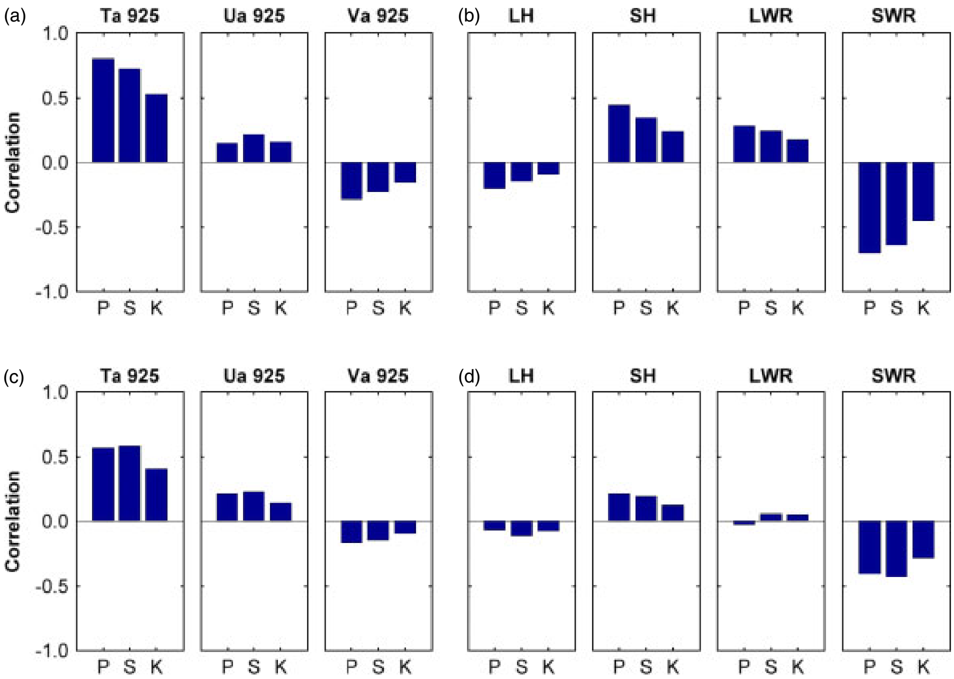

Figure 6. Bar plots for cross-correlation computed between probable predictors in NCEP data and observed T_{max} and T_{min}. (a) and (b) denote plots based on Group A and Group B predictors, respectively, for the predictand T_{max}, while (c) and (d) denote plots based on Group A and Group B predictors, respectively, for the predictand T_{min}. P, S and K represent product moment correlation, Spearman's rank correlation and Kendall's tau, respectively. This figure is available in colour online at www.interscience.wiley.com/ijoc

plots that are prepared by using NCEP and GCM data at NCEP grid point 5, are presented in (i). The same results using GCM data for the future (2001–2100), for the four scenarios A1B, A2, B1 and COMMIT are shown in (ii), (iii), (iv) and (v), respectively (Figure 7).

The impact of the temporal trend in predictor variables on downscaled temperature was studied. For a variable, the trend is determined by comparing the mean of the historical (observed) values with the mean estimated for future projections simulated by GCM, using 20-year intervals (2001–2020, 2021–2040, 2041–2060, 2061–2080 and 2081–2100). It can be seen from Figure 7(a) that the predictor variable, Ta 925, shows an increasing trend, while the rest of the predictors show no trend. The projected increase in Ta 925 is high for A2 scenario (Figure 7(a) (iii)), while it is least for B1 scenario (Figure 7(a) (iv)), whereas no trend is discerned with the COMMIT scenario (Figure 7(a) (v)). This is because among the scenarios considered, the scenario A2 has the highest concentration of carbon dioxide (CO_2) of 850 ppm, while the same for A1B, B1 and COMMIT scenarios are 720, 550 and ≈370 ppm respectively. Rise in the concentration of CO_2 in the atmosphere causes the earth's average temperature to increase. In the COMMIT scenario, where emissions are kept at the same levels as in the year 2000, no significant trend in the pattern of projected future temperature could be discerned. Analysis of land surface temperature data extracted from GCM shows a similar trend as Ta 925 for all the scenarios.

Mean, and variance (which is reflected by interquartile range of each box in the box plot) estimated for each of the probable predictor variables in NCEP and GCM datasets are presented in part (i) of Figure 7 for grid point 5, for brevity. Bias is seen in the mean and the variance of the GCM data relative to the NCEP data for almost all the predictor variables. The magnitude of this bias is found to vary from one predictor to another, and from one grid point to another. The mean statistic estimated for Va 925, LH and SH fluxes simulated by the GCM is deflated with respect to that estimated for the respective NCEP variables. On the other hand, the statistic computed for SWR and LWR simulated by the GCM are inflated. Further, it may be noted that the interquartile ranges for Ua 925, Va 925, SWR and LWR simulated by the GCM are large compared to those for respective NCEP variables. The relative bias observed for predictor variables in Group A is less than that estimated for the variables in Group B. This is in agreement with observations based on visual interpretation of scatter plots (Figure 3). Hence the standardization of predictor variables prior to developing the downscaling models is justified. The standardization is useful to reduce bias in the mean and variance of GCM predictors relative to NCEP data, while maintaining the trend in the predictor variables.

