Chapter 6
Multiple Linear Regression Based Statistical Downscaling of Daily Precipitation in a Canal Command

Prabhash Kumar Mishra, Deepak Khare, Arun Mondal, and Sananda Kundu

Abstract The climate impact studies, particularly in hydrology, often require climate information at fine scale for present as well as future scenario. Global Climate Model (GCM) estimates climate change scenarios on coarse spatial resolution. Therefore, different techniques have been evolved to downscale the coarse-grid scale GCM data to finer scale surface variables of interest. In the present study, the Statistical Downscaling Model (SDSM) has been applied to downscale daily precipitation from simulated GCM data. SDSM utilizes Multiple Linear Regression (MLR) technique. The daily precipitation data (1961–2001) representing Tawa region has been considered as input (predictand) to the model. The model has been calibrated (1961–1991) and validated (1992–2001) with screened large-scale predictors of (National Centre for Environmental Prediction (NCEP) reanalysis data. The prediction of future daily rainfall for the study area has been carried out for the period 2020s, 2050s and 2080s corresponding to HadCM3 A2 variables. The calibration and validation results confirm the SDSM model acceptability slightly at a lower degree. The results of the downscaled daily precipitation for the future period indicate an increasing trend in the mean daily precipitation.

Keywords GCM data • Scenario generation • SDSM • Statistical downscaling • Tawa command

P.K. Mishra (✉)
Department of Water Resources Development and Management,
Indian Institute of Technology Roorkee, Roorkee, India

Scientist, Water Resources Systems Division, National Institute of Hydrology,
Roorkee, India
e-mail: erprabhash@gmail.com

D. Khare • A. Mondal • S. Kundu
Department of Water Resources Development and Management,
Indian Institute of Technology Roorkee, Roorkee, India
e-mail: kharefwt@gmail.com; arun.iirs@gmail.com; sanandakundu@gmail.com

M. Singh et al. (eds.), *Climate Change and Biodiversity: Proceedings of IGU Rohtak Conference, Vol. 1*, Advances in Geographical and Environmental Sciences,
DOI 10.1007/978-4-431-54838-6_6, © Springer Japan 2014

6.1 Introduction

Climate is a complex system, and is very difficult to quantify its' variables. Precipitation is an important parameter (variable) for climate change impact studies. A proper assessment of precipitation for past events, and its' future scenarios is needed for water resources planning. Global Circulation Models (GCMs) are tools available to simulate the ongoing and future changes in climate at global scale. GCMs are numerical models representing the physical processes of the earth-atmosphere-ocean system (Robock et al. 1993; Hewitson and Crane 1996; Wilby and Wigley 1997; Prudhomme et al. 2003; Crawford et al. 2007). These models are of coarse-grid resolution, and of high accuracy at large spatial scales (Bardossy 1997; Ojha et al. 2010; Hassan and Harun 2012). However, impact studies by hydrologists and water resources planner require local/regional-scale hydrological variability to represent local climate phenomena. Hence, different approaches are evolved to downscale the coarse-grid scale GCM data to finer scale surface variables in last few decades. Such methods include canonical correlation analysis, multiple linear regressions, artificial neural networks and support vector machines (Murphy 2000; Lall et al. 2001; Huth 2002; Aksornsingchai and Srinilta 2011; Ghosh and Mujumdar 2006; Raje and Mujumdar 2009; Ghosh 2010; Kannan and Ghosh 2010; Raje and Mujumdar 2011; Hashmi and Shamseldin 2011; Kodra et al. 2012). Recently, downscaling of precipitation has found wide utility for scenario generation on different time scales. SDSM is one of the statistical downscaling tools that implement the multiple linear regression model, and provides scenario of daily surface weather variables under the present and future climate forcing. The tool also performs ancillary tasks of data quality control and transformation, prescreening of predictor variables, model calibration and validation, scenario generation, statistical analysis and its representation of climate data (Wilby and Dawson 2007).

