ORIGINAL PAPER

PLS regression-based pan evaporation and minimum–maximum temperature projections for an arid lake basin in India

Manish Kumar Goyal · C. S. P. Ojha

Abstract Climate change information required for impact studies is of a much finer scale than that provided by Global circulation models (GCMs). This paper presents an application of partial least squares (PLS) regression for downscaling GCMs output. Statistical downscaling models were developed using PLS regression for simultaneous downscaling of mean monthly maximum and minimum temperatures (T_{max} and T_{min}) as well as pan evaporation to lake-basin scale in an arid region in India. The data used for evaluation were extracted from the NCEP/NCAR reanalysis dataset for the period 1948–2000 and the simulations from the third-generation Canadian Coupled Global Climate Model (CGCM3) for emission scenarios A1B, A2, B1, and COMMIT for the period 2001–2100. A simple multiplicative shift was used for correcting predictand values. The results demonstrated that the downscaling method was able to capture the relationship between the premises and the response. The analysis of downscaling models reveals that (1) the correlation coefficient for downscaled versus observed mean maximum temperature, mean minimum temperature, and pan evaporation was 0.94, 0.96, and 0.89, respectively; (2) an increasing trend is observed for T_{max} and T_{min} for A1B, A2, and B1 scenarios, whereas no trend is discerned with the COMMIT scenario; and (3) there was no

trend observed in pan evaporation. In COMMIT scenario, atmospheric CO_2 concentrations are held at year 2000 levels. Furthermore, a comparison with neural network technique shows the efficiency of PLS regression method.

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 (Anandhi et al. 2009). General circulation models (GCMs), representing physical processes in the atmosphere, ocean, cryosphere, and land surface, are the most advanced tools currently available to simulate time series of climate variables for the world, accounting for the effects of the concentration of greenhouse gases in the atmosphere and to obtain information about an altered global environment and climate system (Prudhomme et al. 2003). However, in most climate change impact studies, such as hydrological impacts of climate change, impact models are usually required to simulate sub-grid scale phenomenon and therefore, require input data (such as temperature) at similar sub-grid scale. The methods used to convert GCM outputs into local meteorological variables required for reliable hydrological modeling are usually referred to as "downscaling" techniques. Hydrologic variables, such as temperature, evaporation, etc., are significant parameters for climate change impact studies. A proper assessment of probable future temperature and their variability are to be made for various hydroclimatology scenarios.

More recently, downscaling has found wide application in hydroclimatology for scenario construction and simulation/prediction of (1) low-frequency rainfall events (Wilby 1998),

M. K. Goyal (✉) · C. S. P. Ojha
Department of Civil Engineering, Indian Institute of Technology,
Roorkee, India
e-mail: vipmkgoyal@rediffmail.com

M. K. Goyal
e-mail: mkgoyal@uwaterloo.ca

M. K. Goyal
Department of Civil and Environmental Engineering,
University of Waterloo,
Waterloo, Canada

(2) mean temperature (Benestad 2001), (3) potential evaporation rates (Weisse and Oestreicher 2001), (4) daily T_{max} and T_{min} (Wilby et al. 2002), (5) streamflows (Cannon and Whitfield 2002), (6) runoff (Arnell et al. 2003), (7) soil erosion and crop yield (Zhang et al. 2004), (8) mean, minimum, and maximum air temperature (Kettle and Thompson 2004), (9) precipitation (Tripathi et al. 2006), (10) daily T_{max} and T_{min} (Schoof and Pryor 2001), (11) streamflow (Ghosh and Mujumdar 2008), (12) T_{max} and T_{min} (Anandhi et al. 2009), and (13) precipitation (Vimont et al. 2009).

Downscaling models make use of a strong observed empirical relationship between one or several large-scale predictors and a variable of interest at regional scale, the predictand. The relationships between these scales can be determined by a number of methods including regression (Kilsby et al. 1998), partial least squares (PLS) regression (Bergant and Kajfež-Bogataj 2005), canonical correlation analysis (Heyen et al. 1996; Xoplaki et al. 2000), K-nearest neighbor (Gangopadhyay et al. 2005), and artificial neural networks (Hewitson and Crane 1994; Gardner and Dorling 1998; Cannon and Lord 2000; Schoof and Pryor 2001; Goyal and Ojha 2010a; Ojha et al. 2010). In the literature, authors have not found application of PLS regression technique for simultaneous downscaling of maximum and minimum temperatures as well as evaporation specifically for Indian region.

In this paper, we present a downscaling methodology based on PLS projection to latent structures regression technique to study climate change impact over Pichola lake basin in an arid region. The objectives of this study include: (1) predictor selection, based on variable Importance in the Projection (VIP) score; (2) downscaling of mean monthly maximum temperature (T_{max}), minimum temperature (T_{min}), and pan evaporation using PLS regression approach; (3) an application of simple multiplicative shift to correct the bias of mean monthly GCM-simulated variables, and (4) comparing results with neural network approach from simulations of Canadian Coupled Global Climate Model (CGCM3) for latest Intergovernmental Panel on Climate Change (IPCC) scenarios. The scenarios which are studied in this paper are relevant to IPCCs fourth assessment report which was released in 2007.

The remainder of this paper is structured as follows: section 2 provides a description of the study region and reasons for its selection. Section 3 provides details of various data used in the study. Section 4 describes briefly the PLS regression and the reasons for selection of the predictor variables for downscaling. Section 5 explains the proposed methodology for development of the PLS regression downscaling models for downscaling T_{max}, T_{min}, and pan evaporation to the lake basin and introduction of multiplicative shift for bias correction. Section 6 presents the results and discussion. Finally, section 7 provides the conclusions drawn from the study.

2 Study region

The area of the this study is the Pichola lake catchment in Rajasthan state in India that is situated from 72.5° to 77.5° E and 22.5° to 27.5° N. The Pichola lake basin, located in Udaipur district, Rajasthan is one of the major sources for water supply for this arid region. During the past several decades, the streamflow regime in the catchment has changed considerably, which resulted in water scarcity, low agriculture yield and degradation of the ecosystem in the study area. Regions with arid and semi-arid climates could be sensitive even to insignificant changes in climatic characteristics (Linz et al. 1990). Temperature affects the evapotranspiration (Jessie et al. 1996), evaporation, and desertification processes and is also considered as an indicator of environmental degradation and climate change. Understanding the relationships among the hydrologic regime, climate factors, and anthropogenic effects is important for the sustainable management of water resources in the entire catchment; hence, this study area was chosen because of the aforementioned reasons.

The mean monthly T_{max} in the catchment varies from 19°C to 39.5°C and mean annual T_{max} is 30.6°C. The mean monthly T_{min} ranges from 3.4°C to 29.8°C based on decadal (1990–2000) observed value. The observed mean monthly T_{max} and T_{min} as well as pan evaporation have been shown in Fig. 1a, b for various months of year 2000, respectively. The location map of the study region is shown in Fig. 2.
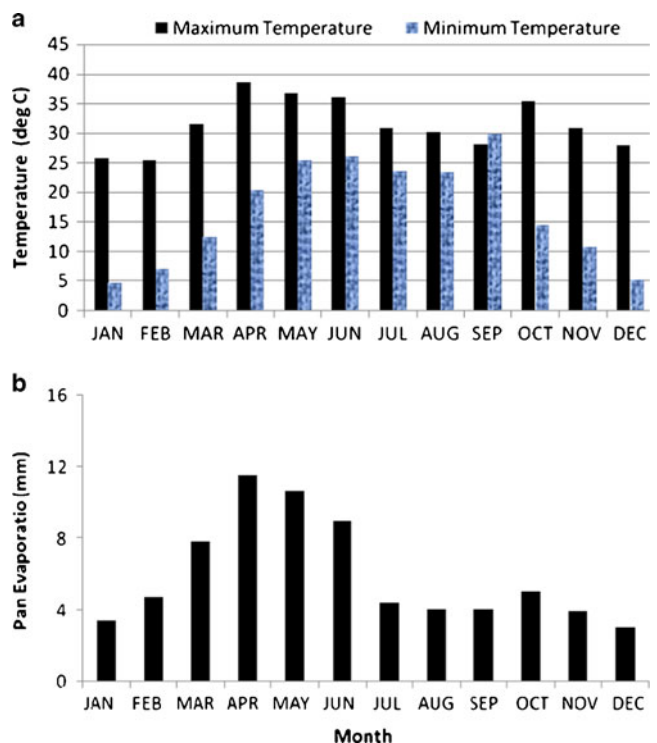