7.3. Selection of the potential predictors

For downscaling each of the two predictands (T_{max} and T_{min}), the potential predictor variables are identified for each group of probable predictors by using scatter plots and the three measures of dependence described in Section 6. The selected potential predictors, which are listed in Table IV, are used to develop the SVM downscaling models. From the Table it can be observed that air

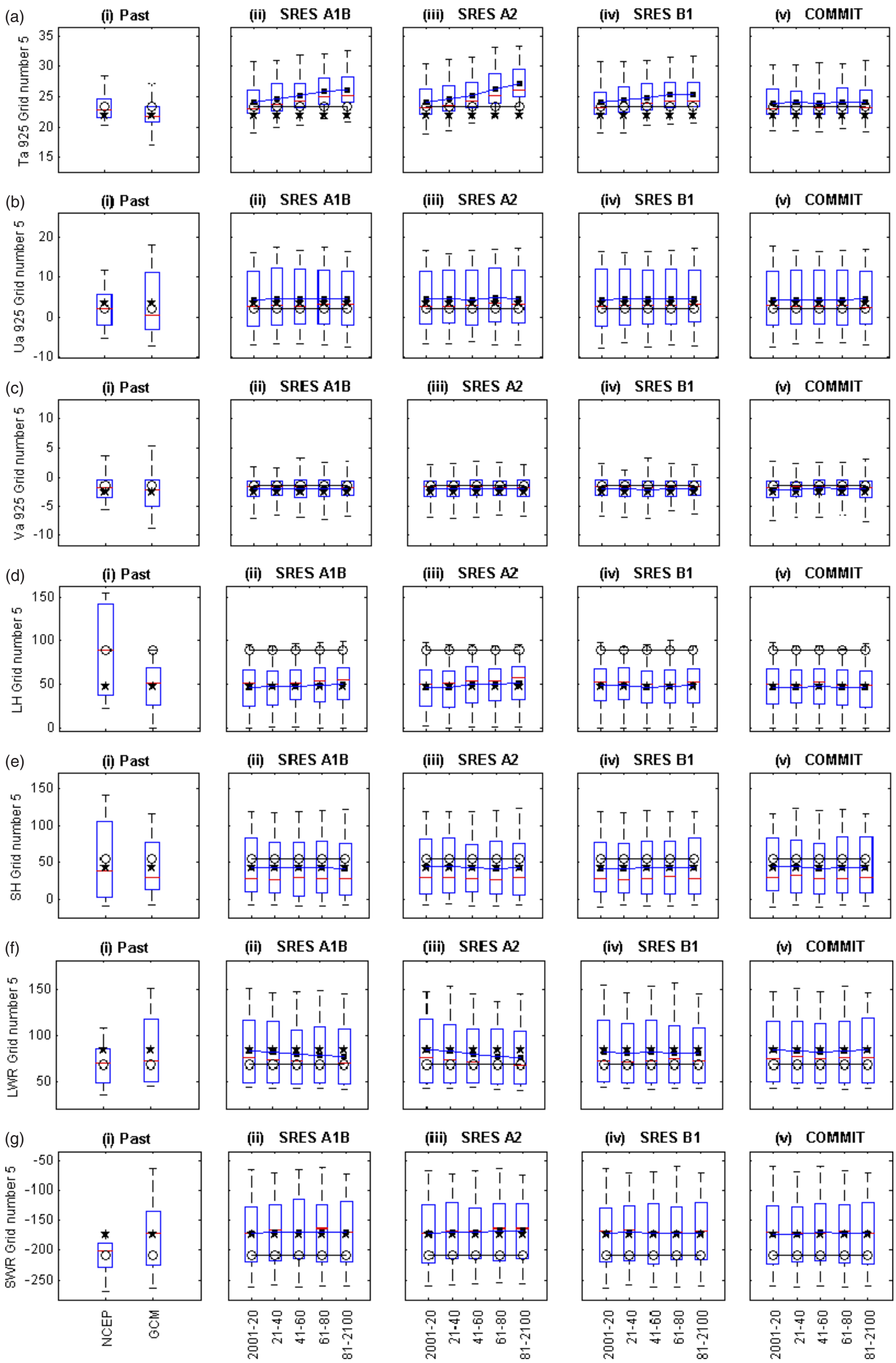

Figure 7. Typical results for determining the trend of the predictor variables. Air temperature at 925 mb, zonal wind velocity at 925 mb, meridional wind velocity at 925 mb, latent heat flux, sensible heat flux, longwave radiation flux, and shortwave radiation flux for grid point 5 are denoted as (a), (b), (c), (d), (e), (f) and (g), respectively. The horizontal line in the middle of the box represents median. The circle and star denote the mean values of predictor variable for NCEP and GCM datasets respectively. The gap between star and circle denotes bias in the predictor. The line joining squares depicts the mean trend projected by GCM for the predictor variable. In (ii), (iii), (iv) and (v) the line that joins the circles indicates the historical trend of the predictor variable. This figure is available in colour online at www.interscience.wiley.com/ijoc

temperatures and meridional wind velocities at 925 mb are selected as potential predictors from Group A. For downscaling maximum temperature SH, longwave and shortwave radiation fluxes are selected as potential predictors from Group B, whereas air temperatures and meridional wind velocities at 925 mb, SH and shortwave

DOI: 10.1002/joc

Table IV. List of probable and potential predictors selected for use in this study for downscaling T_{max} and T_{min}. The model numbers are defined in Table III.

Model number	Probable predictors	Potential predictors selected	
		Names	NCEP grid points
1	ta_925, ua_925, va_925	ta_925	2,3,5,6
		va_925	2,5
2	ta_925, ua_925, va_925	ta_925	2,3,5,6
		va_925	2,5
3	ta_925, ua_925, va_925	ta_925	1,2,3,4,5,6,7,8,9
		va_925	1,2,4,5,7,8,9
4	ta_925, ua_925, va_925	ta_925	1,2,3,4,5,6,7,8,9
		va_925	1,2,4,5,7,8,9
5	LH, SH, LWR, SWR	SH	2,3,5,6
		LWR	2,6
		SWR	2,3,5,6
6	LH, SH, LWR, SWR	SH	2,3,5,6
		LWR	2,3,6
		SWR	2,3,5,6
7	LH, SH, LWR, SWR	SH	1,2,3,5,6,8,9
		LWR	3,6
		SWR	1,2,3,4,5,6,7,8,9
8	LH, SH, LWR, SWR	SH	1,2,3,5,6,8,9
		LWR	3,6,9
		SWR	1,2,3,4,5,6,7,8,9
9	ta_925, ua_925, va_925, LH, SH, LWR, SWR	SH	2,3,5,6
		SWR	2,3,5,6
		ta_925	2,3,5,6
		va_925	2
10	ta_925, ua_925, va_925, LH, SH, LWR, SWR	SH	2,3,5,6
		SWR	2,3,5,6
		ta_925	2,3,5,6
		va_925	2
11	ta_925, ua_925, va_925, LH, SH, LWR, SWR	SH	1,2,3,5,6,8,9
		SWR	1,2,3,4,5,6,7,8,9
		ta_925	2,3,4,5,6,7,8,9
		va_925	2
12	ta_925, ua_925, va_925, LH, SH, LWR, SWR	SH	1,2,3,5,6,8,9
		SWR	1,2,3,4,5,6,7,8,9
		ta_925	2,3,4,5,6,7,8,9
		va_925	2
13	ta_925, ua_925, va_925	ta_925	2,3,5,6
		va_925	2,5
14	ta_925, ua_925, va_925	ta_925	1,2,3,4,5,6,7,8,9
		va_925	1,2,4,5,7,9
15	LH, SH, LWR, SWR	SH	2,3,6
		SWR	2,3,5,6
16	LH, SH, LWR, SWR	SH	1,2,3,6
		SWR	1,2,3,4,5,6,7,8,9
17	ta_925, ua_925, va_925, LH, SH, LWR, SWR	ta_925	2,3,6
		SWR	2,3,5,6
18	ta_925, ua_925, va_925, LH, SH, LWR, SWR	ta_925	2,3,4,5,6,7,8,9
		SWR	1,2,3,6