The objective of this study is to understand and utilize Linear Multiple Regression (MLR) technique to downscale mean daily precipitation, both for present and future, for crop planning over a command area corresponding to HadCM3 A2 GCM data utilizing SDSM model. The running paper is subsequently structured as follows: Sect. 6.2 provides a brief description of the study area, followed by discussion on data used in Sects. 6.3 and 6.4 describes the procedure to screen predictor variables for downscaling, and the proposed methodology for development of the regression based model for downscaling precipitation for the command area. Section 6.5 presents the results with discussion, and finally, Sect. 6.6 provides the conclusions drawn from the study.

6.2 Study Area

Tawa command is spread over in an area of about 5,273.12 km^2 falling in the district of Hoshangabad, Madhya Pradesh, India. It lies between 22°54′ N to 23°00′ N latitude and 76°457′ E to 78°45′ E. The area is characterized by a hot summer and

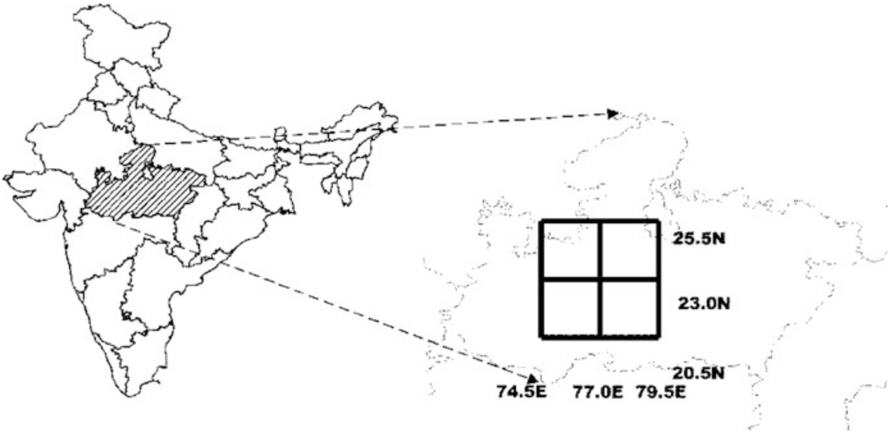

Fig. 6.1 Location map of the Tawa command with NCEP grid (2.5° × 2.5°)

evenly distributed rainfall during the southwest monsoon period. The temperature starts rising from beginning of February and peak is reached in the month of May touching the mercury at 42 °C (Normal). The winter season commences with November and temperature dips to 7.2 °C in the month of December. The relative humidity during summer is low in the month of April i.e. about 18.1 % and is maximum in August i.e., 86.7 %. The annual rainfall varies from 652 mm to 1,898 mm in the command area with average of 1,175 mm based on observations recorded during 1961–2010. The location map of the study area is shown in Fig. 6.1.

6.3 Data Used

6.3.1 Meteorological Data

The daily precipitation data were collected from India Meteorological Department (IMD), Pune for the periods 1961–2001. The daily data were converted to monthly, seasonal and annual time scale before analysis is done.

6.3.2 Reanalysis Data

The daily observed predictor data (re-analysis data) of atmospheric variables, derived from the National Center of Environmental Prediction (NCEP) on 2.5° latitude × 2.5° longitude grid-scale for 41 years (1961–2001) are obtained from the Canadian Climate Impacts Scenarios (CCIS) website (http://www.cics.uvic.ca/scenarios/sdsm/select.cgi).

6.3.3 GCM Data

The large-scale daily predictors of Hadley Center's GCM (HadCM3) for HadCM3 A2 future scenarios for 139 years (1961–2099) on 3.75° latitude × 3.75° longitude grid-scale are obtained from the Canadian Climate Impacts Scenarios (CCIS) website (http://www.cics.uvic.ca/scenarios/sdsm/select.cgi). Among the Special Reports on Emission Scenarios (SRES) A2, being the worst case scenario with high emission projection in future, was considered. HadCM3 is a coupled atmosphere-ocean GCM developed at the Hadley Centre of the United Kingdom's National Meteorological Service. HadCM3 has been chosen because of its' wider acceptance in many climate change impact studies. Further, it provides daily predictor variables, which can be exclusively used for the SDSM model.