Fig. 1 **a** Maximum and minimum temperature in the study region. **b** Observed pan evaporation in the study region
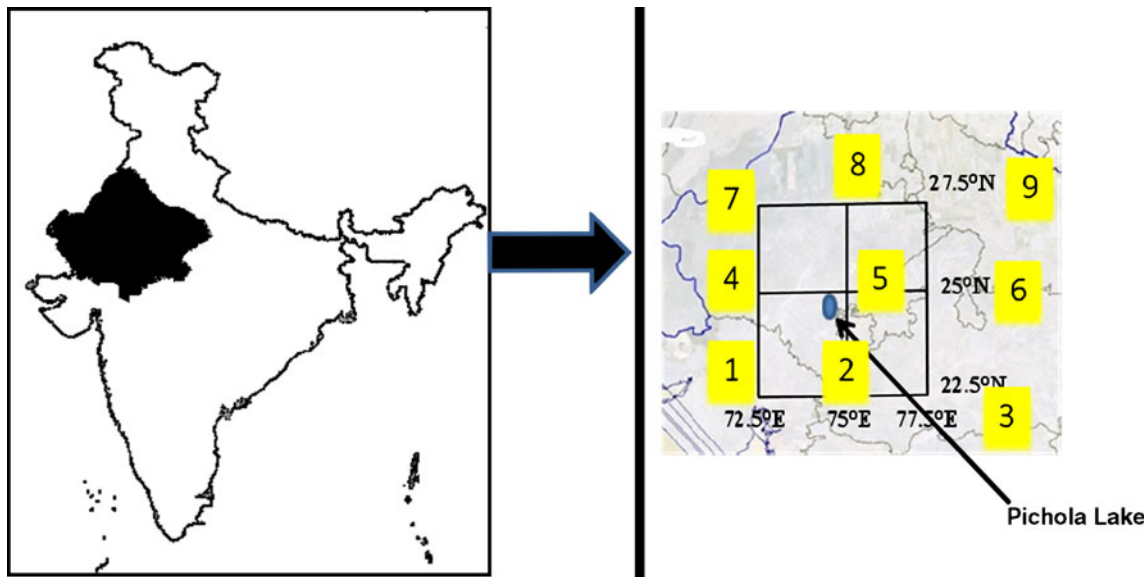
Fig. 2 Location map of the study region in Rajasthan State of India with NCEP grid

3 Data extraction

The monthly mean atmospheric variables were derived from the National Center for Environmental Prediction (NCEP/NCAR; hereafter called NCEP) reanalysis data set (Kalnay et al. 1996) for the period of January 1948 to December 2000. The data have a horizontal resolution of 2.5° latitude×2.5° longitude and 17 constant pressure levels in the vertical. The atmospheric variables are extracted for nine grid points whose latitude ranges from 22.5° to 27.5° N, and longitude ranges from 72.5° to 77.5° E at a spatial resolution of 2.5°. The meteorological data, i.e., T_{max} and T_{min} as well as pan evaporation are used at monthly time scale from records available for Pichola Lake which is located in Udaipur at 24°34′ N latitude and 73°40′ E longitude. The data is available for the period January 1990 to December 2000 (Khobragade 2009). The Canadian Center for Climate Modeling and Analysis (CCCma) (http://www.cccma.bc.ec.gc.ca) provides GCM data for a number of surface and atmospheric variables for the CGCM3 T47 version which has a horizontal resolution of roughly 3.75° latitude×3.75° longitude and a vertical resolution of 31 levels. CGCM3 is the third version of the CCCMA Coupled Global Climate Model which makes use of a significantly updated atmospheric component AGCM3 and uses the same ocean component as in CGCM2. The data comprise of present-day (20C3M) and future simulations forced by four emission scenarios, namely A1B, A2, B1, and COMMIT.

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. The GCM data is re-gridded to a common 2.5° using inverse square interpolation technique (Willmott et al. 1985).The utility of this interpolation algorithm was examined in previous downscaling studies (Shannon and Hewitson 1996; Crane and Hewitson 1998; Tripathi et al. 2006; Ghosh and Mujumdar 2008; Goyal and Ojha 2010b, c). The development of downscaling models for each of the predictand variables T_{max} and T_{min} as well as pan evaporation begins with the selection of potential predictors, followed by the application of PLS regression on downscaling model. The developed model is then used to obtain projections of T_{max} and T_{min} as well as pan evaporation from simulations of CGCM3.

4 PLS regression and selection of predictors

4.1 PLS regression

PLS regression is used to describe the relationship between multiple response variables and predictors through the latent variables. PLS regression can analyze data with strongly collinear, noisy, and numerous X-variables, and also simultaneously model several response variables, Y. In general, the PLS approach is particularly useful when one or a set of dependent variables (or time series) need to be predicted by a (very) large set of predictor variables (or time series) that are strongly cross-correlated (Abdi 2003). This is often the case in empirical downscaling of climate variables (Bergant and Kajfež-Bogataj 2005). For details of PLS regression, readers are referred to Manne (1987, Lindgren et al. (1993), Rannar et al. (1994), and Wold et al. 2001).

