# Rainfall Projection in Yamuna River Basin, India, Using Statistical Downscaling

**Siddharth Chaudhary, A. Agarwal and Tai Nakamura**

**Abstract** This paper exhibits a strategy to build up a downscaling model using the artificial neural network (ANN) to obtain a projection of monthly mean precipitation (MMP) at river basin scale. Moreover, we establish an association between local climate variables and large-scale variables known as atmospheric circulation variables using statistical downscaling. National Centers for Environmental Prediction (NCEP) reanalysis data was applied on the model for calibration and validation for the time frame of January 1971 to December 2005. Similarly, RCP 4.5 scenario of CanCM4 was used for future projection till December 2035. The model developed shows the value of the coefficient of determination as 0.988 for calibration and for validation 0.883. Therefore, we accept the model and it is used for projection and forecasting of precipitation till 2035 on the Yamuna river basin.

**Keywords** Statistical downscaling · Artificial neural network
Global circulation model · Climate change

## 1 Introduction

In recent past, many studies have been conducted on finding an appropriate downscaling technique for projection of monthly precipitation at river basin scale [1–3]. The projected rainfall at basin scale is used for many scientific purposes such as climate change impact assessment [4], drought analysis [3] and flood frequency analysis [5]. In the climate change studies, global circulation models (GCMs) is generally

S. Chaudhary (✉) · T. Nakamura
Remote Sensing and GIS, Asian Institute of Technology, Klong Luang, Thailand
e-mail: chaudharysiddharth@outlook.com

A. Agarwal
Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany

A. Agarwal
Research Domain Transdisciplinary Concepts & Methods, Potsdam Institute for Climate Impact Research, Telegrafenberg, Potsdam, Germany



M. Rathinasamy et al. (eds.), *Water Resources and Environmental Engineering II*,
https://doi.org/10.1007/978-981-13-2038-5_2

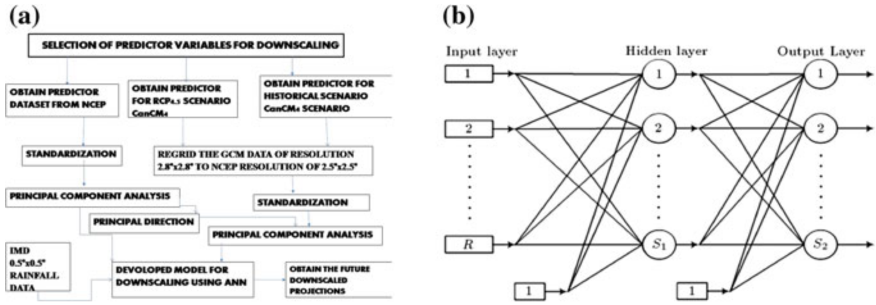


**Fig. 1** **a** Methodology for ANN downscaling; **b** MLP model structure

used because it simulates past, present and future time series of all the climate variables across the globe. The results simulated by the GCM are generated, keeping all the atmospheric forcing both internal and external, various scenarios for greenhouse gases and also the social and economical changes. The resolution of GCM data is at a coarser resolution, and hence, we cannot use it for studying the climate change impact assessment on hydrological regime of a river basin. Downscaling, a statistical and dynamical approach, is used to draw high-resolution information from coarser resolution variables [6]. In dynamic downscaling, we are able to embed the regional climate model (RCM) into GCM. This technique makes use of the numerical meteorological model to show that how local weather conditions are affected due to change in global pattern. There are many advantages of dynamic downscaling but due to high computational time and its complexity, it is not a very popular method. Whereas statistical downscaling is a technique in which an empirical relationship is established between the predictors and predictands. Predictors are the atmospheric variables which are generated by the GCM, and predictands are regional climate data. Predictors like mean sea level pressure, specific humidity, geopotential height, air temperature are related to the point scale variables called predictands like rain gauge stations as shown in Fig. 1. The relationships established are finally used for forecasting the information regarding climate for the future period. Wilby et al. [7] give some implicit assumptions in statistical downscaling. These assumptions are listed below:

1. The predictors are variables of relevance and are realistically modelled by GCM.
2. The predictors employed fully represent the climate change signal.
3. The relationship is valid under altered conditions.

## 2 Methodology

The model used for projecting rainfall over the Yamuna river basin in this study is artificial neural network (ANN). The model used is a non-regression model, and it helps in establishing a relationship between the predictors and predictands. ANN model works similar to the biological neural network system.

