Efficacy of hybrid neural networks in statistical downscaling of precipitation of the Bagmati River basin
Keshav Kumar, Vivekanand Singh and Thendiyath Roshni

ABSTRACT
This study investigates and analyses the present and future scenarios of precipitation using statistical downscaling techniques at selected sites of the Bagmati River basin. Statistical downscaling is achieved by feed forward neural network (FFNN) and wavelet neural network (WNN) models. Potential predictors for the model development are selected based on the performances of Pearson product moment correlation and factor analysis. Different training algorithms are compared and the traincgb training algorithm is selected for development of FFNN and WNN models. The visual comparison and the statistical performance indices were calculated between observed and predicted precipitation. From the analysis of results, it is evident that WNN models were well capable of (training: RMSE 1.61 – 1.67 mm, R 0.94 – 0.952; testing: RMSE 1.68 – 1.78 mm, R 0.93 – 0.95) predicting precipitation followed by FFNN model for all the selected sites. Hence, the projected precipitation (2014 – 2036) is found by WNN model only with inputs as different GCMs data. The projected precipitation results are analysed for the period 2014 – 2036 and find that there is a decrease in precipitation with respect to base period data (1981 – 2013) by 66.62 to 84.21% at Benibad, 4.53 to 21.74% at Dhenge and 6.40 to 22.27% at Kamtaul, respectively.

Key words | factor analysis, FFNN, potential predictor, statistical downscaling, WNN

INTRODUCTION
Statistical downscaling method has been widely used in projection of climatic variables in recent years. Climatic projections help to assess the future precipitation and thereby the future precipitation scenario at the global and regional scale (Orlowsky & Seneviratne 2013). General circulation models (GCMs) are commonly used and are the most efficient tools available for projections of climatic variables. For regional scale analysis, the climatic projections cannot be used directly and hence downscaling techniques are used to fill the gap between the GCM output and the climatic variables (Sachindra et al. 2018). Downscaling is an approach to take known information at large scales to make a prediction at local scales. The downscaling process relies on the premise that there exists a linear or nonlinear relationship between the predictors and the predictand. There are two main approaches, i.e., dynamic and statistical techniques to downscale climate information. Dynamic downscaling effectively increases the spatial resolution, through coupling a higher resolution numerical model (e.g., regional) to a lower resolution model (Giorgi & Mearns 1991; Hostetler et al. 2011; Xu & Yang 2012; Mannig et al. 2013; Ashfaq et al. 2016). In the last few years, many statistical and dynamic downscaling techniques have been developed to take information from GCMs and transfer information at the local scale (Carter & Kenkyy 1994; Anandhi et al. 2009). A statistical downscaling method has the advantage of being simple and easier to downscale the climate variable than the dynamic downscaling method. Various methods of statistical downscaling have different complexity during the downscaling process. The statistical method is categorized into three
categories. First is linear methods, such as regression, quantile regression, canonical correlation analysis, etc. Second is deterministic nonlinear methods like ANN (artificial neural network), SVM (support vector machine) and the third classification method includes weather generator and regression tree (Zorita & Von Storch 1999). The relationship between predictor and predictand can be determined by various methods such as regression analysis (Huth 1999; Haylock et al. 2006; Benestad & Haugen 2007), principal component analysis (PCA) (Cavazos 1997; Schubert 1998; Ghosh & Mujumdar 2006), K-nearest neighbor (KNN) (Gangopadhyay et al. 2005), Kernel regression analysis (Kannan & Ghosh 2013), ANN (Hewitson & Crane 1994; Gardner & Dorling 1998; Cannon & Lord 2000; Schoof & Pryor 2001; Chadwick et al. 2011; Fiseha et al. 2012; Goyal et al. 2012), SVM (Hua et al. 2008; Anandhi et al. 2009) and RVM (Ghosh & Mujumdar 2008; Tareghian & Rasmussen 2013). Although the literature on statistical downscaling shows many studies on the comparison of different machine learning techniques for climate change, it lacks in the strengths and correlation of individual predictors, predictand and the effectiveness of training algorithms in the model development.