radiation fluxes are selected as potential predictors from Group C. The decision on selection of these potential predictors is further justified by scatter plots, and the cross-correlation bar plots for the three measures of dependence prepared for this purpose, but not shown here to save space.

7.4. Developing SVM downscaling models

From the standardized data of potential predictors, PCs are extracted to form feature vectors. These feature vectors are provided as input to develop SVM downscaling model following the procedure described in Section 6.

For obtaining the optimal range of values of SVM parameters, kernel width (σ) and penalty term (C), the grid search procedure is used. Typical results of this analysis are presented in Figure 8. From this figure, the ranges of σ and C having the least NMSE are selected as the optimum parameter ranges. The NMSE values are indicated in the bar code provided close to the figure. Using genetic algorithm, the optimum value of each parameter is selected from its optimum range. For each of the 18 SVM models developed, the selected parameters are shown in Table V.

Typical results of downscaled predictands (T_{max} and T_{min}) obtained from the three groups of predictors are presented in Figures 9 and 10. In part (i) of these figures, the T_{max} and T_{min} downscaled using NCEP and GCM datasets are compared with the observed T_{max} and T_{min} for the study region using box plots. The projected precipitation for 2001–2020, 2021–2040, 2041–2060, 2061–2080 and 2081–2100, for the four scenarios A1B, A2, B1 and COMMIT are shown in (ii), (iii), (iv) and (v) respectively.

Table V. Parameters of SVM downscaling models developed in this study. The model numbers are defined in Table III.

SN	Model number	SVM model parameter	
		Kernel width (σ)	Penalty term (C)
1	1	2050	2050
2	2	2050	2050
3	3	2050	2050
4	4	2050	2050
5	5	50	2050
6	6	50	250
7	7	50	250
8	8	50	2050
9	9	250	850
10	10	450	850
11	11	2050	2050
12	12	250	450
13	13	1050	50
14	14	1050	50
15	15	50	1050
16	16	1050	50
17	17	4050	4050
18	18	1050	50

7.5. Performance of the downscaling models

In this section, investigations are carried out to study three aspects. The first is assessment of the effect of length of calibration period on performance of the downscaling model, and the second is assessment of impact of location-based stratification of predictor variables on downscaling. The sensitivity of the SVM models to the different groups of predictors is the third aspect examined.

On an annual basis, the surface temperature difference between the hottest and coolest months is about 3 °C on the oceans and about 8 °C on land. On the other hand, the same at 925 mb is about 6 °C on oceans and about 8 °C on land. To assess the effect of this variation on the results of downscaling, location-based stratification was carried out to form two domains, one comprising of predictor variables pertaining to only land grid points, and the other containing those pertaining to both ocean and land.

To address the first aspect, the observed records of temperature at two stations are analysed. Santhebastewadi gauging station is located in the study region and has a shorter period of record (1992–2000). Gadag gauging station is located just outside the study region and has a longer period of record (1978–2000). The cross-correlation between contemporaneous records of T_{max} at these stations is found to be high. Therefore, a relationship is established between the contemporaneous records of T_{max} at these stations. This relationship is used to obtain correlative estimates of monthly T_{max} and T_{min} for the missing period for Santhebasthewadi station from the records of Gadag station. Details of the procedure adopted are available in Gupta (1989).

From the results presented in Table VIA and B it can be observed that increasing the period of calibration from

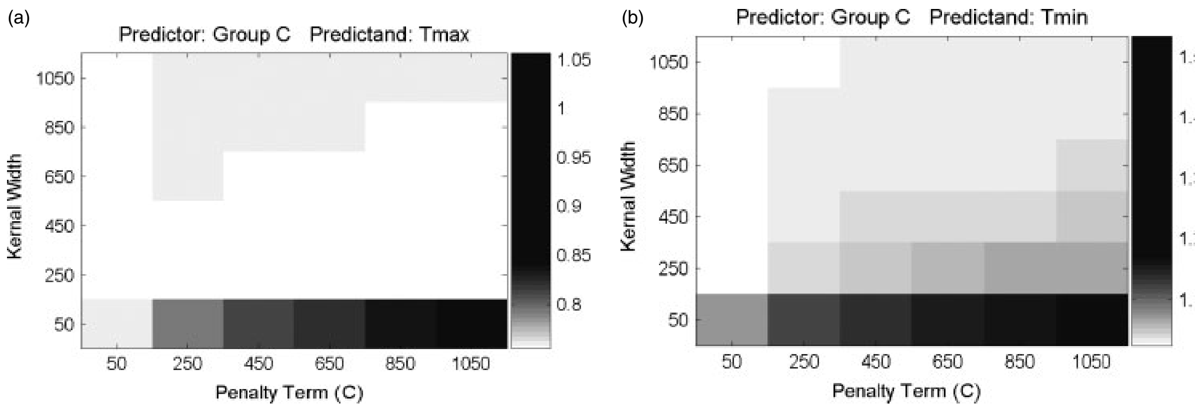

Figure 8. Typical results of the domain search to estimate optimal values of the parameters (kernel width, σ; penalty, C) for downscaling T_{max} and T_{min} from predictor variables in Group C are shown as (a) and (b) respectively. The bar code shows the NMSE values. The ranges of parameters for which NMSE is least are selected.