ANN model can be trained using backpropagation method. It is generally used when we need to do supervised learning. This method is more efficient in finding the gradient of the error function. The methodology of downscaling using the ANN is shown in Fig. 1a.

## 2.1 Multilayer Perception Network

MLP is also known as a feedforward network. It is based on the mapping input pattern space to output space. In this method, the information is transferred with the help of the connections between the neurones in the forward direction only. This kind of network does not have any feedback or memory. Figure 1b shows three layer MLP.

## 2.2 Standardisation

Once the potential predictors are selected, they are needed to be standardised. The main purpose of standardisation is to reduce the systematic biases in the variance and mean of GCM outputs in correspondence to the observed or NCEP data. Standardisation is applied before statistical downscaling is applied. Standardisation is applied on predictor variable for the baseline period by subtracting each data by its mean and dividing it by standard deviation. This method has a major limitation due to its consideration of biases only in mean and variances and not in other statistical parameters.

## 2.3 Principal Component Analysis

For the development of the empirical or statistical relationship, we cannot use the data directly. There are mainly two reasons behind this: (i) multidimensionality and (ii) multicollinearity.

Computational time required for the analysis is increased due to multidimensionality. It is observed that there is an increase in the sparseness of the data and this affects the output. It is really important to know the internal data pattern and its variability. If the data is reduced without considering the above two, then the accuracy of the model will reduce. Multicollinearity shows a high correlation between various predictors and may lead to solving these problems before we establish any statistical relationship.

PCA is a very important tool used in statistical downscaling studies [8]. It is used to for reducing the dimensions and multicollinearity of the data. PCA generates principal components which have almost similar variability as compared to the original data by analysing and identifying the patterns in the data.

## 3 Study Area and Data Used

The study has been conducted for the Yamuna river basin which is a sub-basin of Ganga river basin and is located in north-western part of India as shown in Fig. 2a. Yamuna basin lies between 22.50°N to 32.00°N latitude and 73.20°E to 81.50°E longitude. The total area covered by the basin is 3,66,233 km$^2$ which is approximately 10.7% of Ganga basin. There is great variation in the elevation of land surface ranging from 62 to 6288 m above mean sea level. The main source of Yamuna river lies in the Yamunotri glacier (Mussorie range) in the Uttarakhand, India. The following states fall in the boundary of the Yamuna basin: Himachal, Haryana, Rajasthan, Madhya Pradesh, Uttar Pradesh and Uttarakhand. The percentage of the area covered by the states is shown in Fig. 2b, and the Yamuna basin is shown in Fig. 2c.

Annual rainfall in the Yamuna basin shows variations ranging from 400 to 1200 mm. Moreover, the basin is highly influenced by the event of south-west monsoon during months of June and September due to which the basin gets replenished. Winter rainfall in this region is very less. The rainfall distribution shows an increasing nature from north-west to south-east direction. The climatic condition of the basin can be broadly classified into three categories: (a) Humid: for upstream Himalayan catchment; (b) Semi-Arid: north-west to western catchment; and (c) Sub-Humid: south-west catchments. The mean maximum temperature of the basin is 24 to 42.4 °C and the mean minimum temperature is −1 to 11.0 °C.

## 4 Data Used

The global circulation model (GCM) used in the study is developed by Canadian centre for climate modelling and analysis (CanCM4) [9]. The spatial resolution of GCM used is 2.8125° × 2.8125°. Data for historical scenarios are extracted for the period of January 1971–December 2005, and for RCP 4.5, it is extracted for January 2016–December 2035. The GCM is represented in the form of 20 grid points and is shown in Fig. 3. The list of predictor variables used for downscaling precipitation data over Yamuna river basin is shown in Table 1. The data is available freely and can be downloaded from www.ccma.bc.ec.gc.ca.

Predictors used in the study were selected on the basis of the study conducted by Anandhi et al. [1]. A dataset of the predictors for the period of January 1971 to December 2004 over the area from 22.50°N to 32.00°N and from 73.20°E to 81.50°E is obtained from NCEP. The NCEP data is represented in the form of 12 grid points and is in monthly scale. The resolution of the NCEP data is 2.5° × 2.5°. The precipitation data is in gridded format with a resolution of 0.5° × 0.5° and it is taken from Indian Metrological Department (IMD). For the study purpose, we regrid the GCM data points to NCEP data points.