6.4 Methodology

The Statistical Downscaling Model (SDSM) is a multiple regression-based tool for generating future scenarios to assess the impact of climate change. It has the ability to capture the inter-annual variability better than other statistical downscaling approaches, e.g. weather generators, weather typing. The model requires two types of daily data, the first type corresponds to local data known as 'Predictand' (Precipitation, temperature) and the second type corresponds to large-scale data of different atmospheric variables known as 'Predictors'(NCEP reanalysis data and simulated GCM based data), for downscaling. Formulating an empirical relationship between predictand and predictor is central to the downscaling technique. This can be achieved by methods, both parametric (Multiple Linear Regression) and non-parametric (Artificial Neural Network; Support Vector Machine). The study has been carried out using SDSM tool version 4.2.9.

6.4.1 Selection of Predictors

For downscaling predictand, the selection of suitable predictors is one of the most important and time consuming steps during downscaling. The appropriate predictor variables are selected through scatter plots, correlation and partial correlation analysis performed between the predictand of interest and predictors. The observed daily NCEP reanalysis data set for the periods 1961–2001 was used to identify the predictors.

6.4.2 Model Calibration and Validation

Model calibration involves development of an empirical relationship, here multiple linear regression, between the predict and of interest and identified daily observed predictors. Part of the NCEP reanalysis data for the period 1961–1991 is used for model calibration, and remaining data between 1992 and 2001 for validation. Validation process enables to produce synthetic daily data based on inputs of the data not considered during model calibration and the formulated regression model. The model performance was evaluated based on the coefficient of correlation (R), defined as:

$$R = \frac{\sum \left(X_{obs} - \overline{X}_{obs}\right)\left(X_{\text{mod}} - \overline{X}_{\text{mod}}\right)}{\sqrt{\sum \left(X_{obs} - \overline{X}_{obs}\right)^2 \sum \left(X_{obs} - \overline{X}_{obs}\right)^2}} \tag{6.1}$$

where,

X_{obs} = Observed value; $\overline{X}_{obs}$ = Mean observed value ; X_{mod} = Modelled value; $\overline{X}_{\text{mod}}$ = Mean modelled value.

6.4.3 Scenario Generation

The validated regression model is applied to generate future scenario for the region utilizing the simulated HadCM3 A2 GCMs data. The study assumes that the relationship between predictor and predictand remains valid under the future climate conditions. Twenty ensembles of daily synthetic precipitation for a period of 139 years (1961–2099) have been generated. The ensemble values are averaged and divided into three separate time period viz. 2020s (2011–2040), 2050s (2041–2070) and 2080s (2071–2099).

6.5 Results and Discussions

6.5.1 Selection of Predictor Variables

The selection of predictor variables is the most significant and time consuming step in statistical downscaling. A list of predictor variables (NCEP and GCM) of a grid-box closest to the Tawa region is presented in Table 6.1. A total of 26 large-scale predictor variables have been considered in the initial screening process. These are categorized into six types based on the atmospheric pressure level. The predictors are selected based on correlation and partial correlation analysis of NCEP predictors and observed weather variables for the period 1961–2003 in SDSM. Variables with higher correlation coefficients between precitand (precipitation) and predictors