Hence, the aim of the present study is to place emphasis on the selection of the potential predictors using two statistical methods, i.e., Pearson product moment correlation and factor analysis and to investigate the suitability of the training algorithm by comparing different training algorithms (traincgb, trainlm, traingdx, traingda, traingdm, traingd) using a feed forward neural network (FFNN) model. Further, to develop FFNN and wavelet neural network (WNN) models for downscaling the monthly precipitation, coarse GCM data are collected from CMCC-CMS RCP 4.5 scenario, MPI-ESM-MR RCP 4.5 scenario and MPI-ESM-MR RCP 8.5 scenario to project the precipitation over the period of 2014–2036.

**STUDY AREA AND DATA USED**

The Bagmati River is a perennial river of Nepal and India, especially of north Bihar. The Bagmati River originates from the Shivpuri range of hills in Nepal at latitude 27°47’ N and longitude 85°17’ E, and 16 km north east of Kathmandu at an altitude of 1,500 m above mean sea level. The total length of the river is 589 km and the total catchment area is 14,384 km², of which 6,500 km² falls in Bihar region. It passes through nearly 195 km in Nepal and the remaining 394 km in Bihar. The average annual rainfall of the Bagmati basin is 1,255 mm. The present study area (selected Bagmati river basin) is shown in Figure 1.

The land in the study area is mainly used for horticultural and agricultural purposes with 20% of the area under non-agricultural use such as roads, railways, water-bodies, buildings, etc. No forest cover is available in the study area. The climatic condition of the Bagmati River basin is changeable due to its topography. For this study, monthly rainfall data of 33 years, i.e., from 1981 to 2013 at three rain gauge stations, namely, Benibad, Dhenge Bridge and Kamtaul have been collected from Indian Meteorological
Department (IMD), Pune. Reanalysis data of different variables (air temperature, geopotential height, specific humidity, relative humidity) of 33 years, i.e., from 1981 to 2013 was downloaded from National Centre for Environmental Prediction (NCEP)/National Centre for Atmospheric Research (NCAR) at different pressure levels (100 mb to 1,000 mb) which is shown in the appendix (Table A1). NCEP/NCAR reanalysis variables data are used as inputs and observed precipitation data from IMD, Pune was used as target. Data for two future climate models were obtained from the CMIP5 project (Coupled Model International Project 5): MPI-ESM-MR model and CMCC-CMS model. The input variables for the projected period were downloaded from RCP 4.5 and RCP 8.5 scenarios of the MPI-ESM-MR model and RCP 4.5 scenario of the CMCC-CMS model (2014 to 2036). The homogeneity and stationarity test of observed precipitation data have also been carried out. In this study, 396 observations of 50 climatic variables are considered for the projection of precipitation.

**METHODOLOGY**

Data analysis is necessary to understand the behaviour of data. In this data analysis, homogeneity and stationarity of the observed rainfall data are analysed by Pettitt test and augmented Dickey–Fuller test.

**Pettitt test**

The Pettitt test is a non-parametric test that requires no assumption about the distribution of data.

The Pettitt test is an adaptation of the rank-based Mann–Whitney test that allows identifying the time at which the shift occurs (Pettitt 1979). The Pettitt test does not detect a change in distribution if there is no change of location. In this test, null hypothesis, $H_0$: there is a unit root for the series and alternate hypothesis, $H_a$: there is no unit root for the series, i.e., the series is stationary. The ranks $r_1 \ldots r_n$ of the $Y_1 \ldots Y_n$ are used to calculate the statistic (Pettitt 1979):

$$X_k = 2 \sum_{i=1}^{k} r_i - k(n + 1), \quad k = 1, 2 \ldots \ldots n$$  

If a break occurs in year $k$ then the statistic is maximal or minimal near the year $k$:

$$P_k = \max |X_k| \quad 1 \leq k \leq n$$  

where $P_k$ is Pettitt’s statistic.

**Augmented Dickey–Fuller (ADF) test**

The augmented Dickey–Fuller test consists of the following equation (Dickey & Fuller 1979; Cheung & Lai 1995):

$$\Delta y_t = \mu + \gamma t + \alpha y_{t-1} + \sum_{j=1}^{k-1} \beta_j \Delta y_{t-1} + \mu_t$$  

where $\alpha, \beta$ are constants, $y_t$ is the time series to be tested for unit root, $\gamma$ is the coefficient of time trend, $t$ is the time trend, $\mu_t$ is a white noise error term and $\Delta$ is the difference operator. The test examines the negativity of the parameter $\alpha$ based on its regression $t$ ratio. Dickey & Fuller (1979) found the asymptotic distribution of the statistics. ADF estimates $t$ value of the coefficient, which follows the $\tau$ (tau) statistic. The tau statistic or test is known as the DF test. DF has computed critical tau values and can be obtained in a tabulated form.