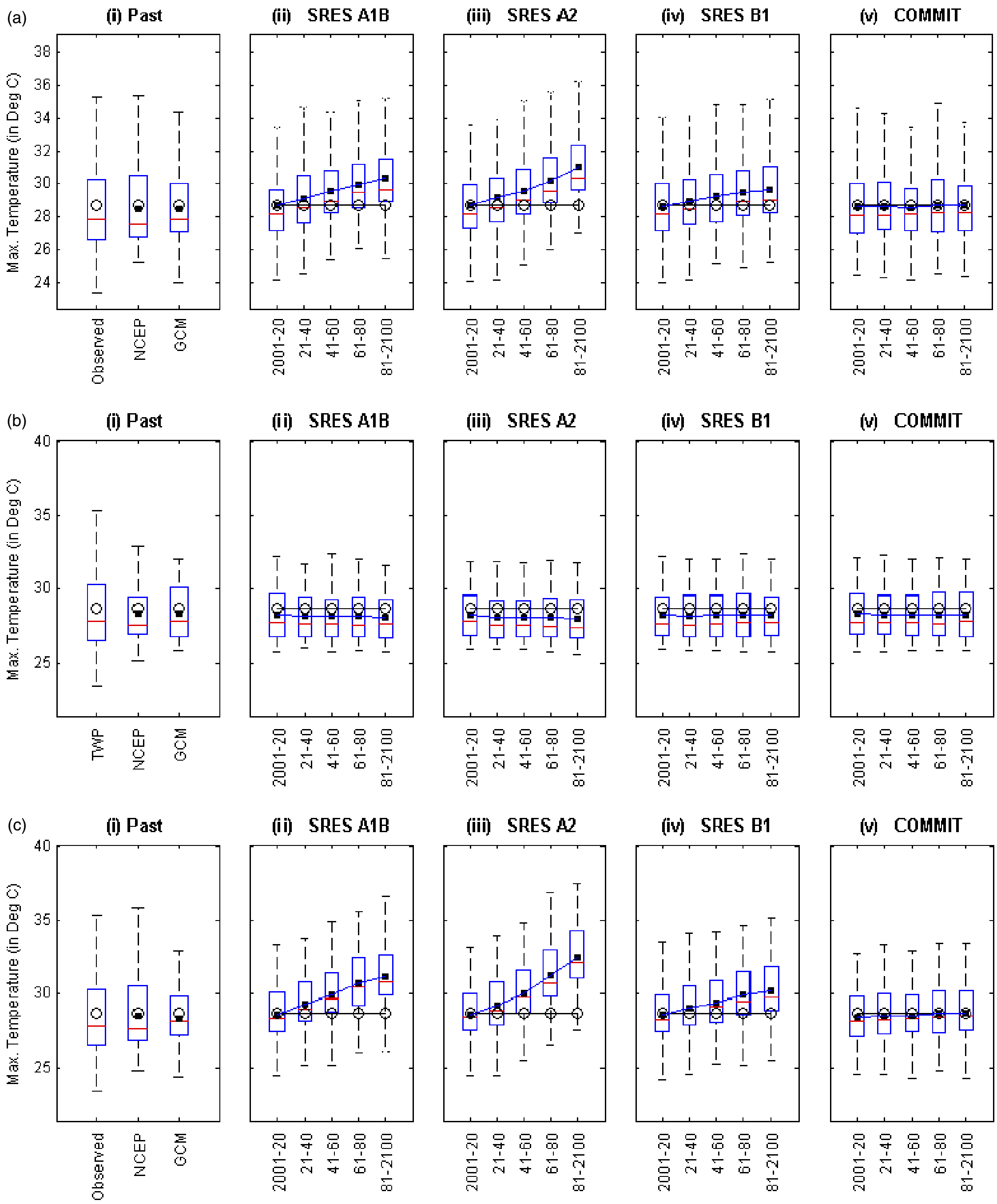

Figure 9. Typical results from the SVM-based downscaling model graphed using box plots for the predictand T_{max}. (a), (b) and (c) denote results based on Group A, Group B and Group C predictors respectively. The horizontal line in the middle of the box represents median. The circles denote the mean value of T_{max}, and the darkened square represents the mean value of simulated T_{max}. The gap between darkened square and circle denote bias in the T_{max} 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 T_{max}, while the line connecting the solid squares depicts the mean trend of T_{max} projected by GCM. This figure is available in colour online at www.interscience.wiley.com/ijoc

6 to 16 years did not result in significant improvement in the performance of the downscaling model. These results indicate that a smaller period of records at Santhebasthewadi station would as well be sufficient to develop an efficient downscaling model using SVM that implements the structural risk minimization principle by striking a right balance between the training error and the ability of the machine to learn any training set without error (Tripathi *et al.*, 2006). Hence, for predictand T_{min}, Santhebasthewadi station data alone was used to develop the downscaling model.

To address the second aspect, the results of downscaling obtained using each of the two domains (land; land and ocean) of the climate variables for each combination of predictor group and predictands are shown in Table VIC and D. It can be seen that use of predictor variables from the smaller spatial domain covering the land area improves the overall performance of the