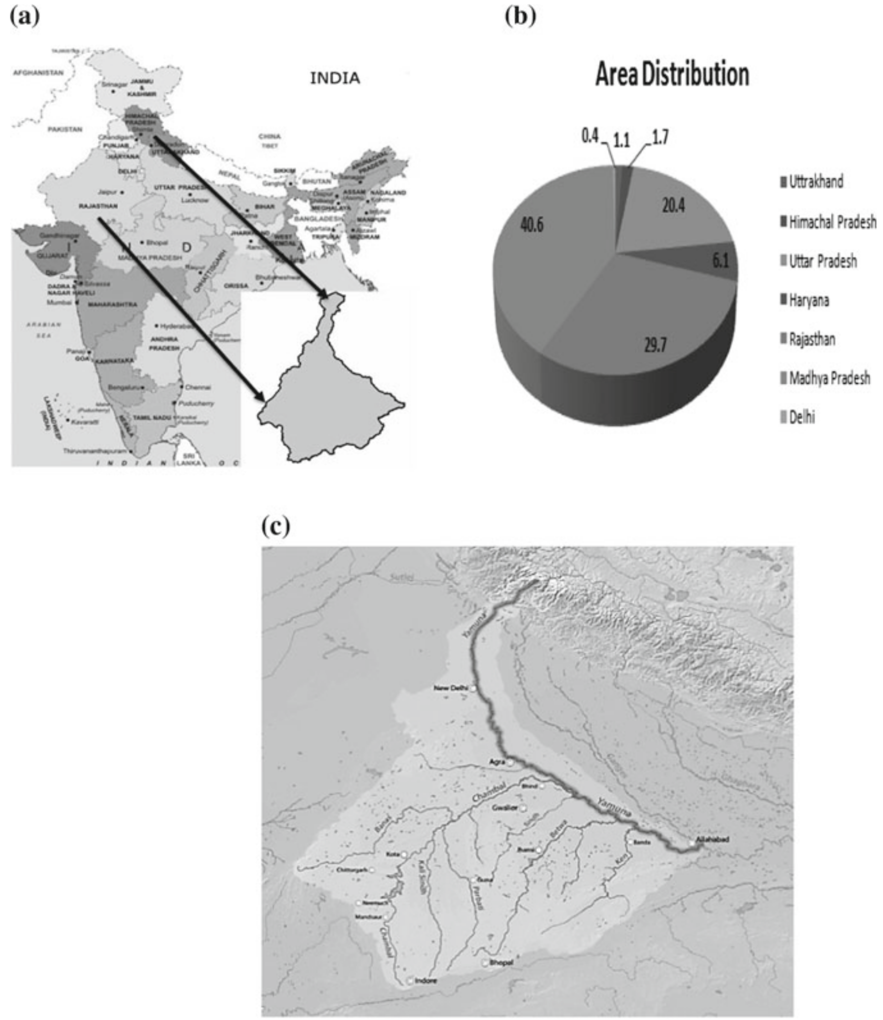


**Fig. 2** **a** Represents the location of study area in the context of Indian topology; **b** represents the area distribution of Yamuna river Basin state-wise; **c** geographical location of Yamuna river basin

# 5 Results

## *5.1 Calibration and Validation*

The ANN model was developed for the Yamuna river basin, and the model was calibrated for the time period (1971–1990), validated for 1991–2005. Maximising the efficiency of the model is the main objective of the method. The time series