However, if the computed absolute value of the tau statistic ($\tau$) is less than the critical tau values, null hypothesis is rejected, in which case, the time series is stationary. On the other hand, if the computed tau statistic is less than the critical tau value, the null hypothesis is not rejected, in which case, the time series is non-stationary.

**Selection of potential predictors**

Selection of predictors for statistical downscaling is crucial for establishing the best relationship between predictors (like relative humidity, geopotential height, relative humidity, specific humidity, etc.) and predictand (like precipitation). Two statistical techniques (Pearson product moment correlation and factor analysis) have been used to select the best predictors for downscaling the precipitation. The main criterion for the selection of suitable predictors for downscaling is that they should
be strongly correlated with the predictands (Grimes et al. 2003). Consequently, Pearson product moment correlation and factor analysis are the most widely used approaches for picking potential predictors (Huang et al. 2012; Hassan et al. 2014; Pervez & Henebry 2014). The complete methodology is explained using a flow chart as shown in Figure 2.

**Pearson product moment correlation**

The correlation coefficient between different variables data sets have been computed based on product moment correlation coefficient (Häne et al. 1993), which is given as:

$$\rho_{xy} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{N\sigma_x\sigma_y}$$  \hspace{1cm} (4)
where, \( \rho_{xy} \) is correlation coefficient, \( x_t \) is predictand, \( y_t \) is predictor, \( \bar{x} \) is mean of predictand, \( \bar{y} \) is mean of predictor, \( \sigma_x \) is standard deviation of predictand, \( \sigma_y \) is standard deviation pf predictand and \( N \) is total number of data.

In statistical analysis, Pearson product moment correlation evaluates the linear correlation between two variables \( x \) (i.e., predictors) and \( y \) (i.e., predictand) that gives a value between \(-1\) and \(+1\) and was developed by Karl Pearson. Positive values (i.e., \( 0 \) to \( 1 \)) of the correlation coefficient shows a positive correlation between the variables \( x \) and \( y \); whereas, negative values (i.e., \(-1 \) to \( 0 \)) shows a negative correlation between variables \( x \) and \( y \) (Ratner 2009; Puth et al. 2014; Rodríguez-Roblero et al. 2019).

**Factor analysis**

Factor analysis has been used to minimize more variables into fewer variables. This method separates the most extreme basic variance from all factors and places them into a typical score. From the record of all factors, this score has been utilized for facilitating the investigation. This method has various assumptions: (1) there is no multicollinearity, (2) there is a linear relationship, (3) there exists a true correlation between factors and variables and (4) it includes relevant variables in the analysis.

The factor analysis can algebraically be described as follows (Rencher 2003). There are \( p \) variables \( X_1, X_2, \ldots, X_p \) measured on a sample of \( n \) subjects, then variable \( i \) can be written as a linear combination of \( m \) factors \( F_1, F_2, \ldots, F_m \) where, as explained above, \( m < p \). Thus,

\[
X_i = a_{i1}F_1 + a_{i2}F_2 + \ldots + a_{im}F_m + e_i
\]

where the \( a_{ij} \) is the factor loadings (or scores) for variable \( i \) and \( e_i \) is the part of variable \( X_i \) that cannot be ‘explained’ by the factors.

Criteria for determining the number of factors: eigenvalue is a qualifying criterion for deciding a factor. If eigenvalue is greater than 1, it is considered as a factor for further analysis and if eigenvalue is less than 1, then that factor is not considered for analysis. Selection criteria for selecting potential predictors by factor analysis are: (a) prepare the data, (b) decide how many components/factors to extract using principal factor analysis, (c) calculation of eigenvalues of components/factors, (d) the factor whose eigenvalue is greater than 1 and variability is nearest to 100% has been selected for further analysis, (e) factors are rotated using orthogonal rotation method and (f) extract the components/factors pattern.

In this study, predictors have been investigated through statistical methods and potential predictors have been selected by comparing two statistical methods as mentioned above.