Table 6.1 Name and description of all NCEP and GCM predictors

Sl No.	Atmospheric pressure level	NCEP variables	Name	Unit
A.	1013.25 hPa (1)	MSL pressure	mslp	Pa
B.	1000 hPa (6)	Wind speed (Geostrophic)	p_f	m/s
		Zonal (Eastward) velocity (U-component)	p_u	m/s
		Meridional (Northward) velocity (V-component)	p_v	m/s
		Vorticity	p_z	s^{-1}
		Wind direction	p_th	degree
		Divergence	p_zh	s^{-1}
C	850 hPa (8)	Wind speed (Geostrophic)	p8_f	m/s
		Zonal (Eastward) velocity (U-component)	p8_u	m/s
		Meridional (Northward) velocity (V-component)	p8_v	m/s
		Vorticity	p8_z	s^{-1}
		Wind direction	p8_th	degree
		Divergence	p8_zh	s^{-1}
		Geopotential height	p850	m
		Relative humidity	r850	%
D	500 hPa (8)	Wind speed (Geostrophic)	p5_f	m/s
		Zonal (Eastward) velocity (U-component)	p5_u	m/s
		Meridional (Northward) velocity (V-component)	p5_v	m/s
		Vorticity	p5_z	s^{-1}
		Wind direction	p5_th	
		Divergence	p5_zh	s^{-1}
		Geopotential height	p500	m
		Relative humidity	r500	%
E	Near surface (3)	Specific humidity	shum	g/kg
		Mean temperature	temp	°C
		Relative humidity	rhum	%

(NCEP) are chosen for model formulation for scenario generation. The selected predictors with their corresponding correlation coefficients, partial correlation and p value are given in Table 6.2. The scatter plot between selected predictors and observed variable are shown in Fig. 6.2. These statistics help to identify the amount of explanatory power that is unique to each predictor. A 5 % significance level ($p < 0.05$) is used to test the significance of predictor-predictand correlation.

6.5.2 Model Calibration and Validation Results

The model calibration process formulates downscaling model based on multiple regressions between the predictand (observed precipitation) and selected NCEP predictors (Table 6.2). Since the predictand-predictor relationship is governed by wet-day occurrence, an intermediate process in the case of precipitation, a threshold value of 0.3 mm rainfall is considered during model calibration. Calibration (1961–1991)

Table 6.2 Selected NCEP predictors with correlation coefficient, partial correlation and p value

Sl No.	Selected predictors	Correlation coefficients	Partial correlation	P value
1	ncepp_zas	0.374	0.128	0.0001
2	ncepp5_zas	0.320	0.089	0.0001
3	ncepp8_zas	0.344	0.062	0.0006
4	ncep_mslp_as	−0.214	−0.058	0.0013
5	ncepp850as	−0.239	0.056	0.0022
6	ncep_shum_as	0.174	−0.047	0.0105
7	ncep_rhum_as	0.172	0.041	0.0285