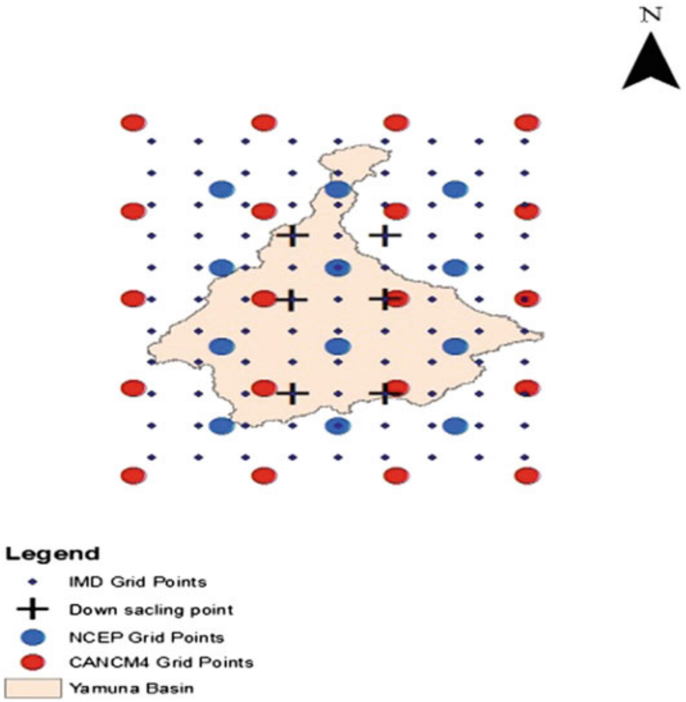


**Fig. 3** Geographical position of NCEP grid points covering the Yamuna basin

**Table 1** List of predictor variables

| Predictor variable | Pressure level (mb) |
|---|---|
| Air temperature | 1000,850,500,250 |
| Eastward wind | 1000,850,500,250 |
| Northward wind | 1000,850,500,250 |
| Geopotential height | 1000,850,500,250 |
| Surface temperature | |
| Mean sea level pressure | |

Team, E. W. (n.d.). Physical Sciences Division. Retrieved, from https://www.esrlnoaa.gov/psd/data/gridded/data.ncep.reanalysis.html

plot for observed and simulated rainfall for both calibration and validation period is shown in Fig. 4c, d.

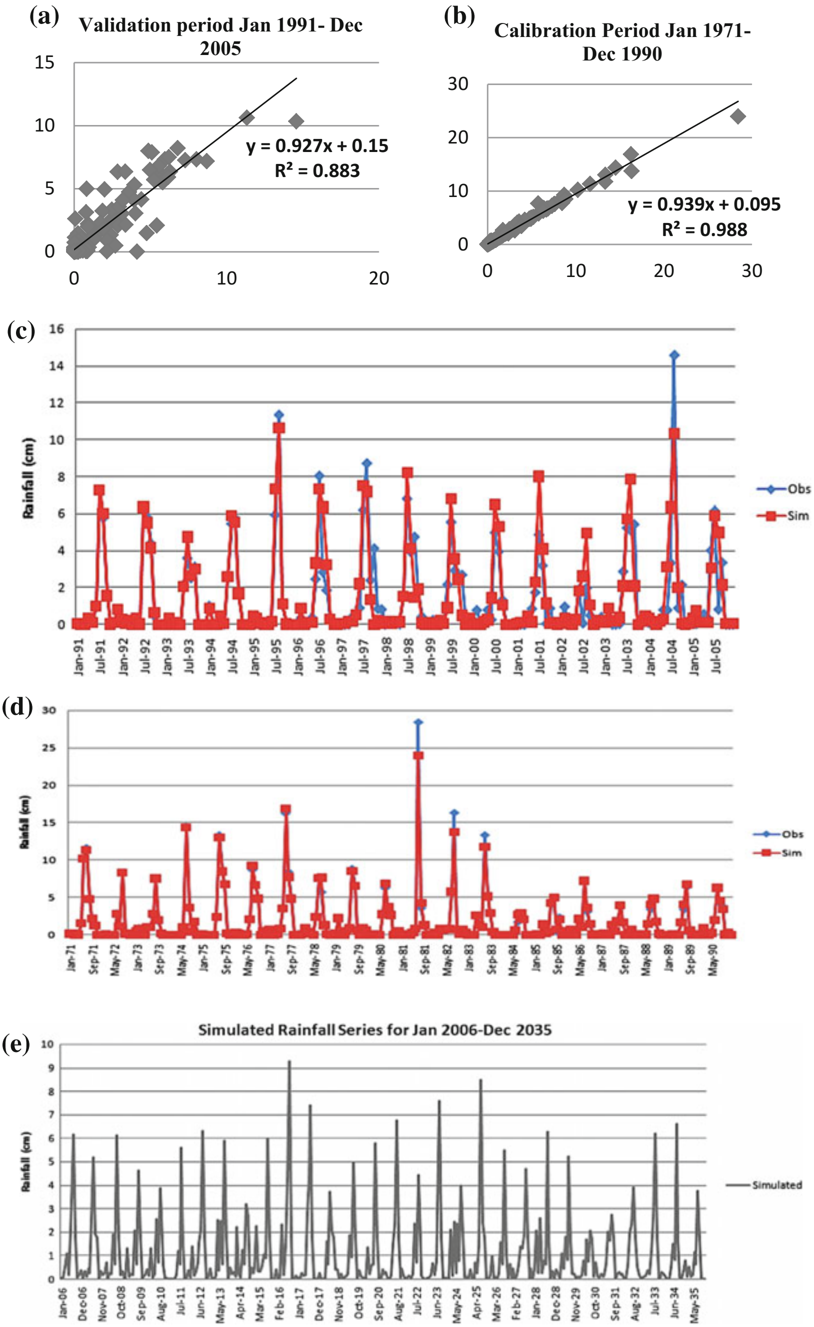


**Fig. 4** **a** Scatter plot for calibration period, **b** scatter plot for validation period, **c** observed and simulated rainfall for 1971–1991, **d** observed and simulated rainfall for 1991–2005, **e** projected rainfall for 2006–2035