**Development of models**

**Feed forward neural network (FFNN)**

The commonly used neural network is a three-layered feed forward network because of its common suitability to a variety of problems. The input layer is the first layer and its contribution is to transfer the input variables onto the forthcoming layers of the network. The last layer consists of the output variables and the layer is known as output layer. The layer in between the input and output layer is called the hidden layer and the presence of this layer increases the network’s efficiency to model complex functions. The processing elements (neurons) in every layer are called nodes. The information flow and processing in this network is done from input layer to hidden layer and from hidden layer to output layer. The problem to be addressed decides the number of nodes present in input and output layers. The number of hidden layers and the number of nodes in each hidden layer are problem dependent and are commonly determined by a trial and error procedure. A synaptic weight is assigned to each link that shows the relative connection strength of two nodes at both ends in predicting the input–output relationship. Three-layered FFNN are taken for the development of the FFNN model for downscaling precipitation. The learning algorithm is adopted with a back propagation algorithm based on the generalized delta rule proposed by Rumelhart et al. (1986). In this algorithm, the connection weight between the nodes is adjusted by the strength of the signal in the connection and the total measure of the error. The total error between input and output is reduced by redistributing this error backwards. The weights in the connection are adjusted in this step. The output is computed from the FFNN model...
using the adjusted weights. Back propagation continues for the number of given cycles or until a prescribed error tolerance is reached. As mentioned by Dawson & Wilby (1998), the transfer function (tansig, logsig and purelin) and internal parameters (max fail, minimum gradient, error, delta values) are to be considered to make the network learning more generalized. The best fit structure of the FFNN model is determined in the training process. The FFNN model structure is shown in Figure 3.

**Training algorithms**

Different training algorithms, as given in Table 1, have been used to train a network. These training functions are used to update weight and bias values of the neural network. Three types of training algorithms have been used for downscaling the precipitation. They are quasi-Newton algorithms (trainlm), conjugate gradient algorithms (traincgb) and gradient descent algorithms (traind, traingdm, traingdx, traingda). The training functions trainlm, traingda, traingdx, traincgb, traingd are used for training of the FFNN model (Coskun & Yildirim 2003; Zhang 2004; Baptista et al. 2013). The training algorithm is used to train any neural network.

**Wavelet denoising method**

Selected potential predictors from NCEP/NCAR reanalysis data are decomposed by wavelet using one dimensional Daubechies discrete wavelets, up to the second level. The decomposition level of the wavelet is selected by \[ \log_{10}(N) \] where \( N \) = total number of observation data. Input variables are decomposed into detail signal and approximate signal up to the second level. Minmax threshold is used to denoise the decomposed signals. Then, these denoised signals are used for development of WNN.

**Wavelet neural network**

The wavelet is used to break the signals into various parts of frequency using wavelet decomposition tool. Every individual component of frequency is compared with the original frequency signal. Wavelet transformation is mainly classified as continuous and discrete wavelets. Wavelet is represented by a small wave function. Waves which decay

---

**Table 1 | Names of training algorithms**

<table>
<thead>
<tr>
<th>Function name</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>trainlm</td>
<td>Levenberg–Marquardt back propagation</td>
</tr>
<tr>
<td>traingda</td>
<td>Gradient descent with adaptive lr back propagation</td>
</tr>
<tr>
<td>traingdx</td>
<td>Gradient descent momentum and adaptive lr back propagation</td>
</tr>
<tr>
<td>traincgb</td>
<td>Powell–Beale conjugate gradient back propagation</td>
</tr>
<tr>
<td>traingdm</td>
<td>Gradient descent with momentum back propagation</td>
</tr>
<tr>
<td>traingd</td>
<td>Gradient descent back propagation</td>
</tr>
</tbody>
</table>
in a finite period of time are known as small waves, and long waves decay repeatedly over an infinite time period. A wavelet function is described by the following equations (Veitch 2005):

\[ \int_{-\infty}^{\infty} \Psi(u) \, du = 0 \]  \hspace{1cm} (6)

\[ \int_{-\infty}^{\infty} \Psi^2(u) \, du = 0 \]  \hspace{1cm} (7)