Fig. 3 a VIP of the predictand variable (T_{max}) of the two-component PLSR model. **b** VIP of the predictand variable (T_{min}) of the two-component PLSR model
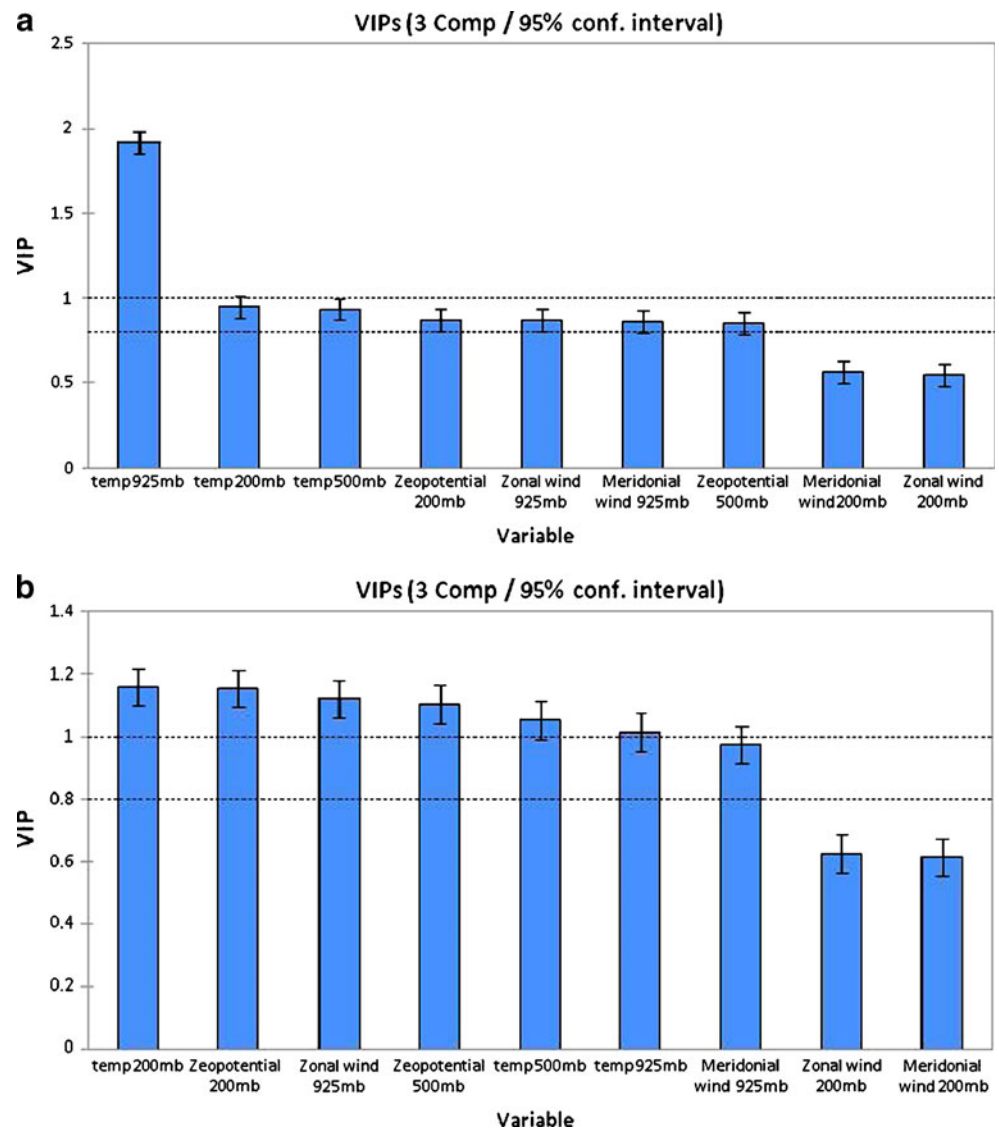
4.2 Different error norms

The different statistical parameters of each model are calculated during calibration to get the best statistical agreement between observed and simulated meteorological variables. For this purpose, various statistical performance measures, such as coefficient of correlation (CC), root mean square error (RMSE) and Nash–Sutcliffe Efficiency Index (Nash and Sutcliffe 1970) were used to measure the performance of various models.

4.3 Selections of predictors

The selection of appropriate predictors is one of the most important steps in a downscaling exercise for downscaling predictands. The predictors are chosen by the following criteria: (1) predictors are skillfully predicted by GCMs; (2) they should represent important physical processes in the

context of the enhanced greenhouse effect; (3) they should not be strongly correlated to each other (Hewitson and Crane 1996; Hellström et al. 2001; Cavazos and Hewitson 2005; Goyal and Ojha 2010d, e). Various authors, such as, Hertig and Jacobeit (2008), Anandhi et al. (2009) have used large-scale atmospheric variables, viz., air temperature, geo-potential height, zonal (u) and meridional (v) wind velocities, as the predictors for downscaling GCM output to temperature over an area. For this study, we have used a total of nine possible predictor variables, namely, air temperature (at 925,500 and 200 hPa pressure levels), geo-potential height (at 200 and 500 hPa pressure levels), zonal (u), and meridional (v) wind velocities (at 925 and 200 hPa pressure levels), as the predictors for downscaling GCM output to mean monthly temperature and pan evaporation over a catchment.

The VIP scores obtained by the PLS regression, has been paid an increasing attention as an importance measure of each explanatory variable or predictor (Chong and Jun

Table 1 Different downscaling model variants used in the study for obtaining projections of predictands at monthly time scale

Predictand	Period of downscaling	Length of the record	Approach	Model
T_{max}	1990–2100	1990–2000	PLS regression	PLSM1
T_{min}	1990–2100	1990–2000	PLS Regression	PLSM2
Pan evaporation	1990–2100	1990–2000	PLS regression	PLSM3
T_{max}	1990–2100	1990–2000	Neural network	NNM1
T_{min}	1990–2100	1990–2000	Neural network	NNM2
Pan evaporation	1990–2100	1990–2000	Neural network	NNM3

2005). The variable selection procedure under PLS is proposed with an application to downscaling technique for identifying influencing variables on understanding the impact of climate change. The VIP scores which are obtained by PLS regression, can be used to select most influential variables or predictors, X (Chong and Jun 2005). The VIP score can be estimated for jth X-variable by

$$\mathrm{VIP}_j = \sqrt{\frac{p}{\sum\limits_{i=1}^{k} R_d(Y, t_i)} \sum\limits_{i=1}^{k} R_d(Y, t_i) w_{ij}^2} \quad (1)$$

where R_d is defined as the mean of the squares of the correlation coefficients (R) between the variables and the component and p is number of predictors.

$$R_d(X, c) = \frac{1}{p} \sum\limits_{i=1}^{k} R^2(x_j, c) \quad (2)$$

Usually the predictor variable whose VIP score is greater than 0.8 and above is considered as an important variable (Wold 1995; Eriksson et al. 2001)

It can be seen form Fig. 3a, b that seven predictor variables, namely, air temperature (925, 500, and 200 hPa); zonal wind (925 hPa); meridional wind (925 hPa); geopotential height (500 and 200 hPa) have their VIP score greater than 0.8. Hence, these variables are used in the prediction model to obtain the projection of predictands. It is noted that different predictors control different local variables and mean temperature is most sensitive to surface and near surface atmospheric factors (Chu et al. 2010).

5 Downscaling of GCM models

PLS regression is used to downscale mean T_{max} and T_{min} as well as pan evaporation in this study. The data of potential

Table 2 Regression coefficients for models PLSM1, PLSM2, and PLSM3

Model	Intercept	A11	A12	A13	A14	A15	A16	A17	A18	A19	A21	A22	A23
PLSM1	30.34	0.57	0.51	0.45	0.46	0.45	0.42	0.34	0.38	0.29	−0.04	−0.06	−0.11
PLSM2	16.00	0.34	0.28	0.24	0.34	0.29	0.27	0.31	0.28	0.26	0.18	0.16	0.11
PLSM3	15.78	0.12	0.87	0.29	0.78	0.19	0.67	0.45	0.78	0.47	0.87	0.26	0.92
Model	A24	A25	A26	A27	A28	A29	A31	A32	A33	A34	A35	A36	A37
PLSM1	0.02	−0.03	−0.06	0.03	−0.08	−0.10	−0.09	−0.16	−0.22	0.00	−0.09	−0.20	−0.04
PLSM2	0.13	0.08	0.02	−0.01	−0.03	0.01	0.02	−0.02	−0.02	0.09	0.07	0.09	0.15
PLSM3	0.56	0.19	0.92	0.34	0.45	0.67	0.28	0.71	0.68	0.38	0.61	0.48	0.24
Model	A38	A39	A41	A42	A43	A44	A45	A46	A47	A48	A49	A51	A52
PLSM1	−0.09	−0.21	−0.07	−0.07	−0.08	−0.09	−0.09	−0.09	−0.08	−0.09	−0.10	0.15	0.17
PLSM2	0.16	0.16	0.06	0.05	0.05	0.05	0.04	0.04	0.05	0.04	0.04	0.23	0.24
PLSM3	−0.67	0.72	0.28	0.19	0.78	−0.45	−0.34	0.12	0.45	0.73	0.82	0.71	0.36
Model	A53	A54	A55	A56	A57	A58	A59	A61	A62	A63	A64	A65	A66
PLSM1	0.18	0.16	0.17	0.18	0.16	0.17	0.18	0.06	0.05	0.05	0.05	0.04	0.04
PLSM2	0.25	0.23	0.24	0.25	0.23	0.25	0.25	0.12	0.12	0.12	0.12	0.12	0.12
PLSM3	−0.48	0.18	−0.26	0.45	0.28	0.48	0.82	0.61	0.53	0.72	0.37	0.49	0.39
Model	A67	A68	A69	A71	A72	A73	A74	A75	A76	A77	A78	A79	
PLSM1	0.05	0.04	0.03	0.05	0.03	0.02	0.04	0.03	0.02	0.02	0.01	0.01	
PLSM2	0.12	0.12	0.12	0.09	0.08	0.08	0.09	0.08	0.07	0.08	0.08	0.07	
PLSM3	0.46	0.78	−0.38	0.78	0.28	0.45	0.61	0.81	0.37	−0.24	0.18	0.71	