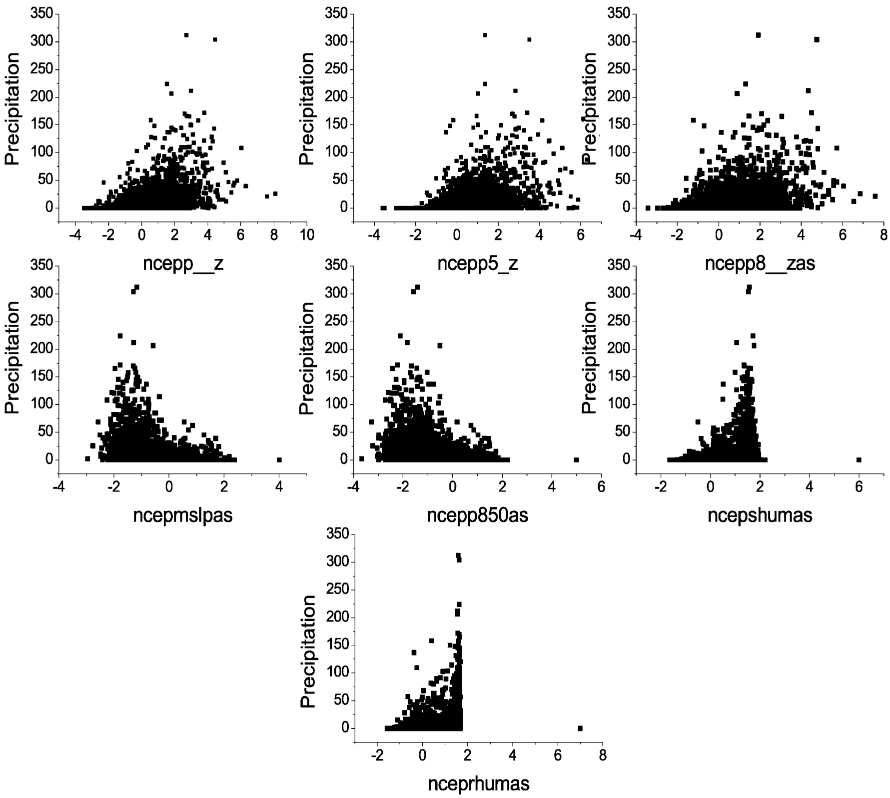

Fig. 6.2 Scatter plots between predictand and selected NCEP predictors

and validation (1992–2001) result of the model downscaling (1961–1991) of daily rainfall is presented in Table 6.3. It can be seen that the SDSM model shows a good agreement between the observed and computed mean daily rainfall, standard deviation and variance with correlation coefficient of 0.57 and 0.50 during calibration and validation respectively. Unlike temperature, the correlation coefficient for the precipitation series is at a lower side. This may be attributed to considerable variation in precipitation with respect to time and space.

Table 6.3 Comparison between daily precipitation (Observed) and daily precipitation (Computed) during model calibration and validation

Type	Period	Mean	SD	Var	Correlation, r
Model calibration	Precp_61-91_Observed	3.02	11.70	136.88	0.57
	Precp_61-91_Computed	3.59	7.24	52.47	
Model validation	Precp_92-01_Observed	3.31	14.46	209.01	0.50
	Precp_92-01_Computed	3.44	6.69	44.81	

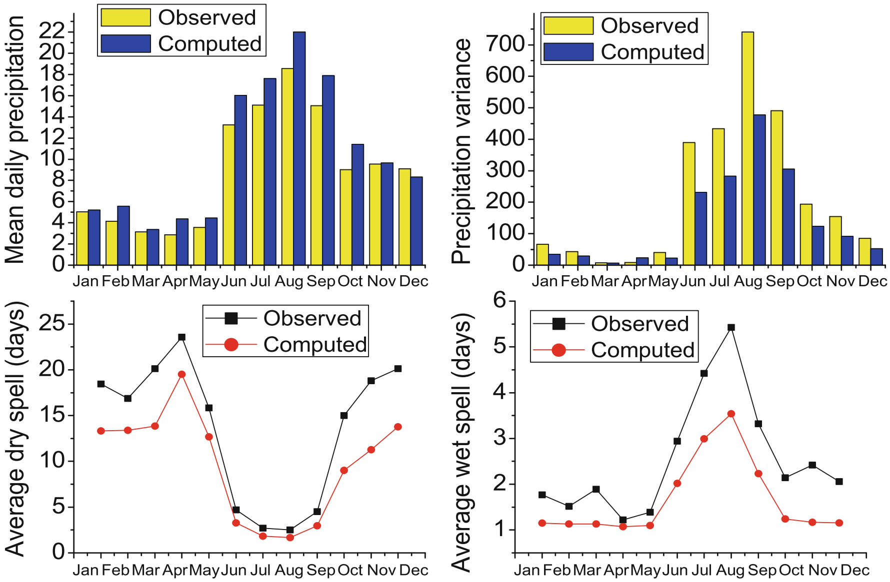

Fig. 6.3 Calibration output of SDSM model downscaling (1961–1991) for daily precipitation

Figure 6.3 highlights the calibration result of the SDSM model with good agreement between observed and computed mean daily precipitation. The computed variance is greater for the monsoon months (June to September). However, there is under-estimation of average dry and wet-spell length. The validation result of the SDSM model for the period 1992–2001 between observed and computed mean daily precipitation, variance, dry-spell length and wet-spell length is shown in Fig. 6.4.

6.5.3 *Future Scenario Generation*

The validated Multiple Linear Regression models between the predictand and large-scale predictors are used to generate the future downscaled data using the HadCM3 GCM data for A2 scenario. The result of the downscaled daily rainfall for different periods is shown in Fig. 6.5. The figure clearly indicates an increasing rainfall trend in the corresponding months for different periods.