Admissibility condition, \[ C_{\Psi} \int_{-\infty}^{\infty} \frac{|\Psi(f)|^2}{f} \, df \]  \hspace{1cm} (8)

where \( 0 < C_{\Psi} < \infty, \) \( u = \) input wavelet signal

\[ \Psi_{\lambda,t}(u) = \frac{\Psi}{\lambda} \left( \frac{u-t}{\lambda} \right) \]  \hspace{1cm} (9)

where \( \Psi(.) \) represents mother wavelet, \( t \) indicates finite translation parameter and \( \lambda \) is dilation parameter, \( \lambda > 0. \)

The right side of Equation (9) is normalized, so that \( \| \Psi_{\lambda,t} \| = ||\Psi|| \) for all \( \lambda \) and \( t. \)

Selection of the mother wavelet depends upon the signal to be analysed. Mainly, Morlet and Daubechies wavelet transform will be used as the ‘mother’ wavelet (Shoaib et al. 2014). Daubechies wavelet shows a good interconnection between parsimony and data abundance, it gives approximately similar events over the observed time sequence and shows up in such a large number of various patterns that most forecast models cannot distinguish them well.

**Continuous wavelet transformation (CWT)**

Wavelet transformation is employed to show the best match for any signal \( x(.) \) by calculating the amplitude coefficient to get a given value of \( \lambda, t \) (Veitch 2005):

\[ (x, \Psi_{\lambda,t}(u)) = \int_{-\infty}^{\infty} \Psi_{\lambda,t}(u) \, x(u) \, du = 0 \]  \hspace{1cm} (10)

for

\[ \int_{-\infty}^{\infty} x^2(t) \, dt < \infty \]  \hspace{1cm} (11)

After that, the original signal is recovered from continuous wavelet transformation by the inverse transfer function.

\[ X(t) = \frac{1}{C_{\Psi}} \int_{-\infty}^{\infty} (x, \Psi_{\lambda,t}(u)) \, du \]  \hspace{1cm} (12)

**Discrete wavelet transform (DWT)**

DWT mainly works with ‘high pass filters’ and ‘low pass filters’ (Krishna et al. 2011). The primary signal is passed through ‘high pass filters’ and ‘low pass filters’ at a different frequency. The primary signal is decomposed into high-frequency low scale signal by high pass filters and low-frequency high scale signal by low pass filters. High-frequency signal (i.e., detail) capture very small features of the interpreted data whereas low-frequency signal (i.e., approximate) carry important features of the data and useful information about the signal. The decomposition process is repeated in every step to break the approximate signal into many lesser resolution components.

To explore more information, CWT analyses the signal over an infinite period of time after many translations and dilations of the primary wavelet. Detail information is usually reduced by taking a part of CWT and the important feature of the signal is preserved using DWT. In this study, DWT is considered for the analysis of input signals. The WNN model architecture is shown in Figure 5.

**Performance evaluation of model**

The performance of a model can be analysed in terms of various characteristics. The three important features of a good model are consistency, accuracy and versatility. The performance indicators are as follows (Kar et al. 2015):

Coefficient of correlation (R) \[ R = \frac{\sum x \, y}{\sqrt{\sum x^2} \sqrt{\sum y^2}} \]  \hspace{1cm} (13)

Nash–Sutcliffe efficiency (NSE) \[ \text{NSE} = 1 - \frac{\sum (t-y)^2}{\sum (t-\bar{t})^2} \]  \hspace{1cm} (14)

Index of agreement (D) \[ D = 1 - \frac{\sum (t-y)^2}{\sum \left[(t-\bar{t}) + (y-\bar{y})\right]^2} \]  \hspace{1cm} (15)

Root mean square error (RMSE) \[ \text{RMSE} = \sqrt{\frac{\sum_{n=1}^{N} (t-y)^2}{N}} \]  \hspace{1cm} (16)
Coefficient of determination ($R^2$)

$$R^2 = \frac{N(\sum ty) - (\sum t)(\sum y)}{\sqrt{[N\sum t^2 - (\sum t)^2][N\sum y^2 - (\sum y)^2]}}$$

where $N$ is the number of observations, $y$ is computed data, $t$ is the observed data, $T = t - \bar{t}$ in which $\bar{t}$ is the mean of the observed data and $Y = y - \bar{y}$ in which $\bar{y}$ is the mean of the output data.