Monthly observed and simulated precipitation was compared at six different locations in the basin. Comparison for station "X" is shown in this paper. It is observed that observed and simulated results show a similar trend for both calibration and validation period. The developed model shows satisfactory results, and the statistical analysis for the calibration and validation period can be seen in the table. The coefficient of determination ($R^2$) is calculated, and the result obtained is within an acceptable range. Due to this, we accept the model and it is used for projection or forecasting of precipitation till 2035 on the river basin which is shown in Fig. 4e. Scatter plot for calibration and validation period is shown in Fig. 4a, b.

## 6 Conclusion

For the present study, we have downscaled GCM output to precipitation at a finer resolution of $0.5 \times 0.5$ over Yamuna river basin. This task is achieved by applying statistical downscaling using the artificial neural network. The results show that simulated rainfall using ANN for CanCM4 historical scenarios provides a good match with the observed data. For the future projection under RCP4.5 scenarios, it is observed that there will be a decrease in rainfall. The projected rainfall may create a hypothesis that wet area will get wetter and dry area will get drier. The projection shows a significant trend and shows that there will be a major change in the magnitude and pattern of rainfall. The model developed simulated satisfactory results but failed to capture extreme events. This method needs to be modified so that it will be able to capture extreme events. The reason for this failure might be due to the use of principal component analysis. PCA neglects few components which show a small percentage of original variability but these components might be important for explaining extreme events. The overall contribution of this study is a projection of rainfall at a finer resolution till the year 2035. Rainfall projection can be used by policymakers for calculating water availability at the regional level and for inter-basin water transfer. It can further be used for management of agricultural water, hydraulic structure design. It is advised to use different types of downscaling models, GCM and emission scenarios when we need to study climate change.

**Acknowledgements** This research was funded by Deutsche Forschungsgemeinschaft (DFG) (GRK 2043/1) within the graduate research training group Natural risk in a changing world (NatRiskChange) at the University of Potsdam (http://www.uni-potsdam.de/natriskchange).

## References


1. Anandhi, A., Srinivas, V.V., Nagesh Kumara, D., Ravi, N.: 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 (2009)
2. Lakhanpal, A.: Statistical downscaling of GCM outputs using wavelet based model. Doctoral dissertation, Indian Institute of Technology, Delhi (2015)
3. Sehgal, V., Sridhar, V., Tyagi, A.: Stratified drought analysis using a stochastic ensemble of simulated and in-situ soil moisture observations. J. Hydrol. **545**, 226–250 (2017)
4. Benestad, R.E., Hanssen-Bauer, I., Forland, E.J.: An evaluation os statistical models for downscaling precipitation and their ability to capture long-term trends. Int. J. Climatol. **2795**, 649–655 (2007)
5. Coulibaly, P., Dibike, Y.B.: Downscaling of global climate model outputs for flood frequency analysis in the Saguenay river system. Final Project Report prepared for the Canadian Climate Change Action Fund, Environment Canada, Hamilton, Ontario, Canada (2004)
6. Chu, J.T., Xia, J., Xu, C.Y., Singh, V.P.: Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China. Theor. Appl. Climatol. **99(**1–2), 149–161 (2010)
7. Wilby, R.L., Wigley, T.M.L.: Downscaling general circulation model output: a review of methods and limitations. Prog. Phys. Geogr. **21**, 530–548 (1997)
8. Agarwal, A.: Hydrologic regionalization using wavelet based multiscale entropy technique. Doctoral dissertation, Indian Institute of Technology, Delhi (2015)
9. Lakhanpal, A., Sehgal, V., Maheswaran, R., Khosa, R., Sridhar, V.: A non-linear and non-stationary perspective for downscaling mean monthly temperature: a wavelet coupled second order Volterra model. Stoch. Env. Res. Risk Assess. **31**(9), 2159–2181 (2017). https://doi.org/10.1007/s00477-017-1444-6