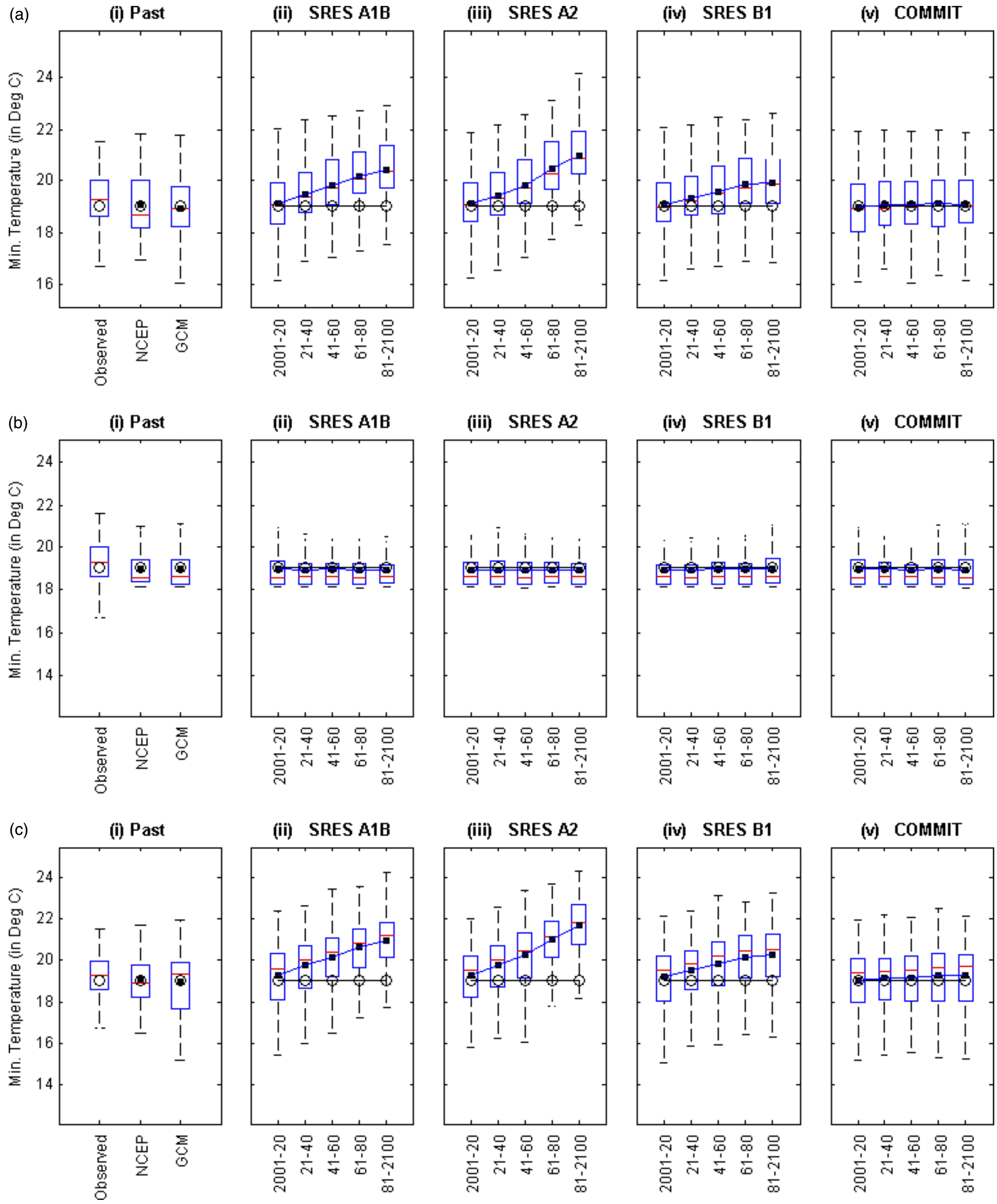

Figure 10. Typical results from the SVM-based downscaling model graphed using box plots for predictand T_{min}. (a), (b) and (c) denote results based on Group A, Group B and Group C predictors, respectively. The horizontal line in the middle of the box represents median. The circles denote the mean value of T_{min}, and the darkened square represents the mean value of simulated T_{min}. The gap between darkened squares and circles denote bias in the T_{min} simulated by the downscaling model for NCEP and GCM datasets. In (ii), (iii), (iv) and (v) the solid line that joins the circles indicates the historical trend of T_{min}, while the line connecting the solid squares depicts the mean trend of T_{min} projected by GCM. This figure is available in colour online at www.interscience.wiley.com/ijoc

downscaling models. These results are strengthened by the fact that variations in patterns of temperature at the earth's surface and at 925 mb are different for the land and the ocean (Table VII).

Finally, to address the third aspect, the sensitivity of the SVM models to the predictor group is studied. The SVM models developed to downscale T_{max} using Group C predictors (Models 9–12) are seen to perform better than those developed based on predictors in the other groups, for both small and large spatial domains. This implies that both surface flux variables and large-scale atmospheric variables have to be considered as predictors for effective downscaling of T_{max}.

Overall, the results of the SVM downscaling models indicate that between the two predictands, T_{max} is better simulated than T_{min} (Figures 11 and 12).

7.6. Impact of trend in predictor variables on downscaled temperature

From the box plots of downscaled predictands (Figures 9 and 10), it can be observed that T_{max} and T_{min} are projected to increase in future for A1B, A2 and B1 scenarios,

Table VIA. Error statistics computed for T_{max} downscaled from predictor variables pertaining to land and ocean grid points using the entire record. Minimum values of MSE, RMSE, NMSE, and maximum of E_f, MAE, MCE and CC indicate optimal values of error statistics. The model numbers are defined in Table III.

Length of record	Model no	MSE	RMSE	NMSE	E_f	MAE	MCE	CC
1992–2000	3	0.7941	0.8911	0.1069	0.8921	0.6792	0.9985	0.9450
	7	0.9147	0.9564	0.1231	0.7570	0.6661	0.9963	0.9380
	11	0.7140	0.8450	0.0961	0.9030	0.7078	0.9998	0.9510
1978–2000	4	0.7765	0.8812	0.1095	0.8901	0.6776	0.9997	0.9440
	8	0.9237	0.9611	0.1303	0.8693	0.6557	0.9974	0.9330
	12	0.7152	0.8457	0.1009	0.8988	0.7025	0.9986	0.9480

Table VIB. Error statistics computed for T_{max} downscaled from predictor variables pertaining to land grid points using the entire record.

Length of record	Model no	MSE	RMSE	NMSE	E_f	MAE	MCE	CC
1992–2000	1	0.8901	0.9434	0.1198	0.8791	0.6354	0.9983	0.9380
	5	1.0477	1.0236	0.1410	0.8576	0.6221	0.9979	0.9270
	9	0.7439	0.8625	0.1001	0.8989	0.7093	0.9996	0.9480
1978–2000	2	1.1886	1.0902	0.1676	0.8318	0.5743	0.9993	0.9120
	6	0.8900	0.9443	0.1258	0.8738	0.6568	0.9979	0.9360
	10	0.7612	0.8725	0.1073	0.8923	0.6757	0.9994	0.9420

Table VIC. Error statistics computed for downscaled predictand T_{max} for different spatial domains of predictor variables for the validation period.

