Downscaling of Precipitation for Lake Catchment in Arid Region in India using Linear Multiple Regression and Neural Networks

C. S. P. Ojha[1], Manish Kumar Goyal[2,*] and A. J. Adeloye[3]

[1]Dept. of Civil Engineering, Indian Institute of Technology, Roorkee, India

[2]Dept. of Civil Engineering, Indian Institute of Technology, Roorkee, India

[3]School of the Built Environment, Heriot-Watt University, Edinburgh, UK

Abstract: In this paper, downscaling models are developed using a Linear Multiple Regression (LMR) and Artificial Neural Networks (ANNs) for obtaining projections of mean monthly precipitation to lake-basin scale in an arid region in India. The effectiveness of these techniques is demonstrated through application to downscale the predictand (precipitation) for the Pichola lake region in Rajasthan state in India, which is considered to be a climatically sensitive region. The predictor variables are extracted from (1) the National Centers for Environmental Prediction (NCEP) reanalysis dataset for the period 1948-2000, 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 2001-2100. The scatter plots and cross-correlations are used for verifying the reliability of the simulation of the predictor variables by the CGCM3. The performance of the linear multiple regression and ANN models was evaluated based on several statistical performance indicators. The ANN based models is found to be superior to LMR based models and subsequently, the ANN based model is applied to obtain future climate projections of the predictand (i.e precipitation). The precipitation is projected to increase in future for A2 and A1B scenarios, whereas it is least for B1 and COMMIT scenarios using predictors. In the COMMIT scenario, where the emissions are held the same as in the year 2000.

Keywords: Artificial neural network, downscaling, precipitation, regression, climate change, IPCC SRES scenarios.

1. INTRODUCTION

General Circulation Models (GCMs) are the most powerful tools available to simulate evolving and future changes in the climate system. GCMs are able to simulate reliably the most important mean features of the global climate at planetary scales. Global circulation models (GCMs) are numerical models that represent the large-scale physical processes of the earth-atmosphere-ocean system and have been designed to simulate the past, present, and future climate [1-3].

Precipitation is an important parameter for climate change impact studies. A proper assessment of probable future precipitation and its variability is to be made for various water resources planning and hydro-climatology scenarios. Recently, downscaling of precipitation has found wide application in hydro-climatology on various time scale for scenario construction and simulation of (i) low-frequency rainfall events [4] (ii) daily precipitation [5] (iii) seasonal precipitation [6] (iv) daily and monthly precipitation [7] (v) monthly precipitation [8] (vi) monthly precipitation [9] (vii) seasonal precipitation [10], daily precipitation [11] and monthly precipitation [12] (viii) monthly precipitation [13] and annual precipitation [14].

The GCMs are usually run at coarse-grid resolution and as a result, fields from GCMs are mostly inappropriate for direct application because of the limited and poor representation of sub-grid-scale features like orography, land use, and dynamics of mesoscale processes. While these models are most accurate at large (continental, hemispheric, and global) spatial scales, smaller-scale (at or near the spatial resolution of the GCMs) climatic details are less well portrayed [1, 15-16]. This makes them unsuitable to many impact modelers, particularly hydrologists and water resources planner interested in local/regional-scale hydrological variability. Hence, a variety of approaches to the ‘downscaling’ of grid-scale (hundreds of km) GCM information to local-scale surface climate have been devised in last few decade [17-19].

Artificial Neural Networks (ANNs) are used in this application to derive relationships between the grid circulation variables and the local climatic variables. This provides a powerful base learner, with advantages such as nonlinear mapping and noise tolerance, increasingly used in the Data Mining (DM) and Machine Learning (ML) fields [20]. An ANN is characterized by an architecture that represents the pattern of connection between nodes, a method for determining the connection weights, and an activation function [21]. ANNs are analogous in application to multiple regression, with the added advantage that they are inherently non-linear, and particularly robust in finding and representing relationships in the presence of noisy data. The application of ANNs and utility for downscaling applications may be found in Sailor *et al.* [3]; Hewitson and Crane [22]; and Schoof and Pryor [23]. ANNs have proved particularly effective in downscaling precipitation and temperature, where there is a

*Address correspondence to this author at the Dept. of Civil Engineering, Indian Institute of Technology, Roorkee, India; Tel: +91-1332-285494; Fax: +91-1332-273560; E-mail: goyaldce@iitr.ernet.in

significant non-linear relationship that more traditional techniques such as regression do not capture well [23].

The objective of this study is to assess the effectiveness of neural networks to downscale mean monthly precipitation by comparing with linear multiple regression (LMR) on a lake catchment in an arid region from simulations of CGCM3 for latest IPCC scenarios. 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.

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 how the various predictor variables behave for the different scenarios and the reasons for selection of the probable predictor variables for downscaling. Section 5 explains the proposed methodology for development of the regression based and ANN based model for downscaling precipitation to the lake basin. Section 6 presents the results and discussion. Finally, Section 7 provides the conclusions drawn from the study.

2. STUDY REGION

The study area of the research is the Pichola lake catchment in Rajasthan state in India that is situated from 72.5°E to 77.5°E and 22.5°N to 27.5°N. It receives an average annual precipitation of 597 mm. It has a tropical monsoon climate where most of the precipitation is confined to a few months of the monsoon season. The south-west (summer) monsoon has warm winds blowing from the Indian Ocean causing copious amount of precipitation during June-September months. The location map of the study region is shown in Fig. (**1**). The observed monthly precipitation has been shown in Fig. (**2**). for various months of year 2000.

Pichola lake is about 3.62 km in length from north to south and 2.41 km in width from east to west with a mean depth of 5.6 m. It is estimated that the lake contains 418 million cubic feet of water and covers an area of 9.71 sq km. It is fed mainly from rainwater and also from the Sisarma tributary [24].

The Pichola lake basin is one of the major sources for water supply for this arid region. During the past several decades, the streamflow regime in this catchment has

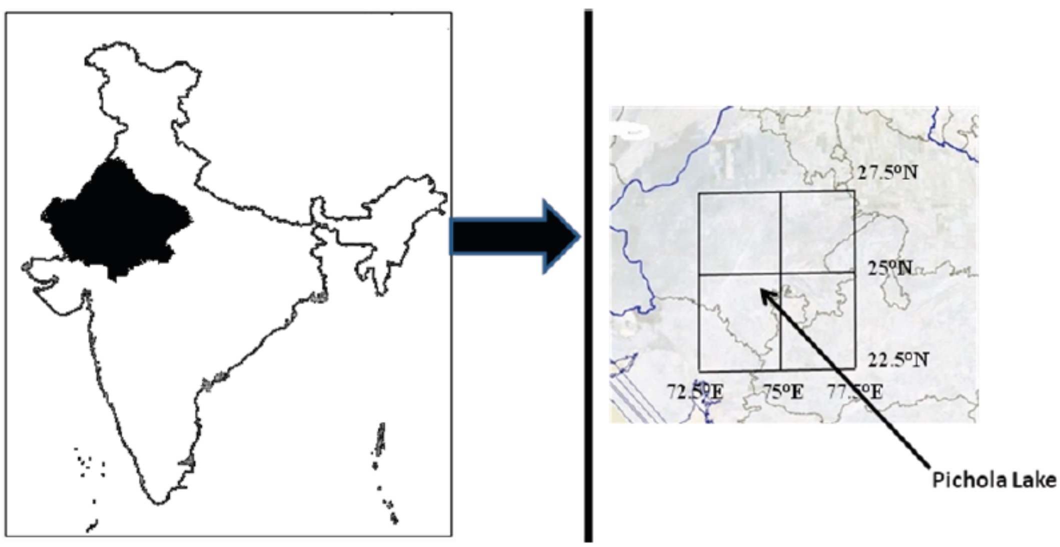

Fig. (1). Location map of the study region in Rajasthan State of India with NCEP grid.