Table 3 Various quality measures of PLS regression model

Index	Max temp (PLSM1)			Min temp (PLSM2)			Pan evaporation (PLSM3)		
	Comp1	Comp2	Comp3	Comp1	Comp2	Comp3	Comp1	Comp2	Comp3
Q^2 cum	0.495	0.888	0.909	0.947	0.951	0.921	0.612	0.772	0.912
R^2X cum	0.727	0.883	0.756	0.883	0.928	0.931	0.823	0.873	0.892
R^2Y cum	0.506	0.893	0.910	0.950	0.956	0.929	0.543	0.792	0.941

predictors is first standardized. Standardization is widely used prior to 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). Standardization is done for a baseline period of 1948 to 2000 because it is of sufficient duration to establish a reliable climatology, yet not too long, nor too contemporary to include a strong global change signal (Wilby et al. 2004; Ghosh and Mujumdar 2008).

To develop downscaling models, the feature vectors (i.e., predictors) which are prepared from NCEP record, are partitioned into a training set and a validation set. Feature vectors in the training set are used for calibrating the model, and those in the validation set are used for validation. The 11-year mean monthly observed maximum and minimum temperatures as well as pan evaporation data series were broken up into a calibration period and a validation period. Table 1 summarizes the certain details of models. The various error criteria are used as an index to assess the performance of the model. Based on the latest IPCC scenario, models for mean monthly T_{max} and T_{min} as well as *pan evaporation* were evaluated based on the accuracy of the predictions for validation data set. The criteria such as Q^2cum index, R^2Xcum and R^2Ycum index of PLS regression models were chosen in this study (Wold 1995; Eriksson et al. 2001; Wold et al. 2001).

Regression coefficients (Aij) for each predictor have been shown in Table 2 where *i* ranges from 1 to 7 indicating Ta 925, Ua 925, Va 925, Ta 500, Ta 200, Zg 200, and Zg

500, respectively, while *j* ranges from 1 to 9 representing location of points in grid, as shown in Fig. 2.

5.1 Correcting bias by a multiplicative shift

Many GCMs either overestimate or underestimate maximum and minimum temperature. The correction scheme brings the distributions close to the observed pattern. A simple multiplicative shift is used to correct the bias of the mean monthly GCM-simulated variable as follows:

$$X_i^{'} = X_i \frac{\overline{X}_{obs}}{\overline{X}_{GCM}} \tag{3}$$

where $X_i^{'}$, X_i refers to raw and corrected GCM-simulated variable, and $\overline{X}_{GCM}$ and $\overline{X}_{obs}$ are long term mean monthly variable from the GCM and the observations for given month (Amor and Hansen 2006).

6 Results and discussions

Seven predictor variables, namely, air temperature (925, 500, and 200 hPa); zonal wind (925 hPa); meridional wind (925 hPa); geo-potential height (500 and 200 hPa) at nine NCEP grid points with a dimensionality of 63, are used as the standardized data of potential predictors. These feature vectors are provided as input to the PLS regression downscaling model. Model quality indexes Q^2cum index, R^2Xcum and R^2Ycum index have been shown in Table 3. It is clear

Table 4 Various performance statistics of models using PLS regression

Model	CR		RMSE		N-S Index	
	Training	Validation	Training	Validation	Training	Validation
PLSM1	0.96	0.94	1.23	1.63	0.92	0.87
PLSM2	0.98	0.96	1.55	2.26	0.93	0.87
PLSM3	0.95	0.89	1.78	2.78	0.89	0.85
PLSM1 (corrected)	0.97	0.96	1.13	1.43	0.94	0.89
PLSM2 (corrected)	0.99	0.98	1.15	2.16	0.94	0.91
PLSM3 (corrected)	0.97	0.93	1.74	2.67	0.90	0.86

RMSE is in degree Celsius for PLSM1 and PLSM2 and in millimeters for PLSM3

Table 5 Mann–Kendall statistics for T_{max} based on 2001–2100 for June

T_{max}	SRESA1B	SRESA2	SRESB1	COMMIT	Historical
S	340	267	667.0	378	−3
Test statistics	0.77	0..45	4.51	0.78	−0.15

that all three indices are highest for the first three components of the predictands. For predictand T_{max}, Q^2cum index, R^2Xcum and R^2Ycum index are 0.921, 0.931, and 0.929; respectively. For T_{min}, Q^2cum index, R^2Xcum and R^2Ycum index are 0.951, 0.928, and 0.956, respectively. Similarly, for predictand pan evaporation, Q^2cum index, R^2Xcum and R^2Ycum index are obtained as 0.912, 0.892, and 0.941, respectively. Hence, model quality can be considered as good. PLS regression is performed on this dataset. Results of the different PLS regression models (viz. *PLSM1*, *PLSM2*, and *PLSM3*), as discussed in Table 1, are tabulated in Table 4. Neural network (NN) models have been developed for each predictand. A comprehensive search of neural network architecture is done by varying the number of nodes in hidden layer. The network is trained using back-propagation algorithm. Results of the different models of neural network technique (*NNM1, NNM2*, and *NNM3* for T_{max}, T_{min}, and *pan evaporation*, respectively) were imported from previous study of Goyal and Ojha (2009). The calibration and validation results are described next.

6.1 Calibration/training results

It can be observed from Table 4 that for predictand T_{max}, CC, RMSE and N-S Index were 0.96 1.23, and 0.92, respectively, using PLS regression model *PLSM1* while CC, RMSE and N-S Index were 0.99, 0.96, and 0.98, respectively using neural network model NNM1. For predictand T_{min}, values of CC, RMSE, N-S Index and MAE were 0.98, 1.55, and 0.93, respectively, while for model NNM2, values of CC, RMSE, N-S Index and MAE were 0.98, 0.91, and 0.96, respectively. The coefficient of correlation and N-S Index for the PLSM3 model were 0.95 and 0.89, respectively, whereas the values of the coefficient of correlation and N-S Index for the model

NNM3 were 0.94 and 0.90, respectively, from predictand pan evaporation.

6.2 Validation/testing results

For predictand T_{max}, values of CC, RMSE, N-S Index were 0.94, 1.63, and 0.92, respectively for PLSM1 model while values of CC, RMSE, N-S Index were 0.96, 2.31, and 0.91, respectively, for NNM1 model. For predictand T_{min}, the values of CC and RMSE were 0.96 and 2.26, respectively, for PLSM2 model while the values of CC and RMSE were 0.94 and 1.62, respectively, for NNM2 model. The value of N-S Index was same, i.e., 0.87 for both models. The coefficient of correlation and N-S Index for the PLSM3 model were 0.89 and 0.85, respectively whereas the values of the coefficient of correlation and N-S Index for the model NNM3 were 0.90 and 0.84, respectively, for predictand pan evaporation.