Spatial domain	Model no	SSE	MSE	RMSE	NMSE	E_f	MAE	MCE	CC
Land (small domain)	1	28.41	1.1836	1.0879	0.1945	0.7971	0.5348	0.9923	0.8983
	5	31.84	1.3267	1.1518	0.2180	0.7725	0.5431	0.9971	0.9060
	9	41.87	1.7446	1.3208	0.2867	0.7009	0.4681	0.9921	0.8439
Land + ocean (large domain)	3	29.70	1.2376	1.1125	0.2033	0.7878	0.5312	0.9978	0.8948
	7	33.24	1.3851	1.1769	0.2276	0.7625	0.5399	0.9972	0.9182
	11	34.22	1.4262	1.1942	0.2343	0.7555	0.5202	0.9984	0.8701

Table VID. Error statistics computed for downscaled predictand T_{min} for different spatial domains of predictor variables for the validation period.

Spatial domain	Model no	SSE	MSE	RMSE	NMSE	E_f	MAE	MCE	CC
Land (small domain)	13	112.9514	1.0458	1.0227	0.5788	0.4173	0.3719	0.9987	0.761
	15	171.7722	1.5905	1.2611	0.6347	0.3594	0.1430	0.9946	0.604
	17	136.9800	1.2683	1.1262	0.5061	0.4891	0.2539	0.9991	0.703
Land + ocean (large domain)	14	141.5507	1.3107	1.1448	0.5230	0.4721	0.2733	0.9971	0.693
	16	184.4284	1.7077	1.3068	0.6814	0.3122	0.1232	0.9972	0.560
	18	246.1000	0.8900	0.9443	0.1258	0.8738	0.6568	0.9979	0.936

Note: Optimal values of error statistics are highlighted in grey. They are used to identify the SVM model providing best performance.
SSE, sum of squares of errors; MSE, mean square error; RMSE, root mean square error; NMSE, normalized mean square error; E_f, Nash-Sutcliffe error estimate; MAE, mean absolute error; MCE, mean cumulative error; CC, correlation coefficient.

Table VII. Mean monthly temperatures computed using the records at nine NCEP grid points in the study region.

Month	Mean monthly temperature in °C			
	At 925 mb		Earth's surface	
	Land	Ocean	Land	Ocean
Jan	22.64	22.24	23.48	26.91
Feb	24.50	22.60	25.54	26.69
Mar	28.02	25.26	28.64	27.70
Apr	29.48	26.30	30.11	28.70
May	29.42	25.57	29.74	29.02
Jun	25.08	22.60	25.66	27.98
Jul	22.35	21.18	23.39	26.98
Aug	21.87	20.81	23.05	26.57
Sep	22.52	21.61	23.33	26.89
Oct	22.88	22.55	23.55	27.52
Nov	22.55	22.76	23.32	28.05
Dec	21.86	22.17	22.52	27.37

whereas no trend is discerned with the COMMIT scenario by using predictors in Groups A and C. The projected increase in predictands is high for A2 scenario, whereas it is least for B1 scenario. In contrast, projections for the predictands using the predictors in Group B did not show any trend for the SRES scenarios.

No trend is seen in the predictands that are projected using predictors in Groups A and C, when Ta 925 was excluded from the predictor groups. Therefore, the projected increase in trend of predictands for the Groups A and C is attributed to the increasing trend evident in Ta 925.

As the SVM downscaled predictand is affected by trend in the predictors, this trend should be compared with the trend in the predictand over historical and future time periods considered. For this purpose, the trend in land surface maximum and minimum temperature data extracted from GCM for the period 1978–2100 was analysed for each of the scenarios considered in the study. The results show a similar trend as the predictor variable Ta 925 extracted from GCM, for all the scenarios considered. Thus it is essential to consider Ta 925 as a predictor for downscaling the predictands. Herein it is to be mentioned that the GCM-simulated values are not considered acceptable because of the coarse resolution of the model. However, the trend in the GCM-simulated values is considered acceptable as these are related to large-scale changes such as global increase of greenhouse gases (GHG) concentrations.

The projections obtained for temperature in the present study strengthen the inferences drawn in Anandhi *et al.* (2008) for precipitation in the study region. In the referred work, the projected increase in precipitation was high for A2 scenario, whereas it was least for B1 scenario. This could be because the rate of evaporation is proportional to the increase in the earth's surface temperature, and the evaporated water would eventually precipitate.

8. Summary and conclusions

The SVM downscaling model is developed for obtaining projections of monthly mean maximum and minimum temperatures (predictands) at river-basin scale. The effectiveness of the model is demonstrated through application to the catchment of Malaprabha reservoir in India. The predictands are downscaled from simulations of CGCM3 for four IPCC scenarios, namely SRES A1B, A2, B1 and COMMIT. The results of validation indicate that the SVM model is a feasible choice for downscaling the predictands.

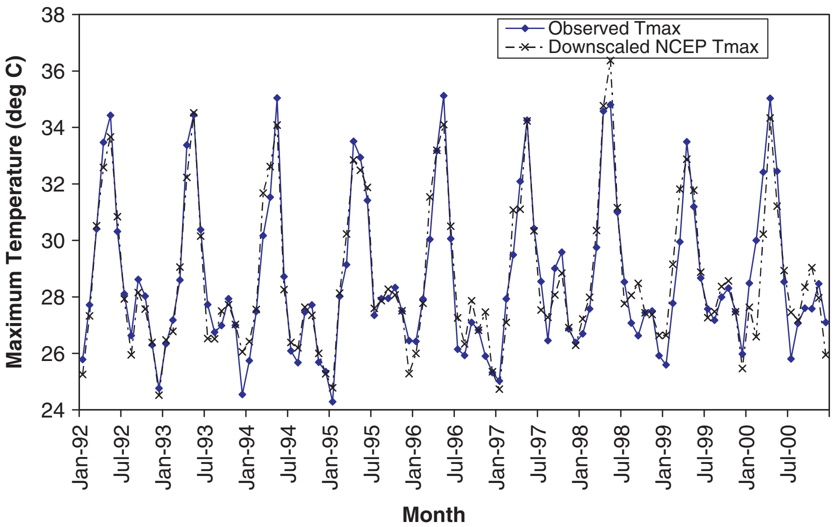

Figure 11. Typical results for comparison of the monthly observed T_{max} with T_{max} simulated using SVM downscaling model 11 for NCEP data. In the figure, calibration period is from 1992 to 1997, and the rest is validation period. This figure is available in colour online at www.interscience.wiley.com/ijoc

