

Rainfall estimation using multiple linear regression based statistical downscaling for Piperiya watershed in Chhattisgarh

S. K. CHANDNIHA* and M. L. KANSAL

Department of Water Resources Development and Management, Indian Institute of Technology, Roorkee, 247667, Uttarakhand, India

**E-mail : chandniha.surendra@gmail.com*

ABSTRACT

Climatic variability and its behavior is a complex phenomenon that is directly associated with uncertainties. In the climate change study, particularly in hydrological aspects, it is necessary to identify the parameters (predictors) that are directly or indirectly associated with predictands. The forecasted results are directly associated with the selection of predictors. In the present study, the statistical downscaling model (SDSM) has been advocated to downscale the daily rainfall in Piperiya watershed of Chhattisgarh state. SDSM is based on multiple linear regression (MLR) technique. The daily rainfall data (1961-2001) of the Piperiya watershed in Chhattisgarh is considered as input (predictand) to the model. The model has been calibrated and validated on the basis of rainfall period of 1961-1990 and 1991-2001 respectively with large scale predictors of National Centre for Environmental Prediction (NCEP) reanalysis data. Finally, monthly rainfall is predicted on the basis of forecasted future daily rainfall for the periods of 2020s, 2050s and 2080s under the consideration of HadCM3 A2 and B2 emission scenarios.

Keywords: Climate change, HadCM3, SDSM, statistical downscaling, Piperiya watershed

Climate change is a long drawn process which is contributed by various factors that affect directly or indirectly the various hydrological processes such as runoff. Climate change and its impact on water resources is difficult to assess but is definitely felt at basin and regional scale. Global Circulation Models (GCMs) reflect the climatic conditions at global as well as local levels and are based on concentration of greenhouse gases (Mitchell *et al.*, 1995). GCMs provide the future forecasting of meteorological parameters under different climatic scenario conditions. In case of precipitation, GCMs have limited ability to forecast good result for daily future time series (Mishra *et al.*, 2014). Downscaling is the technique, which provide the fine scale numerical values from coarse resolution. Statistical downscaling is a regression based downscaling which works on empirical relationship between the local scale predictands and regional scale predictor(s) (Bardossy, 1994; Bárdossy, 1997; Boroneant *et al.*, 2006; Goyal and Ojha, 2012; Goyal and Ojha, 2010). Individual downscaling schemes differ depending on the choice of mathematical transfer function, predictor variables and statistical fitting procedure. The selection of predictors indicate its behavior, dependency, correlation and the principle components (Wilby and Wigley, 2000). Further, regression models assume steady-state parameters under future climate conditions. The future climatic scenarios are

highly sensitive to the choice of predictor variables and statistical transfer function (Horton *et al.*, 2006). It is noticed that the regression-based statistical downscaling model (SDSM) is an appropriate downscaling model which resolve the coarse spatial resolution to fine GCM resolution (Dibike and Coulibaly, 2005; Khan *et al.*, 2006; Samadi *et al.*, 2011; Wilby *et al.*, 2002). In such statistical downscaling, a statistical empirical relationship between atmospheric variables (predictors) such as specific humidity, temperature, geo-potential height, etc. and individual local scale meteorological variables (predictands) such as precipitation and temperature is framed (Wetterhall *et al.*, 2006; Wilby *et al.*, 2002). Keeping this in mind, the present study uses the statistical downscaling technique which is based on multiple linear regression (MLR). Out of the 26 possible predictors, seven most appropriate parameters have been identified as the contributing parameters for the Piperiya watershed in Chhattisgarh. For calibration purpose, 30 year rainfall data (1961-1990) has been used while 1991-2001 data were considered for validation. Further, prediction of future rainfall for the study area has been evaluated for the period 2020s, 2050s and 2080s on the basis of HadCM3 A2 and B2 variables. The results are presented on monthly and annual basis.

MATERIAL AND METHODS

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. This approach involves three sub-classes such as weather typing, weather generator and regression/transform function. The model requires two types of daily data, i.e., (i) the local data known as 'Predictand' (rainfall) and (ii) the different atmospheric variables known as 'Predictors'. Formulating an empirical relationship between predictand and predictor is central to the downscaling technique. This can be derived by various methods such as parametric (multiple linear regression) and non-parametric (artificial neural network; support vector machine). This study uses the Multiple Linear Regression method which falls under parametric methods. The downscaling has been carried out using SDSM tool version 4.2.9.