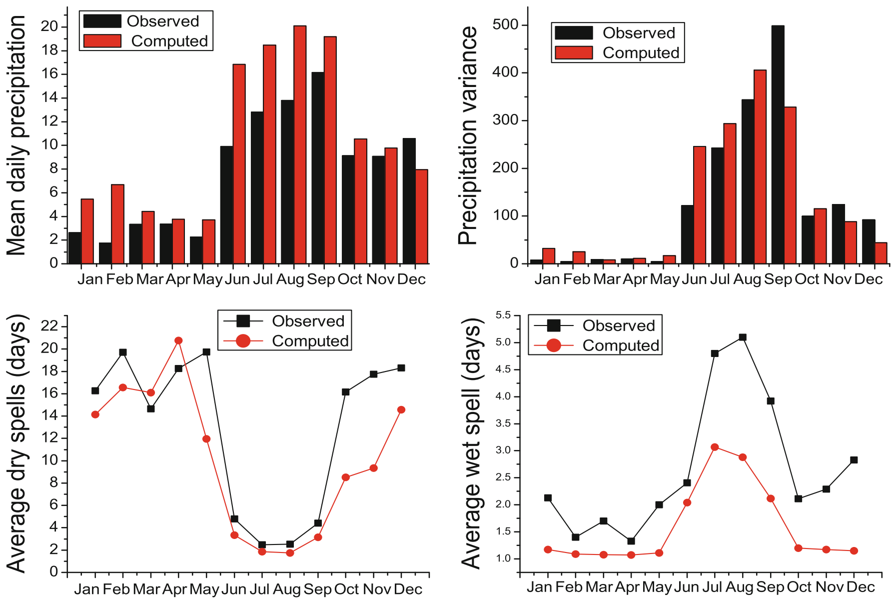

Fig. 6.4 Validation output of SDSM model downscaling (1992–2001) for daily precipitation

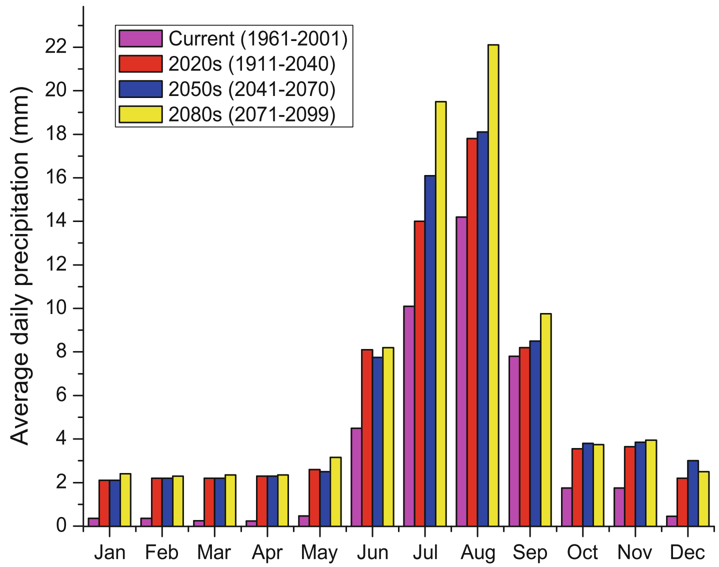

Fig. 6.5 General trend of mean daily precipitation corresponding different scenarios

Table 6.4 Annual average precipitations for present and downscaled precipitation corresponding to HadCM3 A2 scenario

Scenario	Annual average precipitation (mm)
Present	1,129.58
HadCM3 A2 scenario	
2020s	1,327.17
2050s	1,515.92
2080s	1,692.70

The annual precipitation corresponding to future emission is presented in Table 6.4. The result clearly indicates an increase in trend of annual precipitation for successive scenarios. In the 2020s, the simulated annual precipitation is about 200 mm higher than the mean annual precipitation for the present scenario which stands at 1,129 mm. Similarly for 2050s and 2080s, the annual mean precipitations are 1,515.92 and 1,692.70 mm respectively.