Thus multiplicative shift is used to correct the bias of GCM of models PLSM1, PLSM2 and PLSM3 corresponding to T_{max}, T_{min}, and pan evaporation, respectively. All the corrected models performed better than uncorrected in terms of various performance measures, as shown in Table 5. It can be inferred that the performance of PLS regression models bias corrected (viz. PLSM1 (corrected), PLSM2 (corrected), and PLSM3 (corrected)) for predictands (T_{max}, T_{min} as well as pan evaporation) performed well and are competitive in downscaling predictands values with neural network models and the comparsion shows that PLS regression is a reasonable choice.

A comparison of mean monthly observed T_{max} and T_{min} as well as pan evaporation with T_{max} and T_{min} as well as pan evaporation simulated using PLS regression models PLSM1 (corrected), PLSM2 (corrected), and PLSM3 (corrected) have been shown from Figs. 4, 5, and 6, respectively, for calibration and validation period.

Fig. 4 Typical results for comparison of the monthly observed T_{max} with T_{max} simulated using PLR regression downscaling model PLSM1 for NCEP data
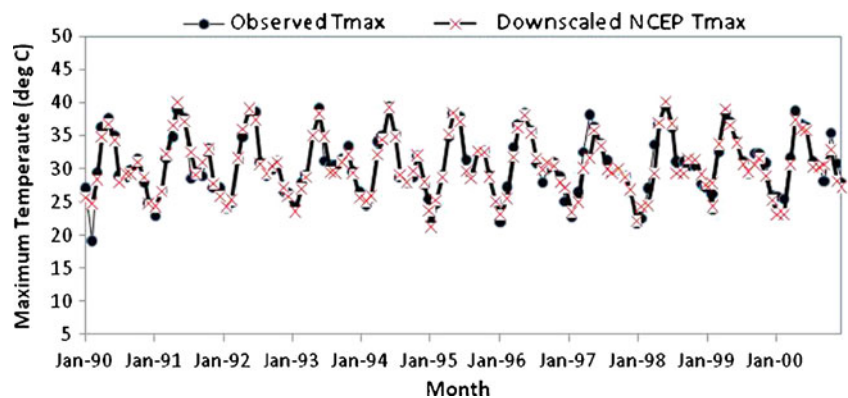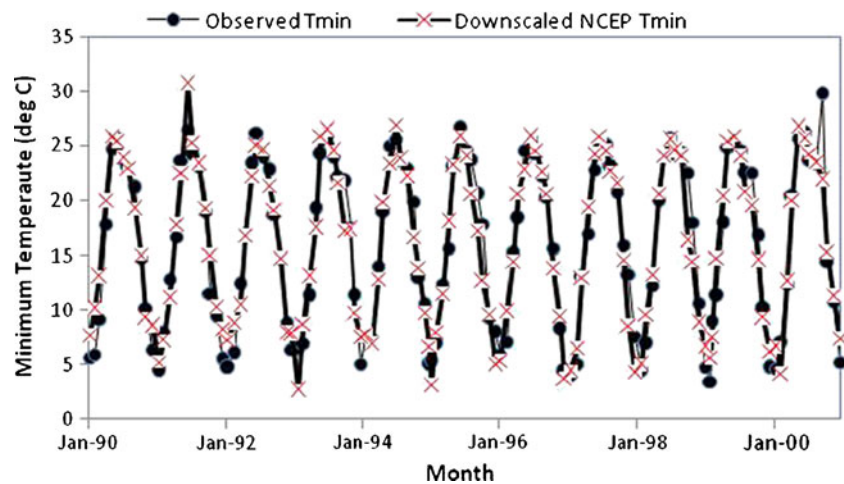
Fig. 5 Typical results for comparison of the monthly observed T_{min} with T_{min} simulated using PLS regression downscaling model PLSM2 for NCEP data

Once the downscaling models have been calibrated and validated, the next step is to use these models to downscale the control scenario simulated by the GCM. The GCM simulations are run through the calibrated and validated PLS regression models to obtain future simulations of predictand. The predictands (viz. T_{max} and T_{min} as well as *pan evaporation*) patterns are analyzed with box plots for 20-year time slices. The middle line of the box gives the median whereas the upper and lower edges give the 75 percentile and 25 percentile of the data set, respectively. The difference between the 75 percentile and 25 percentile is known as inter quartile range (IQR). The two bounds of a box plot outside the box denote the value at ×1.5 IQR lower than the third quartile or minimum value, whichever is high and ×1.5 higher than the third quartile or the maximum value whichever is less. Typical results of downscaled predictands (T_{max} and T_{min}) obtained from the predictors are presented in Figs. 7, 8, and 9. 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 T_{max} and T_{min} as well as an evaporation 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) of Figs. 7, 8, and 9, respectively. From the box plots of downscaled predictands (Figs. 7 and 8), it can be observed 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 scenario by using predictors.

Furthermore, the Mann–Kendall test was employed for trend analysis in the present study (Mann 1945; Kendall 1975). This nonparametric test has been extensively used to test randomness against trend. The test was performed for all the scenarios based on GCM downscale predictands. A value of 0.05 was chosen as the local significance level. Based on this significance level, values larger than 1.96 or lower than −1.96, respectively, indicate a significant positive or negative trend (Mishra et al. 2009). The results of the Mann–Kendall test statistics based on the various scenarios for period 2001–2100 are shown in Tables 5 and 6.