DOI: 10.1002/joc

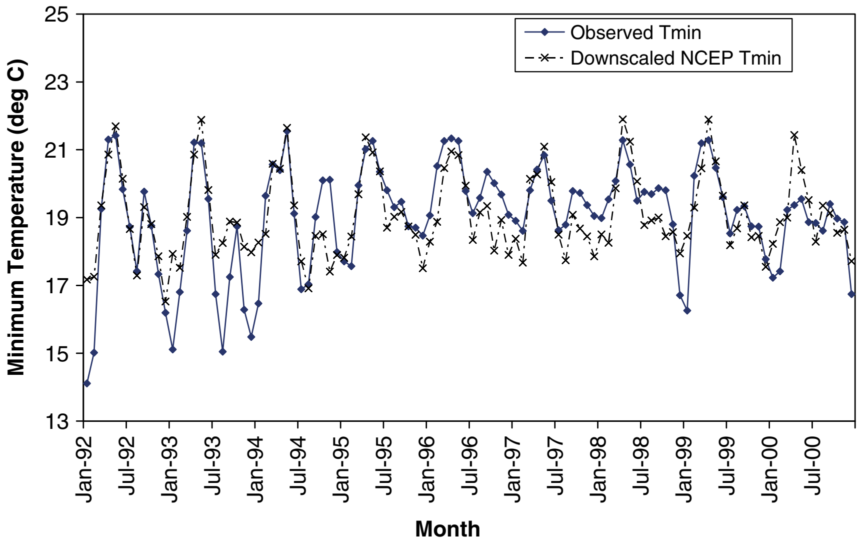

Figure 12. Typical results for comparison of the monthly observed T_{min} with T_{min} simulated using SVM downscaling model 13 for NCEP data. In the figure, calibration period is from 1992 to 1997, and the rest is validation period. This figure is available in colour online at www.interscience.wiley.com/ijoc

The selected predictor variables are classified into three groups namely A, B and C. Large-scale atmospheric variables such as air temperature, zonal and meridional wind velocities at 925 mb which are often used for downscaling temperature are considered as predictors in Group A. Surface flux variables such as LH, SH, shortwave radiation and longwave radiation fluxes are tried as plausible predictors in Group B. Group C comprises of all the variables in both Groups A and B.

Scatter plots and cross-correlations used for studying the reliability of the simulation of the predictor variables by the GCM, and to study the predictor–predictand relationships indicate that the Group A predictors are better simulated by the GCM than Group B predictors.

Eighteen SVM models are developed, one for each combination of predictor group, predictand, calibration period and spatial domain of the climate variables. The performance of the models is evaluated using the statistical measures SSE, MSE, RMSE, NMSE, E_f, MAE, MCE and CC.

The performance of the downscaling model did not change significantly when the calibration period was increased from 6 to 16 years indicating that SVM can offer effective performance even with shorter records. Further, the SVM models based on predictor variables pertaining to land-based stratification showed better performance than those based on predictor variables pertaining to both land and ocean. Furthermore, the SVM models developed using Group C predictors performed better than those based on predictors in the other groups indicating that surface flux variables are also necessary for downscaling the predictands.

The results of downscaling show that T_{max} and T_{min} are projected to increase in future for A1B, A2 and B1 scenarios, whereas no trend is discerned with the COMMIT using predictors in Groups A and C. The projected increase in predictands is high for A2 scenario, whereas it is least for B1 scenario. These results are in agreement with those obtained for precipitation in Anandhi *et al.* (2008) for the same study area.

In contrast, projections obtained for the predictands using the predictors in Group B did not show any trend for the four scenarios. This projected increase in trend of predictands for Groups A and C is attributed to the increasing trend in air temperature at 925 mb which is one of the predictors in these groups. A similar trend was observed in monthly surface temperature simulated by GCM at grid points considered on land. The results suggest that it is necessary to consider predictor variables having trends similar to that of the predictand to be downscaled.

Overall, the results of the SVM downscaling models indicate that between the two predictands, T_{max} is better simulated than T_{min}. Although the present analysis is confined to only one river basin, the methodology developed for downscaling temperature using LS-SVM can be extended to other river basins, as well.

Acknowledgements

This work is partially supported by INCOH, Ministry of Water Resources, Govt. of India, through Project No. 23/52/2006-R&D. The support from the Drought Monitoring Cell, Government of Karnataka, is also acknowledged. 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 Modelling and Analysis
CGCM3	Third-generation Canadian Global Climate Model
GCM	Global Climate Model
IPCC	Intergovernmental panel on climate change
LH	Latent heat flux
LS-SVM	Least square-support vector machine
LWR	Longwave radiation flux
MAE	Mean absolute error
MCE	Mean cumulative error
MSE	Mean square error
MMP	Mean monthly precipitation
NMSE	Normalized mean square error
PCA	Principal component analysis
PC	Principal component
RBF	Radial basis function
RMSE	Root mean square error
SH	Sensible heat flux
SWR	Shortwave radiation flux
SRES	Special report of emission scenarios
SVM	Support vector machine
Ta 925	Air temperature at 925 mb
Ua 925	Zonal wind at 925 mb
Va 925	Meridional wind at 925 mb

Appendix: Abbreviations used in Tables I, II and VII

Predictor Names

afs	Surface airflow strength
di	Divergence
E_f	Nash-Sutcliffe error estimate
F	Geostrophic airflow
geos	Meridional component of geostrophic flow
geow	Zonal component of geostrophic flow
hus	Specific humidity
LH	Latent heat
LWR	Longwave radiation
MAE	Mean absolute error
MCE	Mean cumulative error
MSE	Mean square error
mslp	Mean sea level pressure
NMSE	Normalized mean square error
pr	Precipitation
prw	Precipitable water content
ps	Pressure
RMSE	Root mean square error
rh	Relative humidity
SH	Sensible heat
SSE	Sum of squares of errors
SWR	Shortwave radiation
T_{mean}	Mean temperature
ta	Air temperature
ua	Zonal wind
va	Meridional wind
wd	Wind direction
Z	Vorticity
zg	Geopotential height
zgt	Geopotential height thickness

Note: M preceding the predictor variable name indicates that the mean was used.