6.6 Conclusions

SDSM is one of the downscaling tools widely used to downscale simulated GCM data into local fine-scale data. In the present study, multiple linear regression based SDSM model has been used to downscale daily precipitation data corresponding to HadCM3 A2 GCM (1961–2099). The model calibration and validation has been performed using NCEP reanalysis data for the duration 1961–1991 and 1992–2001 respectively. The calibration and validation results indicate that the model can be used in the Tawa region to downscale climate variables at different temporal and spatial scale. Daily precipitation for the region has been predicted for the study area for the periods 2020s (1911–2040), 2050s (2041–2070) and 2080s (2071–2099). The study indicates an increasing mean daily, monthly and annual precipitation suggesting a wetter climate in the future.

Acknowledgments The authors, thankfully, acknowledges the *Tawa Command Authority* for providing rainfall station data due to which a quality analysis of rainfall in the study have been made possible. Lastly, we acknowledge our thankfulness to State Data Centre, Bhopal, India for supporting during the study. Also the author is thankful to the Ministry of Human Resources Development, GoI for providing financial assistance during the PhD programme.

References

Aksornsingchai P, Srinilta C (2011) Statistical downscaling for rainfall and temperature prediction in Thailand. In: Proceedings of the international multiconference of engineers and computer scientists, vol I. Hong Kong, March 16–18

Bardossy A (1997) Downscaling from GCM to local climate through stochastic linkages. J Environ Manage 49:7–17

Crawford T, Bretts NS, Favis-Mortlock D (2007) GCM grid-box choice and predictor selection associated with statistical downscaling of daily precipitation over Northern Ireland. Clim Res 34:145–160

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

Ghosh S (2010) SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output. J Geophys Res 115, D22102. doi:10.1029/2009JD013548

Hashmi MZ, Shamseldin AY (2011) Comparison of SDSM and LARS-WG for simulation and downscaling of extreme precipitation events in a watershed. Stoch Environ Res Risk Assess 25:475–484

Hassan Z, Harun S (2012) Application of statistical downscaling model for long lead rainfall prediction in Kurau river catchment of Malaysia. Malays J Civ Eng 24(1):1–12

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

Huth R (2002) Statistical downscaling of daily temperature in Central Europe. J Climate 15:1731–1742

Kannan S, Ghosh S (2010) Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output. Stoch Environ Res Risk Assess. doi:10.1007/s00477-010-0415-y

Kodra E, Ghosh S, Ganguly AR (2012) Evaluation of global climate models for Indian monsoon climatology. Environ Res Lett. doi:10.1088/1748-9326/7/1/014012

Lall M, Nozawa T, Emori S, Harasawa H, Takahashi K, Kimoto M, Abe-Ouchi A, Nakajima T, Takemura T, Numaguti A (2001) Future climate change: implications for Indian summer monsoon and its variability. Curr Sci 81(9):1196–1207

Murphy J (2000) Predictions of climate change over Europe using statistical and dynamical downscaling techniques. Int J Climatol 20:489–501

Ojha CSP, Goyal MK, Adeloye AJ (2010) Downscaling of precipitation for lake catchment in arid region in India using multiple linear regression and neural networks. Open Hydrol J 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

Raje D, Mujumdar PP (2009) A conditional random field-based downscaling method for assessment of climate change impact on multisite daily precipitation in the Mahanadi basin. Water Resour Res 45:1–20

Raje D, Mujumdar PP (2011) A comparison of three methods for downscaling daily precipitation in the Punjab region. Hydrol Process 25:3575–3589

Robock A, Turco RP, Harwell MA, Ackerman TP, Andressen R, Chang HS, Sivakumar MVK (1993) Use of general circulation model output in the creation of climate change scenarios for impact analysis. Clim Change 23:293–335

Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. Progr Phys Geogr 21:530–548

Wilby RL, Dawson CW (2007) SDSM 4.2 - a decision support tool for the assessment of regional climate change impacts. UK