Fig. 6 Typical results for comparison of the monthly observed *pan evaporation* with *pan evaporation* simulated using PLR regression downscaling model PLSM3 for NCEP data
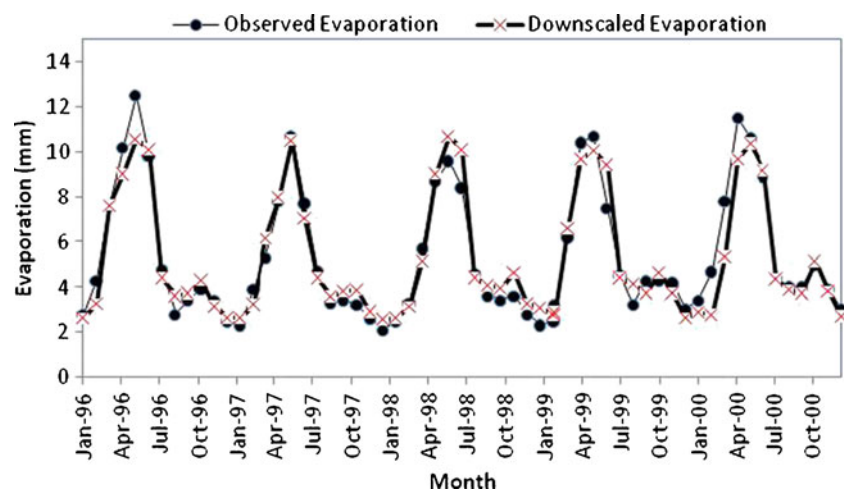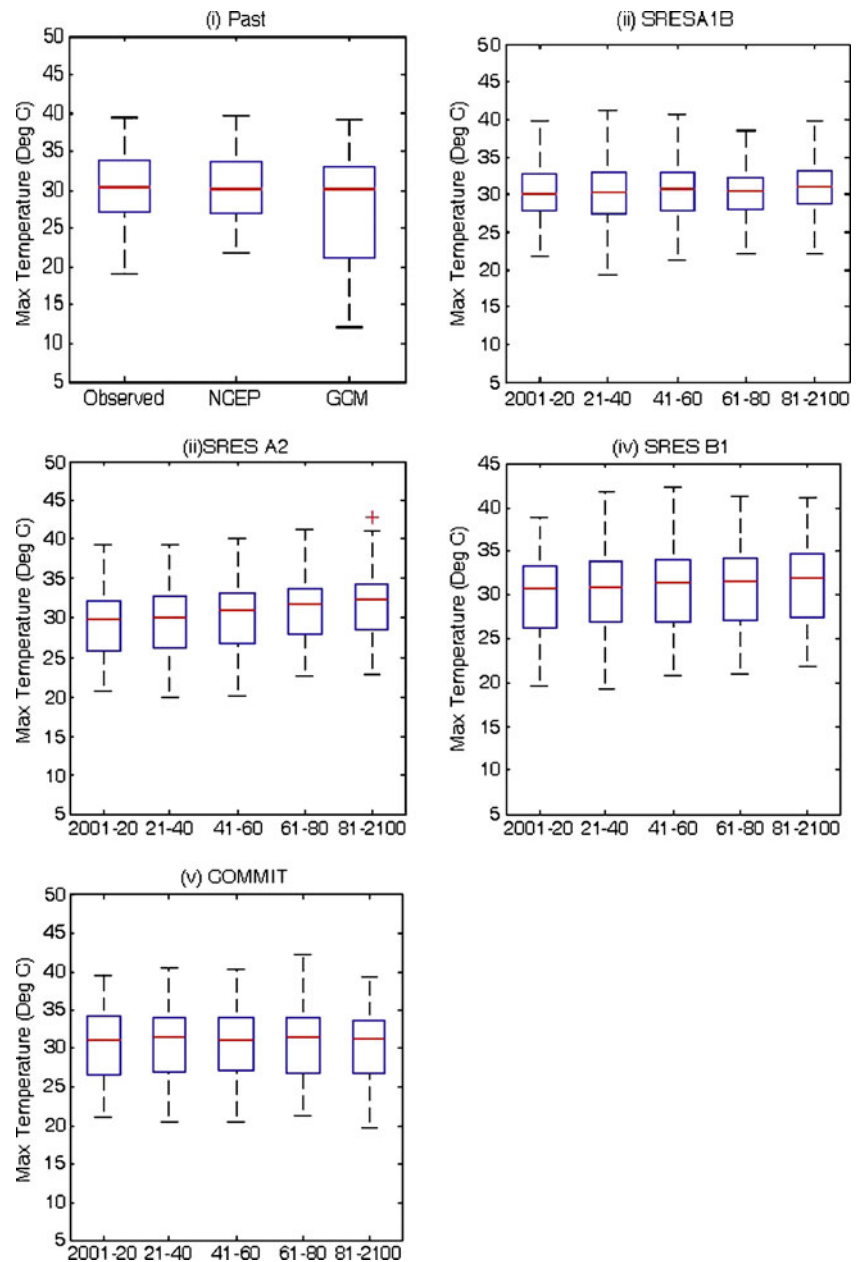
Fig. 7 Box plots results from the PLS regression-based downscaling model PLSM1 for the predictand T_{max}


Historically, T_{max} was observed in the month of June while T_{min} was observed in the month of January in this region. Hence, these months were chosen as a part of this study for trend analysis. It is observed that there is no significant trend, either positive or negative, historically for both the predictands (T_{max} and T_{min}).

For predictand T_{max}, it can be inferred from Table 5 that there is a significant rising trend during May month for SRESB1 and SRESA1B scenario. For predictand T_{min}, it can be observed from Table 6 that there is a significant rising trend for SRESA2 and SRESB1 scenarios for January month for the period of 2001–2100.

Furthermore, it can be concluded that climate would be warmer in the future years. This will increase the vulnerability of the water resource system and further affect the safety of water in the lake catchment. Increase in temperature would result in increase in evapotranspiration which is a major cause of water depletion from riverine systems in arid and semi-arid climates (Dahm et al. 2002). While projected increase in temperatures may enhance the rate of evaporation in the study region since evaporation is proportional to the increase in the earth's surface temperature (Anandhi et al. 2009). However, temperature is only one of the factors that determines the evaporative demand of the atmosphere, the others being vapor 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). Furthermore, increase

Fig. 8 Box plots results from the PLS regression-based downscaling model PLSM2 for the predictand T_{min}
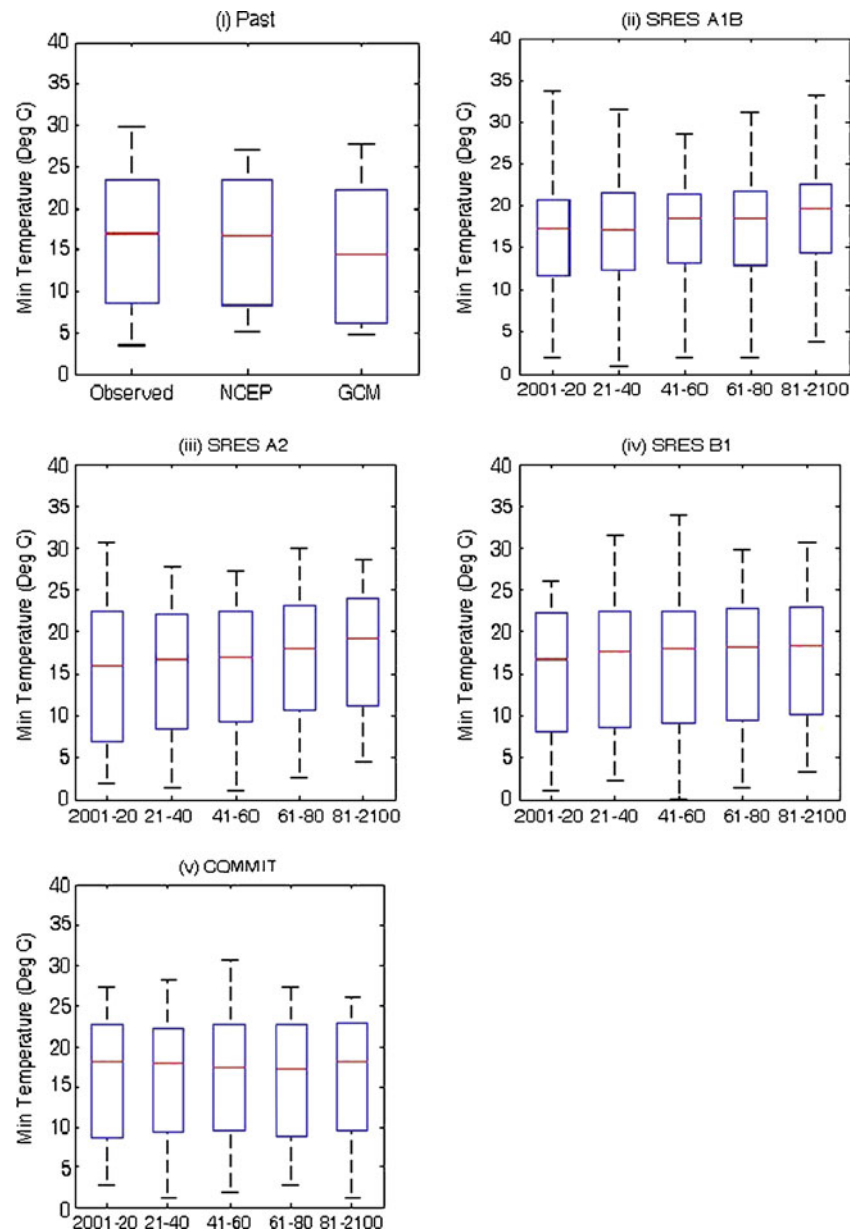
in evaporation may lead to increase in precipitation since the evaporated water would eventually precipitate.