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, positive and negative correlation and partial correlation analysis between predictand (rainfall) and predictors (most appropriate out of 26 possible parameters). The observed daily NCEP reanalysis data set for the periods 1961-2001 was used for the selection of predictors.

Model calibration and validation

Model calibration is carried out to development of an empirical relationship between the predictand and the predictors using multiple linear regression. NCEP reanalysis data for the period 1961-1990 was used for model calibration, and rest of the data from 1991 to 2001 is used for validation purpose. Validation process enables to produce synthetic daily data based on inputs of the data considered during the model calibration. The model performance was evaluated based on the coefficient of correlation (R) between the observed values during the validation period and the modeled values, which can be calculated as

$$R = \frac{\sum (X_{obs} - \bar{X}_{obs})(X_{mod} - \bar{X}_{mod})}{\sqrt{\sum (X_{obs} - \bar{X}_{obs})^2 \sum (X_{mod} - \bar{X}_{mod})^2}}$$

Where,

X_{obs} = observed value; \bar{X}_{obs} = mean observed value;
 X_{mod} = Modeled value; \bar{X}_{mod} = Mean modeled value

Scenario generation

The validated regression model is applied to generate future scenario for the watershed utilizing the simulated HadCM3 A2 and B2 GCMs data. The study assumes that the relationship between predictor and predictand remains valid under the future climate conditions. Twenty ensembles of daily synthetic rainfall for a period of 139 years (1961-2099) have been generated. The ensemble values are averaged and divided into four separate time period viz. past (1961-2010), 2020s (2011-2040), 2050s (2041-2070) and 2080s (2071-2099).

Study area

The Piperiya watershed is located in Hasdeo river catchment of Mahanadi basin in Chhattisgarh state of India. It lies between Northern latitude of 22°37'46" to 23°35'40" and Eastern longitude of 82°01'48" to 82°37'29". Area of the watershed is about 2414 km². The watershed covers the three districts (Koriya, Korba and Bilaspur) of Chhattisgarh and partially intersects the Annappur district of Madhya Pradesh. However the major part of the watershed is covered under Koriya district of Chhattisgarh. The topography is hilly in the northern part and it has plains in south. Elevation of the watershed varies from 324 to 1062m with Northern part at higher elevation. The Koriya district receives about 1212 mm rainfall in a year. In summer, temperature reaches up to 47°C, however the average temperature varies between 17-32°C. The drainage outlet of the watershed is in Manendragarh block of the Koriya district. About 50 percent of the watershed area is covered with forest and agricultural land.

Data used

Meteorological data : The daily rainfall data (1°x1° grid) was collected from India Meteorological Department (IMD), Pune for the periods 1961-2001. Statistical downscaling has been performed using the daily rainfall time series as input predictand in SDSM software. Further, daily observed and estimated time series converted into monthly and annual time series.

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 x 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>).

Table 1: Name and description of all NCEP and GCM predictors

S. No.	Atmospheric pressure level	NCEP Variables Descriptions	Code	Unit
A	1013.25 hPa (1)	Mean sea level pressure	ncepmslpas	Pa
B	1000 hPa (6)	Surface airflow strength	ncepp_fas	m/s ⁻¹
		Surface zonal velocity	ncepp_uas	m/s ⁻¹
		Surface meridional velocity	ncepp_vas	m/s ⁻¹
		Surface vorticity	ncepp_zas	s ⁻¹
		Surface wind direction	ncepp_thas	degree
		Surface divergence	ncepp_zhas	s ⁻¹
C	850 hPa (8)	850 hPa airflow strength	ncepp8_fas	m/s ⁻¹
		850 hPa zonal velocity	ncepp8_uas	m/s ⁻¹
		850 hPa meridional velocity	ncepp8_vas	m/s ⁻¹
		850 hPa vorticity	ncepp8_zas	s ⁻¹
		850 hPa wind direction	ncepp8thas	degree
		850 hPa divergence	ncepp8zhas	s ⁻¹
		850 hPa geopotential height	ncepp850as	m
		Relative humidity at 850 hPa	ncepr850as	%
D	500 hPa (8)	500 hPa airflow strength	ncepp5_fas	m/s ⁻¹
		500 hPa zonal velocity	ncepp5_uas	m/s ⁻¹
		500 hPa meridional velocity	ncepp5_vas	m/s ⁻¹
		500 hPa vorticity	ncepp5_zas	s ⁻¹
		500 hPa wind direction	ncepp5thas	
		500 hPa divergence	ncepp5zhas	s ⁻¹
		500 hPa geopotential height	ncepp500as	m
		Relative humidity at 500 hPa	ncepr500as	%
E	Near surface (3)	Surface specific humidity	ncepshumas	g kg ⁻¹
		Mean temperature at 2m	nceptempas	0C
		Near surface relative humidity	nceprhumas	%