The RMSE is used to measure the differences between predicted values and the observed values. A correlation coefficient value lies between −1 and +1, that quantifies a type of correlation and dependence, by developing meaningful statistical relationships between two or more variables in fundamental statistics. Nash–Sutcliffe efficiency is used to quantify how well a model simulation can predict the outcome variable and it shows the predictive power of any model. Nash–Sutcliffe efficiency can range from $-\infty$ to 1. Nash–Sutcliffe efficiency, $NSE = 1$, corresponds to a perfect match of the model to the observed data; $NSE = 0$, indicates that the model predictions are as accurate as the mean of the observed data; $-\infty < NSE < 0$, indicates that the observed mean is a better predictor than the model.

**Prediction of precipitation using RCP 4.5 and RCP 8.5 scenarios**

The best model has been selected after comparison of performance parameters of the two models. Different scenarios of CMIP5 climate model, i.e., RCP 4.5 and RCP 8.5 from the years 2014 to 2036, have been used in the selected model for prediction of precipitation. RCPs are different scenarios adopted by the IPCC for its Fifth Assessment Report. RCP shows the trajectory of greenhouse gas emissions.

**RESULTS AND DISCUSSION**

**Homogeneity and stationarity test**

The homogeneity and stationarity test for monthly rainfall time series at three stations in the Bagmati River basin were tested with Pettitt’s and Dickey–Fuller tests. Using the Pettitt test, all three rainfall stations were found to be homogenous. Acceptance of the null hypothesis ranges from 54% to 91%, which shows the observed data are homogenous.

Dickey–Fuller test is conducted to check the stationarity of the observed rainfall data. Two hypotheses have been selected, null hypothesis, $H_0$: there is a unit root for the series and alternate hypothesis, $H_a$: there is no unit root for the series, i.e., the series is stationary. Using the Dickey–Fuller test, all three stations observed rainfall data were found to be stationary. It is found the $p$-value at all the stations observed and NCEP rainfall data are less than 0.0001, which shows the null hypothesis is rejected. The results are shown in Table 2.

<table>
<thead>
<tr>
<th>Stations</th>
<th>Observed Observations</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. deviation</th>
<th>$P_\alpha$</th>
<th>$k$</th>
<th>$p$-value (one-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dhenge</td>
<td>396</td>
<td>0</td>
<td>892</td>
<td>106.618</td>
<td>158.403</td>
<td>−1.298</td>
<td>52</td>
<td>0.821</td>
</tr>
<tr>
<td>Benibad</td>
<td>0</td>
<td>718.3</td>
<td>99.3557</td>
<td>144.893</td>
<td>−2.339</td>
<td>51</td>
<td>0.546</td>
<td></td>
</tr>
<tr>
<td>Kamtaul</td>
<td>0</td>
<td>723.2</td>
<td>103.722</td>
<td>147.934</td>
<td>−901</td>
<td>52</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

**Augmented Dickey–Fuller (ADF)**

<table>
<thead>
<tr>
<th>Stations</th>
<th>Observed Observations</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. deviation</th>
<th>Tau</th>
<th>Tau (Critical value)</th>
<th>$p$-value (one-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dhenge</td>
<td>0</td>
<td>892</td>
<td>106.618</td>
<td>158.403</td>
<td>−12.347</td>
<td>−0.874</td>
<td>&lt;0.0001 \</td>
<td></td>
</tr>
<tr>
<td>Benibad</td>
<td>0</td>
<td>718.3</td>
<td>99.3557</td>
<td>144.893</td>
<td>−13.532</td>
<td>−0.874</td>
<td>&lt;0.0001 \</td>
<td></td>
</tr>
<tr>
<td>Kamtaul</td>
<td>396</td>
<td>723.2</td>
<td>103.722</td>
<td>147.934</td>
<td>−13.803</td>
<td>−0.874</td>
<td>&lt;0.0001 \</td>
<td></td>
</tr>
</tbody>
</table>
Partial autocorrelation test is performed to identify the extent of the lag in the autoregressive model. Partial autocorrelation plot of rainfall at three rain gauge stations are shown in Figure 4. The result shows significant correlation at lag time $t-10$, $t-11$ and $t-12$ months at all the three stations.