6.3 Comparison with previous downscaling studies

While this is the first study to use PLS regression approach for downsclaing of maximum and minimum temperature as well as pan evaporation prediction in Rajasthan, India, there have been a few studies using other methods in other parts of India. Hence, it is worthwhile to relate the performance of the models presented here with those presented in other studies that closely relate to this study.

In a recent study, Anandhi et al. (2009) developed statistical downscaling models using a support vector machine (SVM) approach for obtaining projections of

monthly mean maximum and minimum temperatures (T_{max} and T_{min}) for a catchment of the Malaprabha reservoir in southern part of India. The analysis reveals that the SVM model is a feasible choice for downscaling the predictands. The resulting models produced similar results to those of this study. For example, 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. However, downscaling of evaporation has not been considered in this study. However, in the case of Anandhi et al. (2009), between the two predictands, T_{max} was better simulated than T_{min}, whereas in this work T_{min} is better simulated than T_{max}. Hence, it has demonstrated that PLS regression downsclaing method used in this study can accurately capture the trend for predictand T_{max} and T_{min}.

Fig. 9 Box plots results from the *PLSM3*-based downscaling model for the predictand pan evaporation
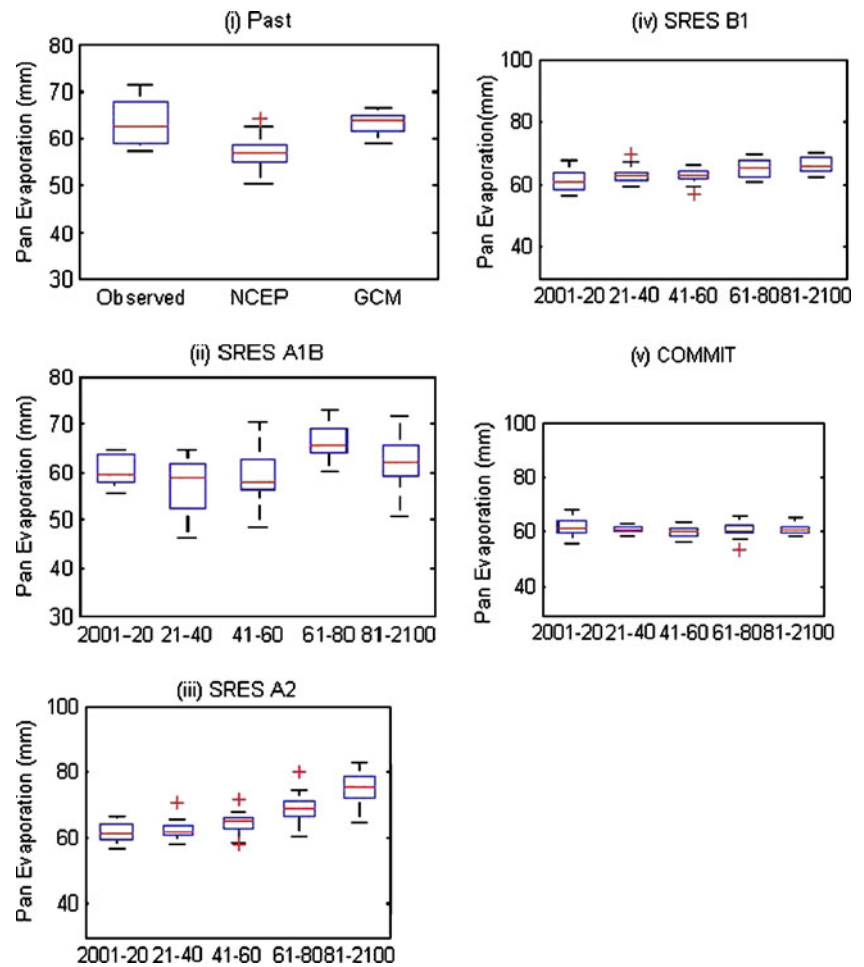
Since, no studies has been reported for downscaling the pan evaporation in India to the best of our knowledge. Hence, a comarison for pan evaporation has been made to a similar study carried out in semi-arid-Haihe River basin, China (Chu et al. 2010). Chu et al. 2010 developed the dowsncaling models for pan evaporation using statistical downsclaing method and results produced are similar to those of this study.

7 Conclusions

Statistical downscaling approaches are generally used to fill the gap between large-scale climate change and local scale response. In this study, PLS regression is applied to the lake catchment in India and we explored its applicability by downscaling mean maximum temperature, mean minimum temperature and pan evaporation simultaneously, which are significant for evaluating the impact of climate change on water resources management. Furthermore, we investigated their trend for future years which would pave the way for the study of hydro-climatological impacts on the lake catchment.

The selection of relevant predictors used for empirical model development plays a crucial role. VIP score obtained from PLS regression has been used for selection of important variables.GCM bias correction procedure improved the overall predictability of predictands. The results of downscaling models using PLS regression 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 scenario. Analysis for months (June for T_{max} while January for T_{min}) with historical T_{max} and T_{min} values reveals that no significant increasing or decreasing trend is found in the observed data at the significance level of 5%.

Table 6 Mann–Kendall statistics for T_{min} based on 2001–2100 for January

T_{min}	SRESA1B	SRESA2	SRESB1	COMMIT	Historical
S	1,020	1,504	1,287	744	7
Test Statistics	2.67	4.26	3.45	2.51	0.46

At the significance level of 5%, it is observed that there is an increasing trend for T_{max} for months of June for various scenarios while there is likely an increasing trend of minimum temperature for all the scenarios for months of January of the year in future. For pan evaporation, it can be concluded that trend is not obvious for future years since the factors working on pan evaporation are complicated.

Appendix

Abbreviations used in text

CCCma: Canadian Center for Climate Modeling and Analysis

CGCM: Canadian Coupled Global Climate Model

CGCM3: Third-generation Canadian Global Climate Model

GCM: Global Climate Model

IPCC: Intergovernmental panel on climate change

NCAR: National Center for Atmospheric Research, USA

RMSE: Root mean square error

SRES: Special report of emission scenarios

Ta 925: Air temperature at 925 hPa

Ua 925: Zonal wind at 925 hPa

Va 925: Meridional wind at 925 hPa

Ta 950: Air temperature at 500 hPa

Va 500: Meridional wind at 500 hPa

Zg 500: geo-potential height at 500 hPa

Ta 200: Air temperature at 200 hPa

Ua 200: Zonal wind at 200 hPa

Va 200: Meridional wind at 200 hPa

References

Abdi H (2003) Partial least squares regression (PLS-regression). In: Lewis-Beck M, Bryman A, Futing T (eds) Encyclopedia for research methods for the social sciences. Sage, Thousand Oaks, pp 792–795

Amor IVM, Hansen JW (2006) Bias correction of daily GCM rainfall for crop simulation studies. Agric For Meteorol 138:44–53

Anandhi A, Srinivas VV, Kumar DN, Nanjundiah RS (2009) Role of predictors in downscaling surface temperature to river basin in India for IPCC SRES scenarios using support vector machine. Int J Climatol 29:583–603

Arnell NW, Hudson DA, Jones RG (2003) Climate change scenarios from a regional climate model: estimating change in runoff in southern Africa. J Geophys Res 108:4519

Benestad RE (2001) A comparison between two empirical downscaling strategies. Int J Climatol 21:1645–1668

Bergant K, Kajfež-Bogataj L (2005) N–PLS regression as empirical downscaling tool in climate change studies. Theor Appl Climatol 81:11–23

Cannon AJ, Lord ER (2000) Forecasting summertime surface-level ozone concentrations in the Lower Fraser Valley of British Columbia: an ensemble neural network approach. J Air Waste Manage Assoc 50:322–339