Measurement height of predictors

_0	Pressure height at 1000 mb
_2	Pressure heights at 200 mb
_2m	2 m from surface
_5	Pressure height at 500 mb
_7	Pressure height at 700 mb
_8	Pressure height at 850 mb
_9	Pressure height at 925 mb
_ns	Near-surface
_s	Surface

Techniques

AM	Analogue method
CCA	Canonical correlation analysis
EOF	Empirical orthogonal function
LS	Local scaling
MLR	Multi-linear regression
PCA	Principal component analysis
SDSM	Statistical downscaling model
SSA	Singular spectrum analysis
TNN	Temporal neural network

Data source

BMRC	Bureau of Meteorology Research Centre
CSIRO	Commonwealth Scientific and Industrial Research Organization, Australia
DOE	Department of Energy, USA
ECMWF	European Centre for Medium-Range Weather Forecasts
LMD	Laboratoire de Météorologie Dynamique du.
NCAR	National Center for Atmospheric Research, USA

Climate models:

CLIGEN	Climate Generator
CGCM	Canadian Coupled Global Climate Model
CSIRO-Mk2	CSIRO climate system model (make/ version 2)
ECHAM4	fourth generation GCM based on the weather forecast model of the **EC**MWF, modified and extended in **Ham**burg, Germany
HadCM3	Third-generation coupled GCM developed by the Hadley Centre of United Kingdom Meteorological Office, UK.
PCM	Parallel Climate Model developed by DOE and NCAR

References

Anandhi A, Srinivas VV, Nanjundiah RS, Kumar DN. 2008. Downscaling precipitation to River Basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology* **28**: 401–420, DOI: 10.1002/joc.1529.

Benestad RE. 2001. A comparison between two empirical downscaling strategies. *International Journal of Climatology* **21**: 1645–1668.

Bergant K, Kajfez-Bogataj L, Trdan S. 2006. Uncertainties in modelling of climate change impact in future: An example of onion thrips (Thrips Tabaci indeman) in Slovenia. *Ecological Modelling* **194**: 244–255.

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.

Chen D, Chen Y. 2003. Association between winter temperature in China and upper air circulation over East Asia revealed by canonical correlation analysis. *Global and Planetary Change* **37**: 315–325.

Collinson NH, Sparks TH. 2004. Nature's changing seasons. 2003 results from the UK Phenology Network. *British Wildlife* **15**: 245–250.

Croxton PJ, Huber K, Collinson N, Sparks TH. 2006. How well do the central England temperature and the England and Wales precipitation series represent the climate of the UK? *International Journal of Climatology* **26**: 2287–2292, DOI: 10.1002/joc.1378.

Dahm CN, Cleverly JR, Coonrod JA, Thibault JR, McDonennell DE, Gilroy DJ. 2002. Evapotranspiration at the land/water interface in a semi-arid drainage basin. *Freshwater Biology* **47**(4): 831.

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, 6–10 December.

Droogers P, Aerts J. 2005. Adaptation strategies to climate change and climate variability: A comparative study between seven contrasting river basins. *Physics and Chemistry of the Earth* **30**: 339–346.

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.

Gupta RS. 1989. *Hydrology and Hydraulic Systems*. Waveland Press: Illinois, USA.

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

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

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, New York.

Jessie CR, Antonio RM, Stahis SP. 1996. *Climate Variability, Climate Change and Social Vulnerability in the Semi-arid Tropics*. Cambridge University Press: Cambridge.

Johnson MS, Coon WF, Mehta VK, Steenhuis TS, Brooks ES, Boll J. 2003. Application of two hydrologic models with different runoff mechanisms to a hillslope dominated watershed in the northeastern US: a comparison of HSPF and SMR. *Journal of Hydrology* **284**: 57–76, DOI:10.1016/j.jhydrol.2003.07.005.

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.

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

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

Marshall SJ, Sharp MJ, Burgess DO, Anslow FS. 2007. Near-surface-temperature lapse rates on the Prince of Wales Icefield, Ellesmere island, Canada: implications for regional downscaling of temperature. *International Journal of Climatology* **27**(3): 385–398, DOI: 10.1002/joc.1396.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models. Part I – a discussion of principles. *Journal of Hydrology* **10**: 282–290.

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.

Pitman AJ. 2003. The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology* **23**: 479–510, DOI: 10.1002/joc.893.

Roderick ML, Farquhar GD. 2005. Changes in New Zealand pan evaporation since the 1970s. *International Journal of Climatology* **25**: 2031–2039, DOI: 10.1002/joc.1262.

Rosenberg NJ, McKenney MS, Martin P. 1989. Evapotranspiration in a greenhouse-warmed world: a review and a simulation. *Agricultural and Forest Meteorology* **47**: 303–320.

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, DOI: 10.1002/joc.1412.

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.

Tatli H, Dalfes HN, Mentes S. 2005. Surface air temperature variability over Turkey and its connection to large-scale upper air circulation via multivariate techniques. *International Journal of Climatology* **25**: 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.

Tripathi S, Srinivas VV, Nanjundiah RS. 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology* **330**(3–4): 621–640, DOI:10.1016/j.jhydrol.2006.04.030.

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 Modelling & 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).

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

Zhang B, Govindaraju RS. 2000. Prediction of watershed runoff using bayesian concepts and modular neural network. *Water Resources Research* **36**(3): 753–762.