GCM data : The large-scale daily predictors of Hadley Center's GCM (HadCM3) for HadCM3 A2 and B2 future scenarios for 139 years (1961-2099) on 3.75° latitude x 3.75° longitude grid-scale are also obtained from the Canadian Climate Impacts Scenarios (CCIS) website (<http://www.cics.uvic.ca/scenarios/sdsm/select.cgi>). 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 in India. Further, it provides daily predictor variables, which can be exclusively used for the SDSM model.

RESULTS AND DISCUSSION

Selection of predictor variables

A list of predictor variables (NCEP and GCM) of a grid-box closest to the Piperiya watershed in Koriya district of Chhattisgarh is presented in Table 1. A total of 26 large-scale predictor variables have been considered in the initial screening process. These are categorized into five types based on the atmospheric pressure levels. The predictors are

selected based on the correlation and the partial correlation analysis of NCEP predictors and observed weather variables for the period 1961-2001 in SDSM model. Variables with best correlation coefficients between predictand and predictors were chosen for model formulation for scenario generation.

The selection of predictors have been carried out using correlation coefficients, partial correlation and p-values between predictand and NCEP predictors. For example, in this case, mean sea level pressure, surface zonal velocity, 500 hPa zonal velocity, 500 hPa vorticity, 500 hPa geopotential height and surface specific humidity are identified as the best suited predictors for this case study. The corresponding correlation coefficients, partial correlation and p-values between the predictand and NCEP predictors are shown in Table 2.

The scatter plot between selected predictors and predictand are shown in Fig. 1. A 5% significance level ($p < 0.05$) is used to test the significance of predictor-predictand correlation