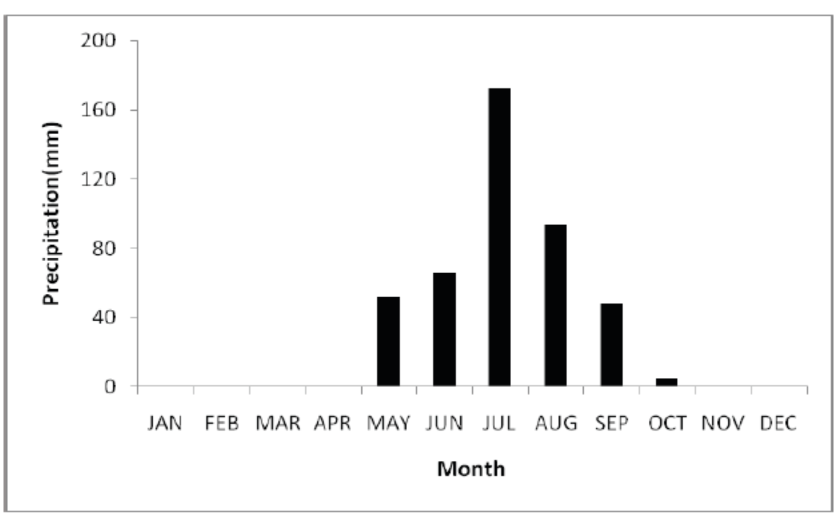

Fig. (2). Observed Precipitation for the study region.

changed considerably, which resulted in water scarcity, low agriculture yield and degradation of the ecosystem in the study area [25]. Regions with arid and semi-arid climates could be sensitive even to insignificant changes in climatic characteristics [26]. Investigations of IPCC (2001) indicate that the mean monsoon precipitation in the region will intensify in future [27]. Understanding the relationships among the hydrologic regime, climate and environmental factors, and anthropogenic effects is important for the sustainable management of water resources in the entire catchment. The motivation of the present study is the need to assess plausible impact of climate change on precipitation in the study region, which has implications on inflows into the Pichola lake which is frequently prone to water shortage and is considered to be a climatically sensitive region and hence, this study area was chosen because of aforementioned reasons.

3. DATA EXTRACTION

Reanalysis Data

The monthly mean atmospheric climatic variables were derived from the National Center for Environmental Prediction (NCEP/NCAR) (hereafter called NCEP) reanalysis data set for a period of January 1948 to December 2000 [28]. The data have a horizontal resolution of 2.5° latitude x longitude and seventeen constant pressure levels in 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°.

Meteorological Data

The precipitation data is used at monthly and annual time scale for Pichola lake which is located in Udaipur district at 24° 34'N latitude and 73°40'E longitude [25]. Data were available for the period 1974 to 2000 at annual time scale and were available for the period January 1990 to December 2000 at monthly time scale.

GCM Data

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 by 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. Data was obtained for CGCM3 climate of the 20th Century (20CM3) experiments used in this study. Herein, it is to be mentioned that the spatial domain of climate variables is chosen following the suggestions in Wilby and Wigley [19].

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 [29]. The utility of this interpolation algorithm was examined in previous downscaling studies [8, 30-34].

The development of downscaling models for the predictand variable precipitation begins with selection of potential predictors, followed by training and validation of the LMR and ANN downscaling model. The developed model is then used to obtain projections of precipitation from simulations of CGCM3.

4. SELECTION OF PREDICTORS

For downscaling predictand, the selection of appropriate predictors is one of the most important steps in a downscaling exercise. The predictors are chosen by the following criteria: (1) they should be skillful in representing large-scale variability that is simulated by the GCMs and are readily available from archives of GCM output and reanalysis data sets; (2) they should strongly correlated with the with the surface variables of interest/predictands i.e. they should be statistically significant contributors to the variability in precipitation; (3) they should represent important physical processes in the context of the enhanced greenhouse effect [33-35]. Various authors have used large-scale atmospheric variables, namely air temperature (at 925, 500 and 200 mb pressure levels), geopotential height (at 500 and 200 mb pressure levels), zonal (u) and meridional (v) wind velocities (at 925 and 200 mb pressure levels), as the predictors for downscaling GCM output to mean monthly precipitation over a catchment [8,17,36,37].

As suggested by Wilby *et al.* [38], predictors have to be selected based both on their relevance to the downscaled predictands and their ability to be accurately represented by the GCMs. Scatter plots and cross-correlations are in use to select predictors to understand the presence of nonlinearity/linearity trend in dependence structure [39]. 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. Scatter plots are prepared and cross-correlations are computed between the predictor variables in NCEP and GCM datasets (Figs. **3** and **4**). The cross correlations are estimated using three measures of dependence namely, product moment correlation [40], Spearman's rank correlation [41,42] and Kendall's tau [43]. 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.

5. DEVELOPMENT OF DOWNSCALING MODEL

For downscaling precipitation, the probable predictor variables that have been selected to develop the models are considered at each of the nine grid points surrounding and within the study region (shown in Fig. **2**). Cross-correlations are computed between the probable predictor variables in NCEP and GCM data sets. Subsequently, a pool of potential predictors is identified by specifying threshold values for the computed cross-correlations. In order to relate the large-scale weather patterns to the local scale, downscaling is necessary. The relationships between these scales can be determined by a number of methods including regression, canonical correlation analysis [44,45], artificial neural networks [23, 46,47].

In this study, linear multiple regression and artificial neural networks (ANNs) are used to downscale mean monthly precipitation. The data of potential predictors is first stan-