Cannon AJ, Whitfield PH (2002) Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models. J Hydrol 259(1):136–151

Cavazos T, Hewitson BC (2005) Performance of NCEP variables in statistical downscaling of daily precipitation. Clim Res 28:95–107

Chong IG, Jun C-H (2005) Performance of some variable selection methods when multicollinearity is present. Chemometr Intell Lab Syst 78:103–112

Chu JT, Xia J, Xu CY, Singh VP (2010) Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China. Theor Appl Climatol 99:149–161

Crane RG, Hewitson BC (1998) Doubled CO2 precipitation changes for the Susquehanna basin: down-scaling from the genesis general circulation model. Int J Climatol 18:65–76

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. Freshw Biol 47(4):831

Eriksson L, Johansson E, Kettaneh-Wold N, Wold S (2001) Multi- and megavariate data analysis: principles and applications. Umetrics Academy, Umeå

Gangopadhyay S, Clark M, Rajagopalan B (2005) Statistical downscaling using K-nearest neighbors. Water Resour Res 41:W02024

Gardner MW, Dorling SR (1998) Artificial neural networks (the multi layer perceptron)—a review of applications in the atmospheric sciences. Atmos Environ 32:2627–2636

Ghosh S, Mujumdar PP (2008) Statistical downscaling of GCM simulations to streamflow using relevance vector machine. Adv Water Resour 31:132–146

Goyal MK, Ojha CSP (2009) "Downscaling of surface temperature for lake catchment in Arid Region in India using linear multiple regression and neural networks", Interim Report, Hydraulic Engg. Section, CED, IIT Roorkee

Goyal MK, Ojha CSP (2010a) Downscaling of Surface Temperature for Lake Catchment in Arid Region in India using Linear Multiple Regression and Neural Networks. Int J Climatology, doi:10.1002/joc.2286

Goyal MK, Ojha CSP (2010b) Evaluation of Linear Regression Methods As Downscaling Tool in Temperature Projections Over Pichola lake Basin in India. Hydrological Processes, doi:10.1002/hyp.7911

Goyal MK and Ojha CSP (2010c) Evaluation of Various Linear Regression Methods for Downscaling of Mean Monthly Precipitation in Arid Pichola Watershed. Natural Resources, Scientific Research, 1(1),11–18 doi:10.4236/nr.2010.11002

Goyal MK, Ojha CSP (2010d) Robust Weighted Regression As A Downscaling Tool In Temperature Projections. International Journal of Global Warming, 2(3): 234–251

Goyal MK, Ojha CSP (2010e) Downscaling of Precipitation on a Lake Basin: Evaluation of Rule and Decision Tree Induction Algorithms. Hydrology Research (Accepted)

Hellström C, Chen D, Achberger C, Räisänen J (2001) Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. Clim Res 19:45–55

Hertig E, Jacobeit J (2008) Downscaling future climate change: temperature scenarios for the mediterranean area. Glob Planet Change 63:127–131

Hewitson BC, Crane RG (eds) (1994) Neural nets applications in geography. Kluwer, Dordrecht

Hewitson BC, Crane RG (1996) Climate downscaling: techniques and application. Clim Res 7:85–95

Heyen H, Zorita E, von Storch H (1996) Statistical downscaling of monthly mean North Atlantic air-pressure to sea level anomalies in the Baltic Sea. Tellus 48A:312–323

Intergovernmental Panel on Climate Change (IPCC) (2007) Climate Change 2007—The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. In: Solomon S et al. (ed) Cambridge University Press, Cambridge

Jessie CR, Antonio RM, Stahis SP (1996) Climate variability, climate change and social vulnerability in the semi-arid tropics. Cambridge University Press, Cambridge

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. Bull Am Meteorol Soc 77(3):437–471

Kendall MG (1975) Rank correlation methods. Charles Griffin, London, p 202

Kettle H, Thompson R (2004) Statistical downscaling in European mountains: verification of reconstructed air temperature. Clim Res 26(2):97–112

Khobragade SD (2009) Studies on evaporation from open water surfaces in tropical climate. Ph.D. thesis, Indian Institute of Technology, Roorkee, India

Kilsby CG, Cowpertwait PSP, O'Connell PE, Jones PD (1998) Predicting rainfall statistics in England and Wales using atmospheric circulation variables. Int J Climatol 18:523–539

Lindgren F, Geladi P, Wold S (1993) The kernel algorithm for PLS I, Many observations and few variables. J Chemom 7:45–59

Linz H, Shiklomanov I, Mostefakara K (1990) Chapter 4 hydrology and water likely impact of climate change IPCC WGII report WMO/UNEP, Geneva

Mann HB (1945) Nonparametric tests against trend. Econometrica 13:245–259

Manne R (1987) Analysis of two partial least squares algorithms for multivariate calibration. Chemom Intell Lab Syst 1:187–197

Mishra AK, Ozger M, Vijay PS (2009) Trend and persistence of precipitation under climate change scenarios for Kansabati basin, India. Hydrol Process 23:2345–2357

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models. Part I—a discussion of principles. J Hydrol 10:282–290

Ojha CSP, Goyal MK, Adeloye AJ (2010) Downscaling of Precipitation for Lake Catchment in Arid Region in India using Linear Multiple Regression and Neural Networks. The Open Journal of Hydrology 4:122–136

Prudhomme C, Jakob D, Svensson C (2003) Uncertainty and climate change impact on the flood regime of small UK catchments. J Hydrol 277:1–23

Rannar S, Geladi P, Lindgren F, Wold S (1994) The kernel algorithm for PLS II, Few observations and many variables. J Chemom 8:111–125

Rosenberg NJ, McKenney MS, Martin P (1989) Evapotranspiration in a greenhouse-warmed world: a review and a simulation. Agric For Meteorol 47:303–320

Schoof JT, Pryor SC (2001) Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. Int J Climatol 21:773–790

Shannon DA, Hewitson BC (1996) Cross-scale relationships regarding local temperature inversions at Cape Town and global climate change implications. S Afr J Sci 92(4):213–216

Tripathi S, Srinivas VV, Nanjundiah RS (2006) Downscaling of precipitation for climate change scenarios: a support vector machine approach. J Hydrol 330(3–4):621–640

Vimont DJ, Battisti DS, Naylor RL (2009) Downscaling Indonesian precipitation using large-scale meteorological fields. Int J Climatol. doi:10.1002/joc.2010

Weisse R, Oestreicher R (2001) Reconstruction of potential evaporation for water balance studies. Clim Res 16(2):123–131

Wilby RL (1998) Modelling low-frequency rainfall events using airflow indices, weather patterns and frontal frequencies. J Hydrol 213(1–4):380–392

Wilby RL, Dawson CW, Barrow EM (2002) SDSM—a decision support tool for the assessment of climate change impacts. Environ Modell Softw 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)

Willmott CJ, Rowe CM, Philpot WD (1985) Small-scale climate map: a sensitivity analysis of some common assumptions associated with the grid-point interpolation and contouring. Am Cartogr 12:5–16

Wold S (1995) PLS for multivariate linear modelling. In: van de Waterbeemd H (ed) QSAR: chemometric methods in molecular design, vol 2. Wiley, Weinheim, pp 195–218

Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemometr Intell Lab Syst 58:109–130

Xoplaki E, Luterbacher J, Burkard R, Patrikas I, Maheras P (2000) Connection between the large-scale 500 hPa geopotential height fields and precipitation over Greece during wintertime. Clim Res 14:129–146

Zhang XC, Nearing MA, Garbrecht JD, Steiner JL (2004) Downscaling monthly forecasts to simulate impacts of climate change on soil erosion and wheat production. Soil Sci Soc Am J 68(4):1376–1385