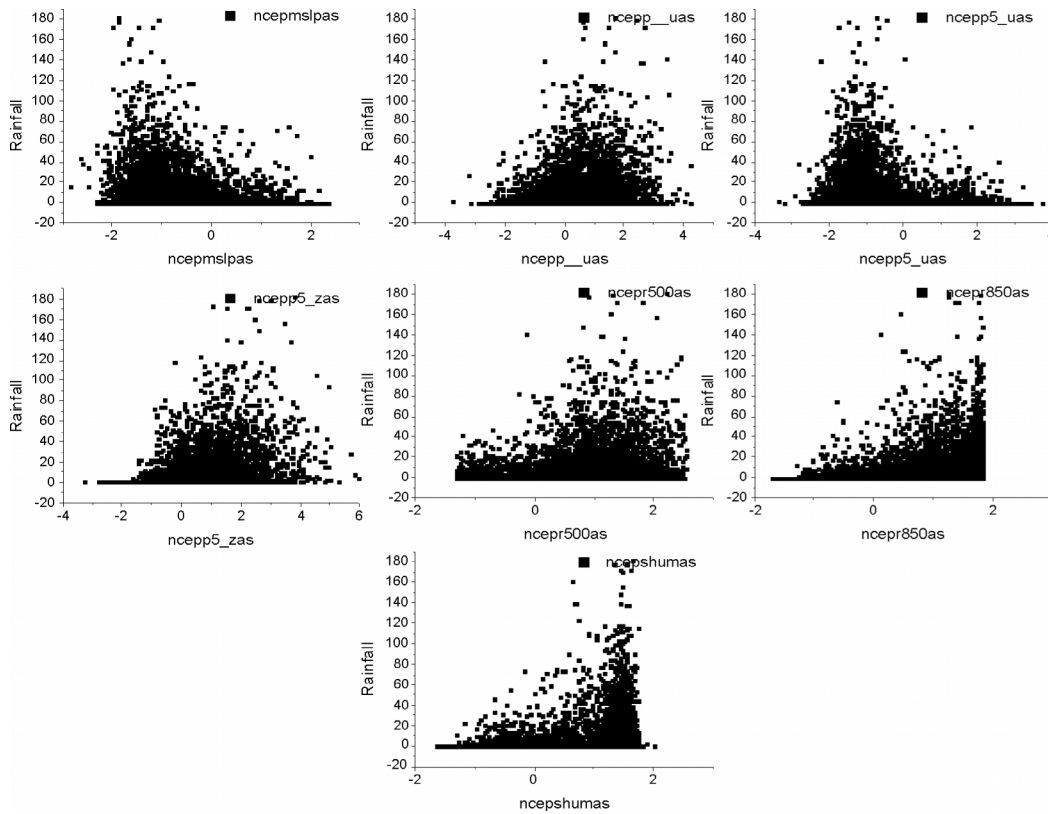


Fig. 1: Scatter plots between predictand and selected NCEP predictors

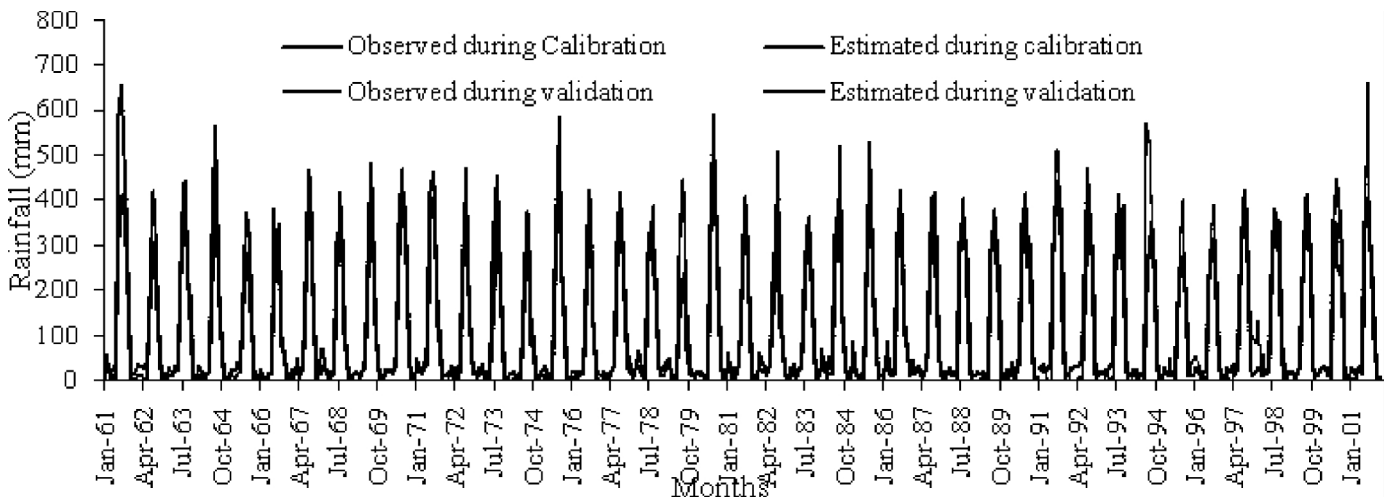


Fig. 2: Observed and estimated monthly rainfall during calibration (1961-1990) and validation (1991-2001)

Table 2 : Selected NCEP predictors and their relationship with rainfall

SI No.	Selected predictors	Correlation coefficients	Partial correlation	P value
1	ncepmslpas	-0.327	-0.070	0.0001
2	ncepp_uas	0.256	0.079	0.0001
3	ncepp5_uas	-0.286	-0.023	0.1987
4	ncepp5_zas	0.384	0.187	0.0001
5	ncepr500as	0.316	0.053	0.0020
6	ncepr850as	0.367	0.122	0.0001
7	ncepshumas	0.361	-0.065	0.0001

Table 3: Calibration and validation of monthly average rainfall with NCEP reanalysis data