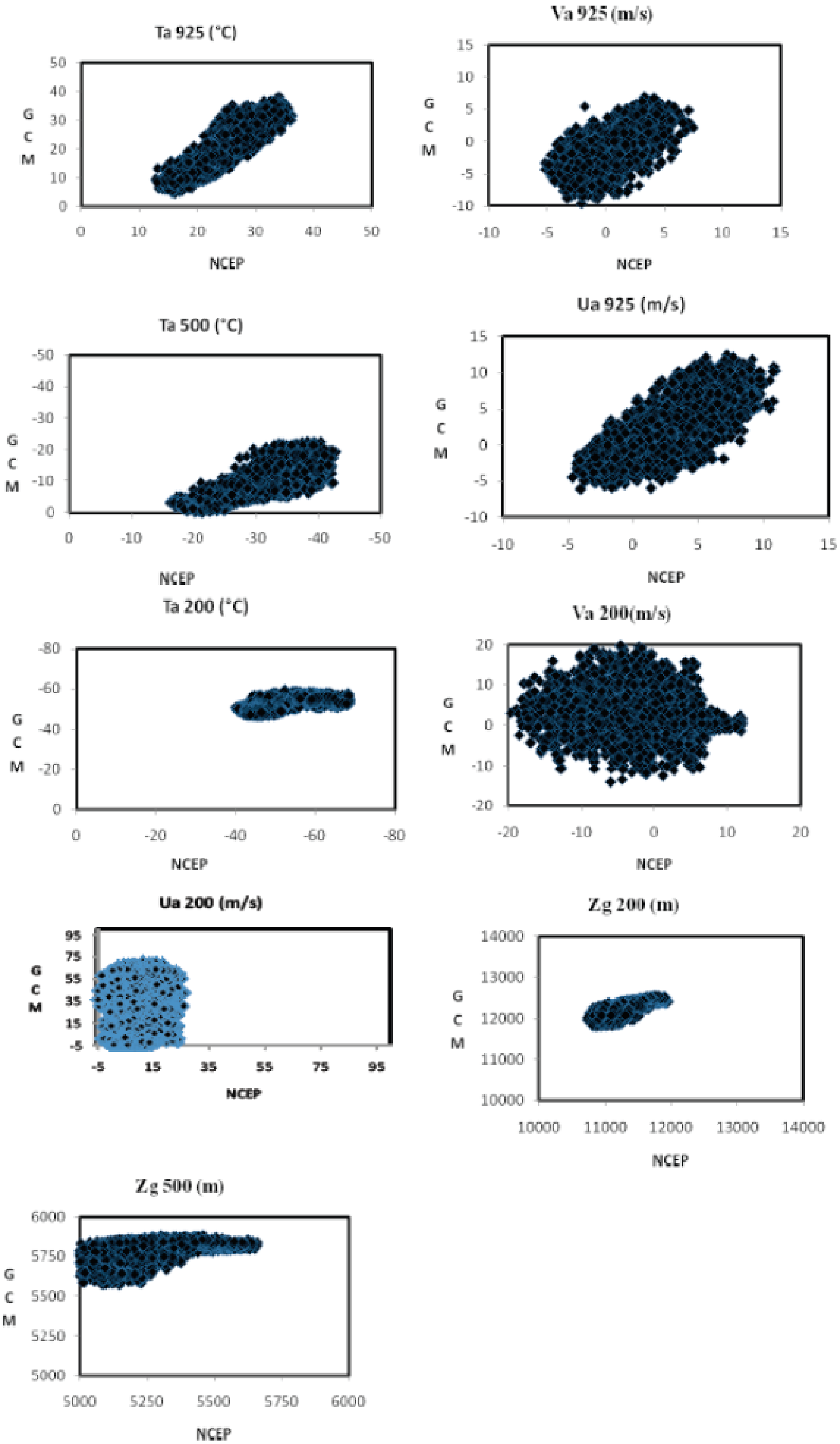

Fig. (3). Scatter plots prepared to investigate dependence structure between probable predictor variables in NCEP and GCM datasets.

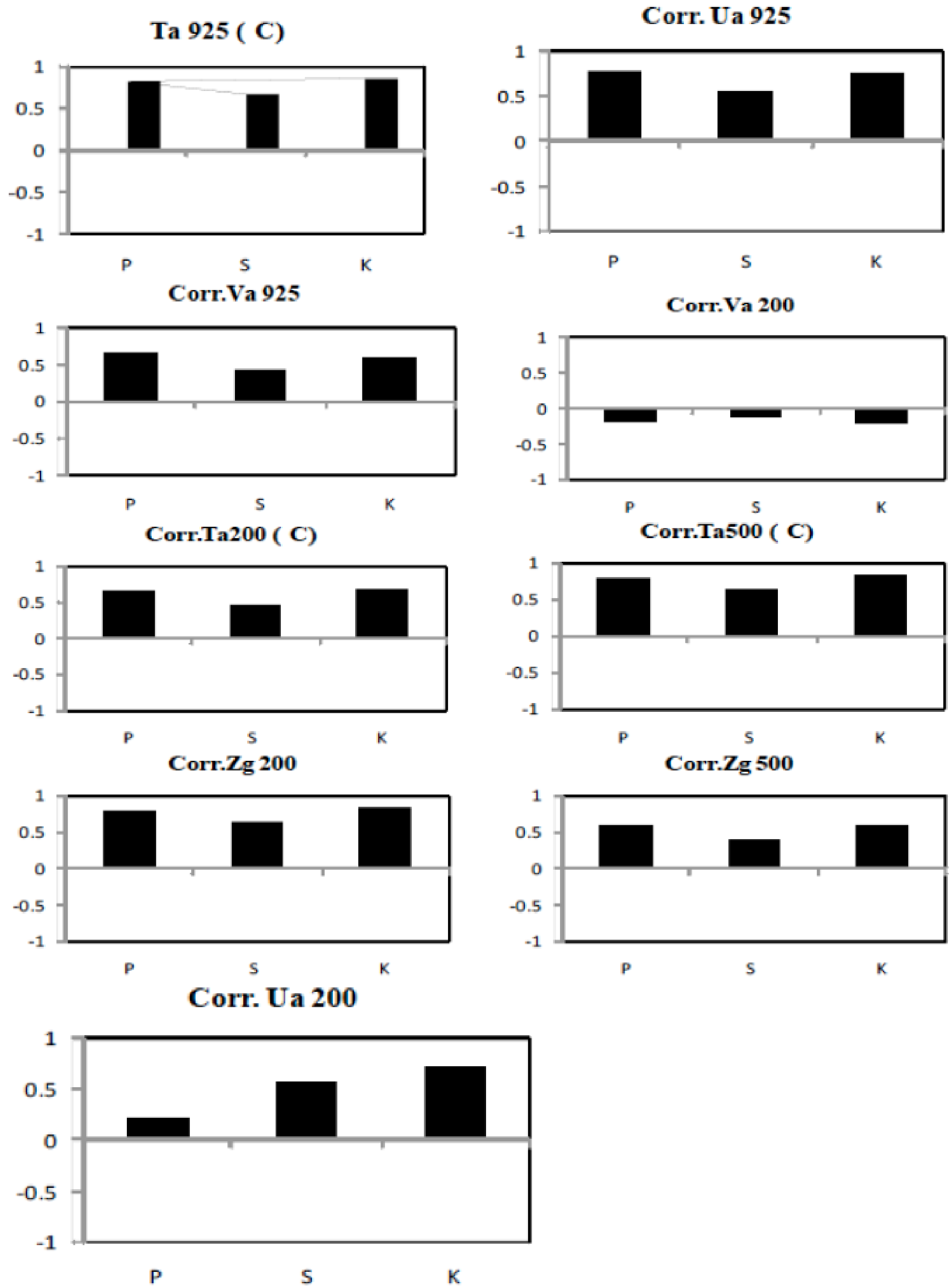

Fig. (4). Bar plots for cross-correlation computed between probable predictors in NCEP and GCM datasets. P, S and K represent product moment correlation, Spearman's rank correlation and Kendall's tau respectively.

dardized. 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 [38]. Fig. (**2**) shows the grid points superposed on the map of Rajasthan state of India. In this study, 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 [38]. The dimension of the GCM output dataset extracted is 9X9=81 (air temperature (at 925, 500 and 200 mb pressure levels), geopotential height (at 500 and 200 mb pressure levels), zonal and meridional wind velocities (at 925 and 200 mb pressure levels) at each of the nine grid points).

Linear multiple regression are performed on dimensionality set of selected predictors. Multi-dimensionality of the predictors may lead to a computationally complicated and large sized model with high muti-collinearity (high correlation between the explanatory variables/regressors). To reduce the dimensionality of the explanatory dataset, Principal Component Analysis (PCA) is performed. The use principal component (PCs) as input to a downscaling model helps in making the model more stable and at the same time reduces it computationally burden. 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/year of the record using the PCs. The feature vector is the input to the linear multiple regression and ANN model, and the contemporaneous value of predictand is the output.

To develop the linear multiple regression and ANN downscaling model, the feature vectors which are prepared from NCEP record are partitioned into a training set and a 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 27-year annual observed temperature-data series was broken up into a calibration period and a validation period. The models were calibrated on the calibration period 1974 to 1994 and validation involved period and 1995 to 2000. The monthly data series was broken from January 1990 to December 1995 as calibration period and from January 1996 to December 2000 as validation period. The various error criteria are used as an index to assess the performance of the model. Based on the latest IPCC scenario, total 10 models were constructed for predictand of both approaches (LMR and ANN). These models for mean monthly and annual precipitation were evaluated based on the accuracy of the predictions for training and testing data set. Table **I** and **III** shows the values of regression coefficients of regression models at annual and monthly scale respectively while Table **II** and **IV** shows certain details of different ANN downscaling models at annual and monthly scale respectively. For linear multiple regression, there will be two models (viz *LMRM1* and *LMRM2*). LMRM1 and LMRM2 refer to linear multiple regression models using principal component analysis at annual and monthly time scale respectively. For NN, there will be eight models (viz. *ANNM1* to *ANNM8*), one for each IPCC scenario at annual and monthly time scale. ANNM1 to ANNM4 denotes to artificial neural network

Table I. Description of Regression Models, Input Values and Model Forms at Annual Time Scale*

Model	Predictand	Equation
LMRM1	Precipitation(P)	$(P)= -34.17+0.9391PC_1-0.0790PC_2+0.1844PC_3+0.0638\ PC_4$

*The predictors in the regression equations (PC#) indicate principal component.

Table II. Different ANN Downscaling Model Variants Used in the Study for Obtaining Projections of Predictand Precipitation at Annual Time Scale

Predictand	Period of downscaling	Length of the record	Scenario	Model
Precipitation(P)	1974-2100	1974-2000	SRESA1B	ANNM1
Precipitation(P)	1974-2100	1974-2000	SRESA2	ANNM2
Precipitation(P)	1974-2100	1974-2000	SRESB1	ANNM3
Precipitation(P)	1974-2100	1974-2000	COMMIT	ANNM4

Table III. Description of Regression Models, Input Values and Model Forms at Monthly Time Scale*

Model	Predictand	Equation
LMRM2	Precipitation(P)	$(P)=-15.94+5.00PC_1-0.93PC_2+0.54PC_3+0.66PC_4+0.09PC_5-0.38\ PC_7$

*The predictors in the regression equations (PC#) indicate principal component.

Table IV. Different ANN Downscaling Model Variants Used in the Study for Obtaining Projections of Predictand Precipitation at Monthly Time Scale

Predictand	Period of downscaling	Length of the record	Scenario	Model
Precipitation(P)	1990-2100	1990-2000	SRESA1B	ANNM5
Precipitation(P)	1990-2100	1990-2000	SRESA2	ANNM6
Precipitation(P)	1990-2100	1990-2000	SRESB1	ANNM7
Precipitation(P)	1990-2100	1990-2000	COMMIT	ANNM8

model at annual time scale and ANNM5 to ANNM8 denotes to artificial neural network model at monthly time scale.

6. RESULTS AND DISCUSSIONS

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

6.1. Potential Predictor Selection

The most relevant probable predictor variables necessary for developing the downscaling models 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. The scatter plots between the probable predictor variables in NCEP and GCM datasets are shown in Fig. (**3**), while the cross correlations computed between the same are shown in Fig. (**4**). In general, the most of predictor variables are realistically simulated by the GCM where CC was greater than 0.65. It is noted that air temperature at 925 mb (Ta 925) is the most realistically simulated variable with a CC greater than 0.8, while meridional wind at 200mb (Va 200) is the least correlated variable between NCEP and GCM datasets (CC = -0.17). It is clear from Figs. (**3** and **4**) that air temperature at 925 mb (Ta 925), air temperature at 500 mb (Ta 500), air temperature at 200 mb (Ta 200), meridional wind at 925mb (Va 925), zonal wind at 925mb (Ua 925) and zeo-potential height at 200mb (Zg 200mb) are better correlated than meridional wind at 200mb (Va 200), zonal wind at 200mb (Ua 200) and zeo-potential height at 500mb (Zg 500).

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 such as to re-gridding from GCM scale to NCEP scale also contribute to low correlation.

6.2. Downscaling and Performance of GCM Models

Six predictor variables namely air temperature (925 mb), zonal wind (925 mb) , meridoinal wind (925 mb), air temperature (200 mb), air temperature (500 mb) and zeo-potential height(200mb) at 9 NCEP grid points with a dimensionality of 54, are used which are highly correlated with each other. Multiple linear regressions were performed on these data sets. Principal Component Analysis (PCA)[32, 48] is performed to transform the set of correlated N-dimensional predictors (N = 54) into another set of N-dimensional uncorrelated vectors (called principal components) by linear combination, such that most of the information content of the original data set is stored in the first few dimensions of the new set. It is observed that the four leading principal components (PCs) of the PCA method explained about 98% of the information content (or variability) of the original predictors. Hence, PCs are extracted to form feature vectors from the standardized data of potential predictors. These feature vectors are provided as input to the linear multiple regression and ANN downscaling model.

The different statistical parameters of each model are adjusted 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 (CR), Standard Error of Estimate (SSE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Normalized Mean square Error (NMSE), Nash–Sutcliffe Efficiency Index and Mean Absolute Error (MAE) were used to measure the performance of various models.

Results of various statistics of linear multiple regression models have been presented in Table **V** and **VI**. It can be inferred from Table **V** and Table **VI** that both linear multiple regression models were not performed well in terms of all performance indicators.

The architecture of ANN is decided by trial and error procedure. A comprehensive search of ANN architecture is done by varying the number of nodes in hidden layer. The network is trained using back-propagation algorithm. Tan sigmoid activation function has been used in hidden layer(s), whereas linear activation function has been used in the output layer. The network error is computed by comparing the network output with the target or the desired output. Mean square error is used as an error function. Results of the different models (*ANNM1* to *ANNM8*) as discussed in Table **II** and Table **IV** are tabulated (see Table **VII** and Table **VIII**):

It can be observed from Table **V** to Table **VIII** that the performance of ANNs for mean monthly and annually precipitation is clearly superior to that of LMR based models. All statistical performance indicators have performed better between predicted and observed value for ANN models.

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 simu-

Table V. Model Evaluation Statistics for Regression Models at Annual Time Scale

Model	CR		SSE		MSE	
	Training	Validation	Training	Validation	Training	Validation
LMRM1	0.61	0.75	388371.93	15027.78	18493.90	17627.37

RMSE		NMSE		N-S Index		MAE	
Training	Validation	Training	Validation	Training	Validation	Training	Validation
135.99	132.77	0.60	0.85	0.37	0.35	0.21	0.23

Table VI. Model Evaluation Statistics for Regression Models at Monthly Time Scale

Model	CR		SSE		MSE	
	Training	Validation	Training	Validation	Training	Validation
LMRM2	0.86	0.80	155775.41	123152.74	2163.55	2052.55

RMSE		NMSE		N-S Index		MAE	
Training	Validation	Training	Validation	Training	Validation	Training	Validation
46.51	45.31	0.26	0.45	0.73	0.54	0.59	0.42

Table VII. Various Performance Statistics for Various ANN Models at Annual Time Scale

Model	Hidden Nodes	CR		SSE		MSE	
		Training	Validation	Training	Validation	Training	Validation
ANNM1	6	0.46	0.08	485050.02	24649.69	23097.62	23485.64
ANNM2	5	0.97	0.02	36194.66	67826.98	1723.56	18249.41
ANNM3	8	0.80	0.53	221204.86	35379.82	10533.56	16745.13
ANNM4	6	0.83	0.54	188980.06	28990.27	8999.05	13996.86