Month	Calibration		Validation	
	Observed	Estimated	Observed	Estimated
January	20.0	23.3	25.6	23.9
February	22.9	25.6	9.0	23.5
March	12.5	14.6	17.9	16.6
April	14.0	14.3	11.9	19.1
May	24.2	27.4	12.5	31.9
June	215.1	211.2	198.7	223.5
July	389.0	388.1	394.3	383.6
August	377.3	374.3	360.3	369.9
September	221.0	222.4	217.7	232.1
October	55.8	62.0	45.5	56.2
November	7.5	10.6	19.3	9.5
December	8.6	11.5	13.9	11.1
Minimum	78.5	976.0	939.0	1015.0
Maximum	2471.0	1688.0	1828.0	1781.0
Average	168.0	1385.0	1326.0	1401.0
SD	296.0	147.0	239.0	2250.
R ²	0.8436		07376	

Table 4: Detailed rainfall statistics for different time steps (scenarios)

Month	A2-Scenario				B2-Scenario			
	Past	2020s	2050s	2080s	Past	2020s	2050s	2080s
January	21.5	30.9	32.3	33.6	21.5	37.5	38.2	43.3
February	19.2	34.4	35.6	34.9	19.2	39.3	43.4	42.7
March	13.9	20.3	22.0	23.7	13.9	19.3	20.7	19.3
April	13.4	24.1	23.3	27.4	13.4	21.5	23.0	22.7
May	21.1	47.7	44.2	51.7	21.1	31.0	34.5	36.6
June	210.7	218.8	232.1	258.2	210.7	224.5	240.2	269.8
July	390.4	339.1	365.4	388.5	390.4	325.4	348.5	378.4
August	372.7	330.8	381.7	465.3	372.7	348.2	400	459.5
September	220.1	262.3	311.4	370.9	220.1	242.7	259.3	310.7
October	53.1	89.9	95.1	91.4	53.1	91.7	96.1	93.1
November	10.6	18.8	19.3	20.0	10.6	21.4	23.3	22.8
December	10.0	23.4	23.9	24.8	10.0	23.6	25.2	26.7
Minimum	785.0	1240.0	1168.0	1536.0	785.0	1226.0	1286.0	1507.0
Maximum	2470.0	1625.0	1769.0	2024.0	2470.0	1618.0	1786.0	2135.0
Average	1357.0	1440.0	1586.0	1790.0	1357.0	1426.0	1552.0	1726.0
SD	114.0	103.0	130.0	127.0	114.0	117.0	135.0	153.0

It may be noted that since the downscaling is carried out on daily basis, the correlation coefficients, partial correlation and p-values are low. However, the purpose of the study is to get the rainfall values on monthly basis, and hence the daily downscaled data is aggregated to present the monthly rainfall values.

Model calibration and validation results

Since the predictand-predictor relationship is governed by wet-day occurrences, a threshold value of 3 mm rainfall is considered during model calibration. Results of the observed and estimated monthly rainfall during calibration and validation period are shown in Table 3. Monthly average and annual statistics of observed and estimated rainfall are also presented in Table 3.

It can be seen that, the SDSM model shows a good agreement between the observed and estimated monthly average and annual (minimum, maximum, average and standard deviation) statistics of rainfall during calibration and validation period. Model efficiency have been estimated with correlation coefficients between observed and estimated in both the cases. However, the values of coefficient of determination during calibration and validation were estimated as 0.8436 and 0.7376 respectively. Observed and estimated monthly rainfall during calibration (1961-1990) and validation (1991-2001) shown in Fig. 2.

It may be noticed that during most of the time period, the predicted and observed values are matching except that of few outliers. In other, words, it can be concluded that the performance of SDSM model using MLR is good on monthly basis for future estimation of rainfall under HadCM3-A2 and B2 emission scenarios.

Future scenario generation

Using the MLR based SDSM model, the future rainfall under HadCM3-A2 and B2 emission scenarios are estimated with the identified predictors. The estimated rainfall is represented into four different time steps, i.e., past (1961-2010), 2020s (2011-2040), 2050s (2041-2070) and 2080s (2071-2099) and the detailed statistics are represented in Table 4.

The monthly rainfall predicted for the study area during the periods 2020s, 2050s and 2080s, indicates an increasing trend. In A2 scenario, annual rainfall varies from 785 to 2470mm in past, 1240 to 1625mm in 2020s, 1168 to 1769mm in 2050s and 1536 to 2024mm in 2080s. Similarly, under B2 scenario, the annual rainfall varies from 785 to 2470mm in past, 1226 to 1618mm in 2020s, 1286 to 1786mm

in 2050s and 1507 to 2135mm in 2080s. Under both A2 and B2 scenarios, there is an increasing trend of rainfall in this watershed. It also has been observed that the future rainfall under the A2 scenario is always higher than under the B2 scenario.