RMSE		NMSE		N-S Index		MAE	
Training	Validation	Training	Validation	Training	Validation	Training	Validation
0.75	1.13	0.22	0.13	0.18	0.13	0.75	1.13
41.52	135.09	0.06	0.88	0.94	0.39	0.78	0.52
102.63	129.40	0.34	0.81	0.64	0.41	0.48	0.31
94.86	118.31	0.29	0.67	0.69	0.50	0.48	0.37

Table VIII. Various Performance Statistics for Various ANN Models at Monthly Time Scale

Model	Hidden Nodes	CR		SSE		MSE	
		Training	Validation	Training	Validation	Training	Validation
ANNM5	6	0.94	0.80	68094.79	120763.50	945.76	2012.73
ANNM6	6	0.96	0.73	51541.18	159217.85	715.85	2653.63
ANNM7	6	0.93	0.73	73288.65	167504.75	1017.90	2791.75
ANNM8	6	0.95	0.77	54361.41	136258.78	755.02	2270.98

RMSE		NMSE		N-S Index		MAE	
Training	Validation	Training	Validation	Training	Validation	Training	Validation
30.75	44.86	0.12	0.44	0.88	0.55	0.76	0.50
26.76	51.51	0.09	0.59	0.91	0.40	0.76	0.43
31.90	52.84	0.12	0.62	0.87	0.37	0.75	0.38
27.48	47.65	0.09	0.50	0.91	0.49	0.75	0.44

lations are run through the calibrated and validated NN downscaling models to obtain future simulations of predictand. The predictand 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 1.5X IQR lower than the third quartile or minimum value, whichever is high and 1.5X higher than the third quartile or the maximum value whichever is less. Typical results of downscaled predictand (precipitation) obtained from the predictors are presented in Figs. (**5**). In part (i) of these figures, the precipitation downscaled using NCEP and GCM datasets are compared with the observed Precipitation 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.

From the box plots of downscaled predictand (Fig. **5**), it can be observed that precipitation are projected to increase in future for A1B, A2 and B1 scenarios. The projected increase of precipitation is high for A1B and A2 scenarios whereas it is least for B1 scenario. This is because among the scenarios considered, the scenario A1B and A2 have the highest concentration of atmospheric carbon dioxide (CO2) equal to 720 ppm and 850 ppm, while the same for B1 and COMMIT scenarios are 550 ppm and ≈370 ppm respectively. Rise in concentration of CO2 in the atmosphere causes the earth's average temperature to increase, which in turn causes increase in evaporation especially at lower latitudes. The evaporated water would eventually precipitate [17]. In the COMMIT scenario, where the emissions are held the same as in the year 2000, no significant trend in the pattern of projected future precipitation could be discerned. The overall results show that the projections obtained for precipitation are indeed robust. A comparison of mean annual observed precipitation with precipitation simulated using several ANN downscaling models(viz. ANNM1 to ANNM4) have been shown in Fig. (**6** to **9**) for calibration and validation period. Calibration period is from 1974 to 1995, and the rest is validation period. Similarly, a comparison of mean monthly observed precipitation with precipitation simulated using several ANN downscaling models (viz. ANNM5 to ANNM8) have been shown in Figs. (**10** to **13**) for calibration

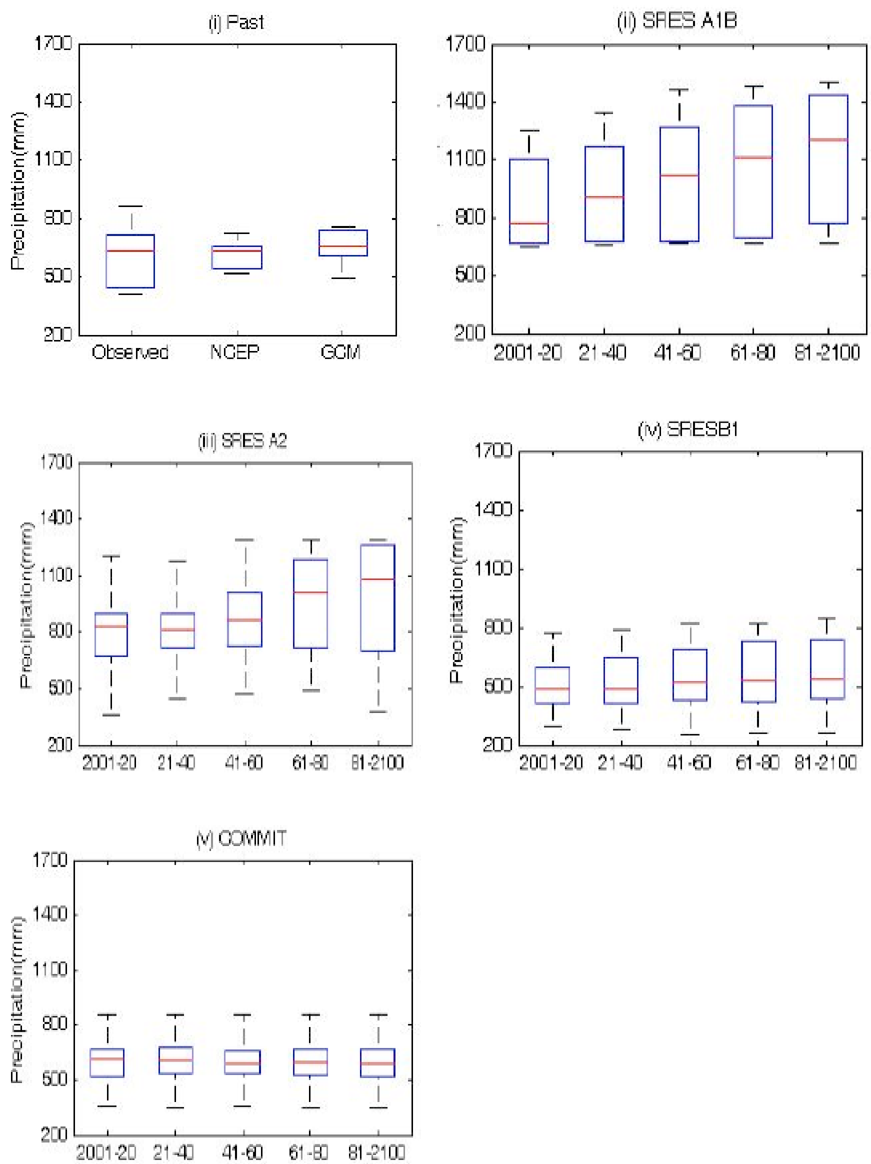

Fig. (5). Box plots results from the ANN-based downscaling model for the predictand precipitaion.

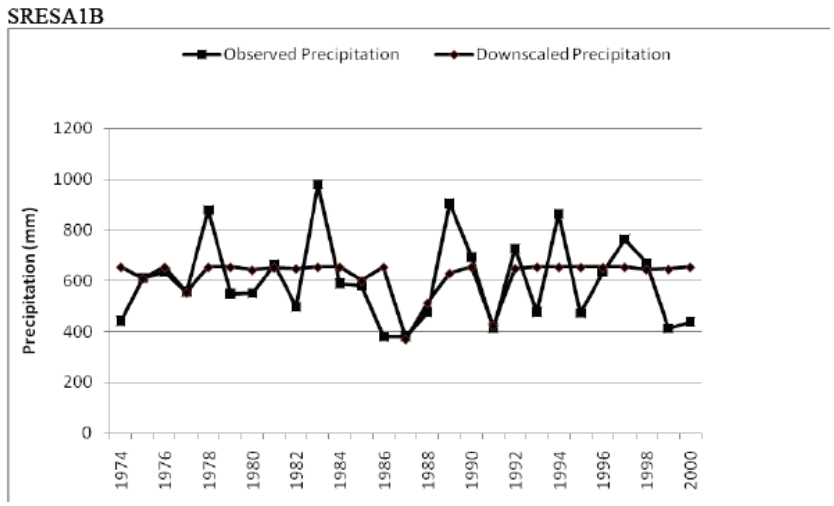

Fig. (6). Typical results for comparison of the mean annual observed precipitation with precipitation simulated using ANN downscaling model ANNM1 for NCEP data. In the figure, calibration period is from 1974 to 1994, and the rest is validation period.

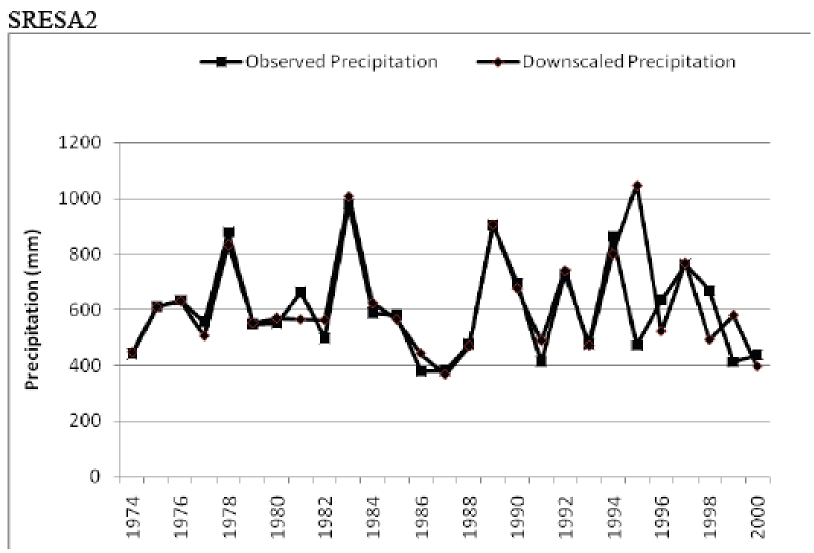

Fig. (7). Typical results for comparison of the mean annual observed precipitation with precipitation simulated using ANN downscaling model ANNM2 for NCEP data. In the figure, calibration period is from 1974 to 1994, and the rest is validation period.

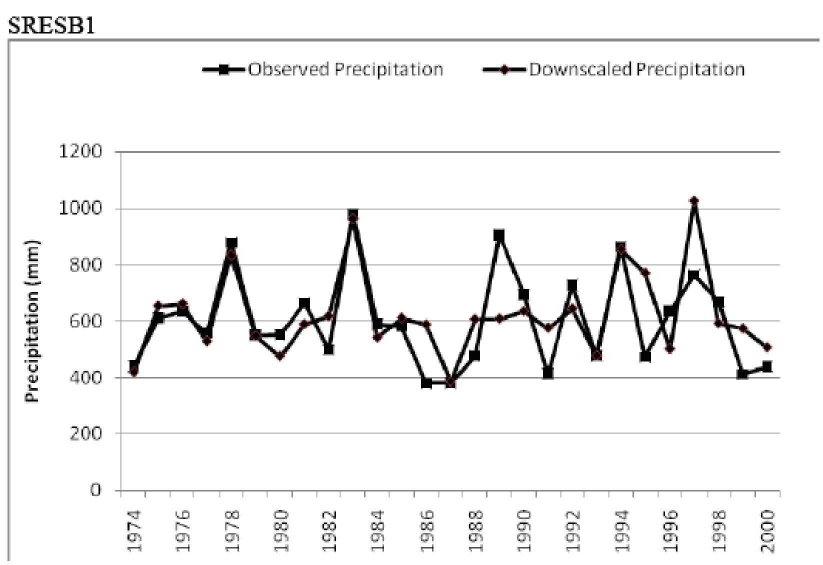

Fig. (8). Typical results for comparison of the mean annual observed precipitation with precipitation simulated using ANN downscaling model ANNM3 for NCEP data. In the figure, calibration period is from 1974 to 1994, and the rest is validation period.

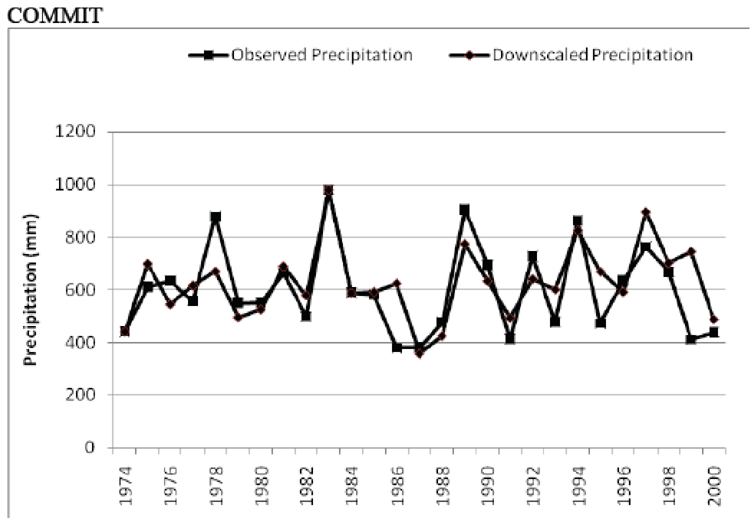

Fig. (9). Typical results for comparison of the mean annual observed precipitation with precipitation simulated using ANN downscaling model ANNM4 for NCEP data. In the figure, calibration period is from 1974 to 1994, and the rest is validation period.

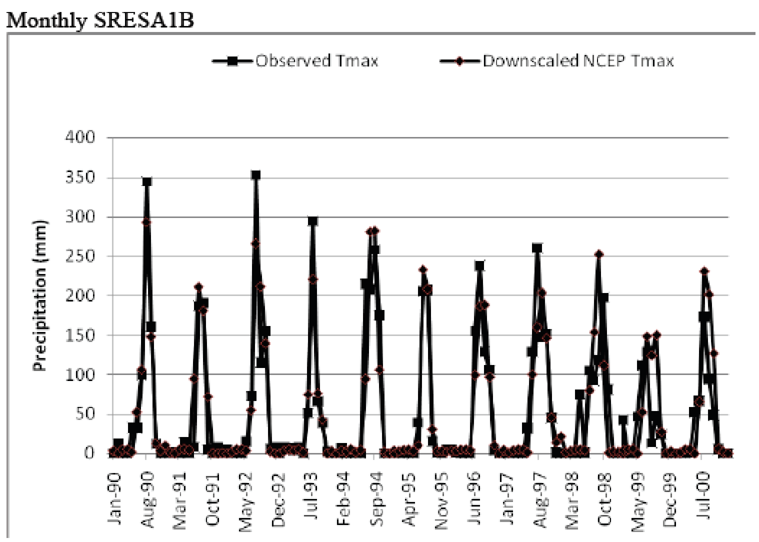

Fig. (10). Typical results for comparison of the amean monthly observed precipitation with precipitation simulated using ANN downscaling model ANNM5 for NCEP data. In the figure, calibration period is from January 1990 to Decmber 1995, and the rest is validation period.