CONCLUSIONS

This study highlights the use of MLR based SDSM technique for assessing the likely future monthly rainfall in Piperiya watershed of Chhattisgarh state in India. The tool is widely used for hydro-meteorological downscaling of GCMs to local fine-scale resolution. In the present study, the daily rainfall time series corresponding to HadCM3 A2 and B2 emission scenario are generated and then used for estimating the monthly rainfall for different future time periods. The model calibration and validation has been accomplished using NCEP reanalysis data for the periods 1961-90 and 1991-2001, respectively. It is hoped that this study will help in the effective management of water resources in the state in general and Piperiya watershed in particular. Also, it will help in study the effect of climate change on the expected rainfall in this particular area.

ACKNOWLEDGMENTS

The authors, thankfully, acknowledge the Indian Meteorological Department, Pune, India for providing rainfall data for the area. Also, the first author is thankful to the Ministry of Human Resources Development, Govt. of India for providing financial assistance for carrying out the study as part of the PhD programme.

REFERENCES

- Bardossy, A. (1994). Estimation of extreme regional precipitation under climate change. In "Engineering Risk in Natural Resources Management", pp. 195-205. Springer.
- Bárdossy, A. (1997). Downscaling from GCMs to local climate through stochastic linkages. *J. Environ. Manage.* 49: 7-17.
- Boroneant, C., Plaut, G., Giorgi, F., and Bi, X. (2006). Extreme precipitation over the Maritime Alps and associated weather regimes simulated by a regional climate model: Present-day and future climate scenarios. *Theor. & Appl. Climatol.* 86: 81-99.
- Dibike, Y. B., and Coulibaly, P. (2005). Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. *J. hydrol.*, 307: 145-163.

- Goyal, M. K., and Ojha, C. (2012). Downscaling of precipitation on a lake basin: evaluation of rule and decision tree induction algorithms. *Hydrol. Res.*, 43: 215-230.
- Goyal, M. K., and Ojha, C. S. P. (2010). Evaluation of various linear regression methods for downscaling of mean monthly precipitation in arid Pichola watershed. *Natural. Resour.*, 1: 11.
- Horton, P., Schaeffli, B., Mezghani, A., Hingray, B., and Musy, A. (2006). Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty. *Hydrol. Processes*, 20: 2091-2109.
- Jiehua, M., Huijun, W., and Ke, F. (2014). Dynamic Downscaling of Summer Precipitation Prediction over China in 1998 using WRF and CCSM4. *Adv. in Atmos. Sci.*, 32(5): 577-584.
- Khan, M. S., Coulibaly, P., and Dibike, Y. (2006). Uncertainty analysis of statistical downscaling methods. *J. Hydrol.*, 319: 357-382.
- Mishra, P. K., Khare, D., Mondal, A., and Kundu, S. (2014). Multiple Linear Regression Based Statistical Downscaling of Daily Precipitation in a Canal Command. In "Climate Change and Biodiversity", pp. 73-83. Springer.
- Mitchell, J., Johns, T., Gregory, J., and Tett, S. (1995). "Climate response to increasing levels of greenhouse gases and sulphate aerosols."
- Samadi, S., Ehteramian, K., and Sarraf, B. S. (2011). SDSM ability in simulate predictors for climate detecting over Khorasan province. *Procedia-Social and Behavioral Sciences*, 19, 741-749.
- Wetterhall, F., Bárdossy, A., Chen, D., Halldin, S., and Xu, C. Y. (2006). Daily precipitation-downscaling techniques in three Chinese regions. *Water Resour. Res.* 42 (11)
- Wilby, R. L., Dawson, C. W., and Barrow, E. M. (2002). SDSM—a decision support tool for the assessment of regional climate change impacts. *Environ. Modelling & Soft.*, 17: 145-157.
- Wilby, R. L., and Wigley, T. (2000). Precipitation predictors for downscaling: observed and general circulation model relationships. *Int. J. Climatol.*, 20: 641-661.