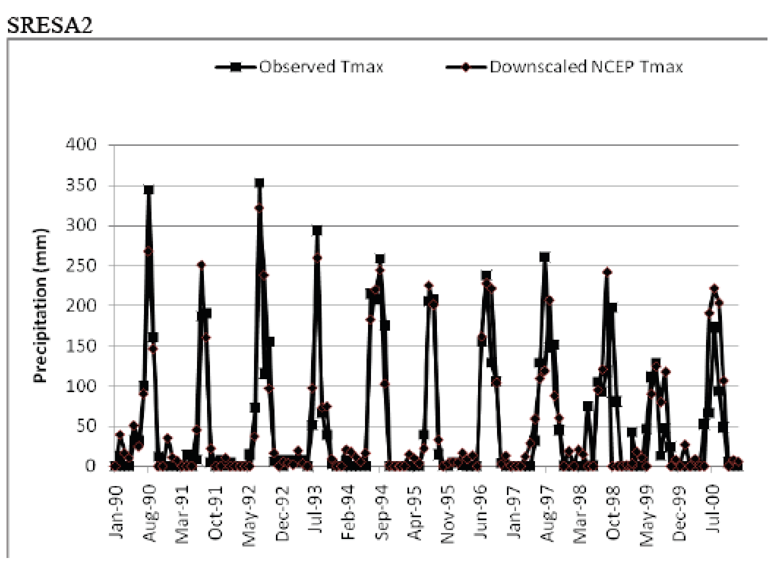

Fig. (11). Typical results for comparison of the amean monthly observed precipitation with precipitation simulated using ANN downscaling model ANNM6 for NCEP data. In the figure, calibration period is from January 1990 to Decmber 1995, and the rest is validation period.

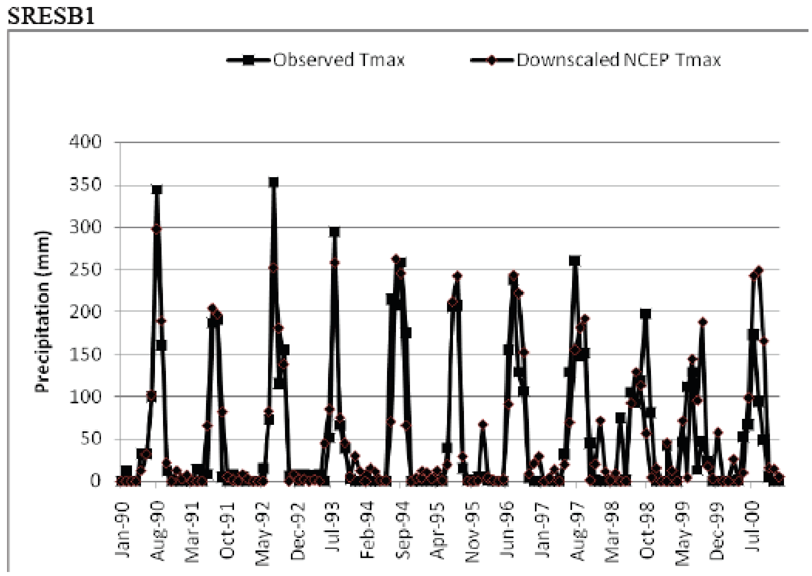

Fig. (12). Typical results for comparison of the amean monthly observed precipitation with precipitation simulated using ANN downscaling model ANNM7 for NCEP data. In the figure, calibration period is from January 1990 to Decmber 1995, and the rest is validation period.

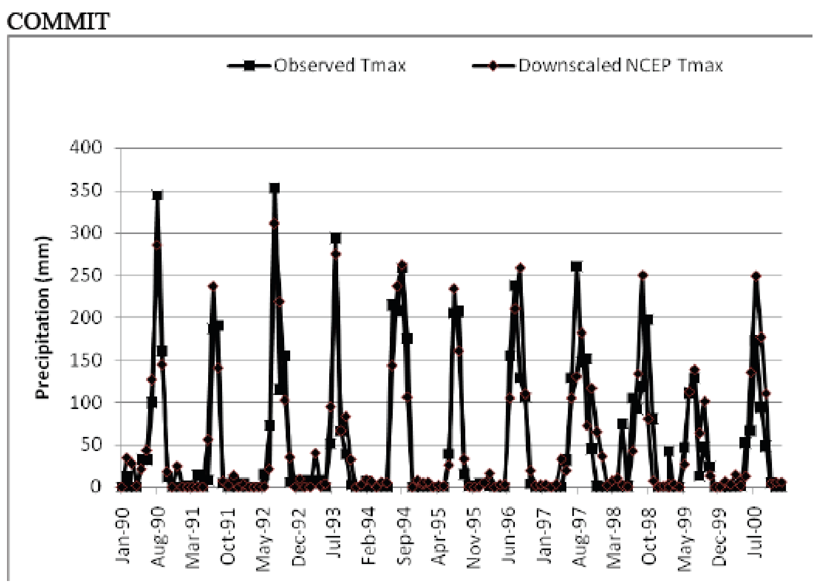

Fig. (13). Typical results for comparison of the amean monthly observed precipitation with precipitation simulated using ANN downscaling model ANNM8 for NCEP data. In the figure, calibration period is from January 1990 to Decmber 1995, and the rest is validation period.

and validation period. Calibration period is from Jaunary 1990 to Decmeber 1995, and the rest is validation period.

7. CONCLUSION

This paper investigates the applicability of the linear multiple regression and neural network for downscaling precipitation from GCM output to local scale. The proposed neural network model is shown to be statistically superior compared to linear multiple regression based downscaling model. The effectiveness of this model is demonstrated through the application of lake catchment in arid region in India. The predictand are downscaled from simulations of CGCM3 for four IPCC scenarios namely SRES A1B, A2, B1 and COMMIT. Scatter plots and cross-correlations are used for studying the reliability of the predictor variables GCM.

The results of downscaling models show that precipitation is projected to increase in future for A2 and A1B scenarios, whereas it is least for B1 and COMMIT scenarios using predictors. These results are in agreement with those obtained for precipitation projections for another river basin in India [12].

APPENDIX I: ABBREVIATIONS

Abbreviations used in text

ANN	=	Artificial Neural Network
CCCma	=	Canadian Center for Climate Modelling 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
MAE	=	Mean absolute error
MSE	=	Mean square error
NMSE	=	Normalized mean square error
PCA	=	Principal component analysis
PC	=	Principal component
RMSE	=	Root mean square error
SRES	=	Special report of emission scenarios
Ta 925	=	Air temperature at 925 mb
Ua 925	=	Zonal wind at 925 mb
Va 925	=	Meridional wind at 925 mb
Ta 500	=	Air temperature at 500 mb
Ua 200	=	Zonal wind at 200 mb
Va 200	=	Meridional wind at 200 mb
Ta 200	=	Air temperature at 200 mb
Zg 200	=	Zeo-potential height at 200 mb
Zg 500	=	Zeo-potential height at 500 mb

Appendix II: Weights and Biases for NN Model ANNM2 Using Back-Propagation Algorithm:

Weights	**h11**	**h12**	**h13**	**h14**	**h15**
i1	-0.2925	0.3092	-0.5335	-0.3635	-0.2925
i2	-0.4455	-0.2372	-0.9856	0.6395	-0.4455
i3	-0.6884	0.4103	0.0524	-1.3407	-0.6884
i4	-0.361	0.4324	-1.4931	-0.2251	-0.361
i5	0.521	-0.476	-0.0041	0.1379	0.521
i6	-0.0193	0.1742	-1.0028	1.3488	-0.0193
i7	-0.3129	-0.6834	0.994	-0.4698	-0.3129
i8	-1.2049	-0.4549	1.5472	0.8928	-1.2049

Biases	b11	b12	b13	b14	b15	b16	b17	b18
	3.2699	2.4241	1.1818	1.1497	0.1561	1.3458	-2.4067	-1.2696

Weights	O1
h21	-1.7789
h22	0.7346
h23	0.893
h24	-0.3502
h25	0.0431
h26	0.2804
h27	-0.6399
h28	-0.9563

Input layer	8nodes
Hidden layer	5 nodes
Output layer	1 node

Biases	bo1
	0.7103

REFERENCES

[1] Grotch SL, MacCracken MC. The use of general circulation models to predict regional climate change. J Clim 1991; 4: 286-303.

[2] Robock A, Turco RP, Harwell MA, *et al*. Use of general circulation model output in the creation of climate change scenarios for impact analysis. Clim Change 1993; 23(4): 293-335.

[3] Sailor DJ, Hu T, Li X, Rosen JN. A neural network approach to local downscaling of GCM output for assessing wind power implications of climate change. Renewable Energy 2000; 19: 359-78.

[4] Wilby RL, Dawson CW, Barrow EM. SDSM – a decision support tool for the assessment of climate change impacts. Environ Modelling Softw 2002; 17: 147-59.

[5] Salathe EP. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rain shadow river basin. Int J Climatol 2003; 23(8): 887-901.

[6] Kim MK, Kang IS, Park CK, Kim KM. Super ensemble prediction of regional precipitation over Korea. Int J Climatol 2004; 24: 777-90.

[7] Wetterhall F, Halldin S, Xu CY. Statistical precipitation downscaling in central Sweden with the analogue method. J Hydrol 2005; 306: 136-74.

[8] Tripathi S, Srinivas VV, Nanjundiah RS. Downscaling of precipitation for climate change scenarios: a support vector machine approach. J Hydrol 2006; 330(3-4): 621-40

[9] Benestad RE. A comparison between two empirical downscaling strategies. Int J Climatol 2001; 21: 1645-68.

[10] Chu JL, Kang H, Tam CY, Park CK, Chen CT. Seasonal forecast for local precipitation over northern Taiwan using statistical downscaling. J Geophys Res 2008; 113:D12118

[11] Paul S, Liu CM, Chen JM, Lin SH. Development of a statistical downscaling model for projecting monthly rainfall over East Asia from a general circulation model output. J Geophys Res 2008; 113: D15117

[12] Anandhi A, Srinivas VV, Nanjundiah RS. Kumar DN. Downscaling precipitation to river basin for IPCC SRES scenarios using Support Vector Machines. Int J Climatol 2008; 28: 401-20

[13] Sen Zekai. Precipitation downscaling in climate modelling using a spatial dependence function. Int J Glob Warming 2009; 1(1-3): 29-42

[14] Daniel JV, David SB, Rosamond LN. Downscaling Indonesian precipitation using large-scale meteorological fields. Int J Climatol 2009; DOI: 10.1002/joc.2010

[15] Mearns LO, Giorgi F, Whetton P, Pabon D, Hulme M, Lal M. Guidelines for use of climate scenarios developed from regional climate model experiments, DDC of IPCC TGCIA Scenarios on economic assessments: An example from U. S. agriculture. Clim Change 2003; 60: 131-48.

[16] Dibike YB, Coulibaly P. Validation of hydrologic models for climate scenario simulation: The case of Saguenay watershed in Quebec. Hydrol Process 2007; 21(23): 3123-235.

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

[18] Carter TR, Parry ML, Nishioka S, Harasawa H, ed. Technical guidelines for assessing climate change impacts and adaptations, report of working group ii of the intergovernmental panel on climate change, University College London and Centre for Global Environmental Research, London and Tsukuba, 1994; 59.

[19] Wilby RL, Wigley TML. Precipitation predictors for downscaling: observed and general circulation model relationships. Int J Climatol 2000; 20(6): 641-61.

[20] Fausett I. Fundamentals of Neural Networks, Prentice Hall, Englewood Cliffs, USA N. Jearsy 1994.

[21] Manish KG, Chandra SPO. Analysis of mean monthly rainfall runoff data of indian catchments using dimensionless variables by neural network. J Environ Prot Sci 2010; doi:10.4236/jep.2010.12020

[22] Hewitson BC, Crane RG, eds. Neural nets applications in geography. Dordrecht: Kluwer Academic Publishers 1994.

[23] Schoof JT, Pryor SC. Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. Int J Climatol 2001; 21: 773-90.

[24] Das BK, Singh M. Water chemistry and control of weathering of Pichola Lake, Udaipur District, Rajasthan, India. Environ Geol 1996; 27: 184-90.

[25] Khobragade SD. Studies on evaporation from open water surfaces in tropical climate, PhD thesis, Indian Institute of Technology, Roorkee, India 2009.

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

[27] Intergovernmental Panel on Climate Change, Climate Change 2001-The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge UK: Cambridge Univ. Press 2001.

[28] Kalnay E, Kanamitsu M, Kistler R, *et al*. The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 1996; 77(3): 437-71.

[29] Willmott CJ, Rowe CM, Philpot WD. Small-scale climate map: a sensitivity analysis of some common assumptions associated with the grid-point interpolation and contouring. Am Cartograph 1985; 12: 5-16.

[30] Crane RG, Hewitson BC. Doubled CO_2 precipitation changes for the Susquehanna Basin: down scaling from the genesis general circulation model. Int J Climatol 1998; 18: 65-76.

[31] Shannon DA, Hewitson BC. Cross-scale relationships regarding local temperature inversions at Cape Town and global climate change implications. S Afr J Sci 1996; 92(4): 213-6.

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

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

[34] Hellstron C, Chen D, Achberger C, Räisasnen J. Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. Clim Res 2001; 19: 45-55.

[35] Cavazos T, Hewitson BC. Performance of NCEP variables in statistical downscaling of daily precipitation. Clim Res 2005; 28: 95-107.

[36] Tolika K, Maheras P, Flocas HA, Papadimitriou AA. An evaluation of a general circulation model(GCM) and the NCEP–NCAR reanalysis data for winter precipitation in Greece. Int J Climatol 2006; 26: 1376-85.

[37] Bardossy A, Bogardi I, Matyasovszky I. Fuzzy rule-based downscaling of precipitation. Theor Appl Climatol 2005; 82: 116-9.

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

[39] Dibike YB, Coulibaly P. Temporal neural networks for downscaling climate variability and extremes. Neural Netw 2006; 19(2): 135-44.

[40] Pearson K. Mathematical contributions to the theory of evolution III regression heredity and panmixia. Philos Trans R Soc Lond Series 1896; 187: 253-318.

[41] Spearman CE. General intelligence' objectively determined and measured. Am J Psychol 1904; 5: 201-93.

[42] Spearman CE. Proof and measurement of association between two things. Am J Psychol 1904; 15: 72-101.

[43] Kendall MG. Regression structure and functional relationship Part I. Biometrika 1951; 38: 11-25.

[44] Heyen H, Zorita E, von Storch H. Statistical downscaling of monthly mean North Atlantic air-pressure to sea level anomalies in the Baltic Sea. Tellus 1996; 48A: 312-23.

[45] Xopalski E, Luterbacher J, Burkard R, Patrikas I, Maheras P. Connection between the large-scale 500 hPa geopotential height fields and precipitation over Greece during wintertime. Clim Res 2000; 14: 129-46.

[46] Cannon AJ, Lord ER. Forecasting summertime surface-level ozone concentrations in the Lower Fraser Valley of British Columbia: An ensemble neural network approach. J Air Waste Manage Assoc 2000; 50: 322-39.

[47] Gardner MW, Dorling SR. Artificial neural networks (the multi layer perceptron)—A review of applications in the atmospheric sciences. Atmos Environ 1998; 32: 2627-36.

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

Received: February 24, 2010 Revised: April 27, 2010 Accepted: May 28, 2010