Selection of potential predictor

Pearson product moment correlation

Potential predictors like air temperature, geopotential heights, relative humidity, specific humidity available at different pressure levels from NCEP/NCAR reanalysis are correlated with observed monthly precipitation. The correlation coefficient of predictors with precipitation at different pressure level is presented in Table 3. Good coefficient of correlation 0.69 and 0.69 of rainfall with air temperature has been obtained at pressure levels 500 mb and 700 mb, respectively. Air temperature at pressure levels 250 to 700 mb, geopotential height at pressure levels 100 to 300 mb, relative humidity at pressure levels 500 to 850 mb and specific humidity at pressure levels 300 to 1,000 mb show more or less the same value of correlation coefficient and thus these predictors have been selected.

Figure 4 | Partial autocorrelation of rainfall data at three rain gauge stations. (a) Benibad, (b) Dhenge Bridge and (c) Kamtaul.
Factor analysis has been done to minimize the number of variables.

**Factor analysis**

Potential predictors available at different pressure level from the NCEP/NCAR Reanalysis model are factor analysed with observed monthly precipitation. Selection of the numbers of factors that influence the precipitation has been made. The eigenvalue and factor pattern between precipitation and predictors are shown in Tables 4 and 5. Among three factors, f1, f2 and f3, the eigenvalue of f1 is greater than 1 and factors f2 and f3 have eigenvalues less than 1. For variability of factor, f1 has 91.02% and the other factors (f2 and f3) have 8.98%. The factor f1 has been selected having an eigenvalue greater than 1 and variability nearest to 100%. Table 5 presents the factor pattern/score of the predictors by considering factor f1. The values of specific humidity are 0.996 and 0.991 at pressure level 600 mb and 700 mb, respectively, which is close to 1. This shows that these two factors influence the precipitation significantly.

Air temperature at pressure level 600 mb and 700 mb, geopotential height at pressure level 50 mb and 70 mb, relative humidity at pressure level 500 mb and 700 mb and specific humidity at pressure level 600 mb and 700 mb have been selected as potential predictors whose values are close to 1 (Table 5).

These selected potential predictors have been used to develop FFNN and WNN models.

**Development of models**

Two models, FFNN with back propagation and WNN with denoised signals have been developed. Pettitt’s test for homogeneity and Dickey–Fuller test for stationarity have been performed for the observed data of rainfall at all the three rain gauge stations from 1981 to 2013. Tests show that the time series data of rainfall are homogeneous and stationary.

**FFNN model**

The monthly data of selected predictors and observed precipitation from 1981 to 2013 have been used in the development of the model. The data from the years 1981 to 2003 have been used for the calibration and from 2004 to 2013 for the validation.

**Selection of training algorithm using FFNN model**

The network has been trained initially by considering selected potential predictors as input vector and precipitation as the output vector. The network has also been trained using different training functions with error tolerance, the number of cycles for learning, learning parameter and the neurons of the hidden layer. The output values from the network have been denormalized and compared with the observed targeted values. Performance criteria such as Nash–Sutcliffe efficiency (NSE), coefficient of correlation (R), index of agreement (D) and root mean square error (RMSE) have been used to examine the performance of the model. For training, different training algorithms (traincgf, trainlm, traindx, trainda, traingdm, traingd) have been used and a comparison is shown in

---

**Table 3** Correlation value of predictors using Pearson product moment correlation

<table>
<thead>
<tr>
<th>Pressure level (mb)</th>
<th>Air temperature</th>
<th>Geopotential height</th>
<th>Relative humidity</th>
<th>Specific humidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>0.55</td>
<td>-0.66</td>
<td>0.66</td>
<td>0.73</td>
</tr>
<tr>
<td>925</td>
<td>0.57</td>
<td>-0.67</td>
<td>0.66</td>
<td>0.73</td>
</tr>
<tr>
<td>850</td>
<td>0.6</td>
<td>-0.68</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td>700</td>
<td>0.69</td>
<td>-0.42</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>600</td>
<td>0.68</td>
<td>0.28</td>
<td>0.74</td>
<td>0.76</td>
</tr>
<tr>
<td>500</td>
<td>0.69</td>
<td>0.54</td>
<td>0.7</td>
<td>0.75</td>
</tr>
<tr>
<td>400</td>
<td>0.68</td>
<td>0.62</td>
<td>0.64</td>
<td>0.74</td>
</tr>
<tr>
<td>300</td>
<td>0.67</td>
<td>0.66</td>
<td>0.44</td>
<td>0.73</td>
</tr>
<tr>
<td>250</td>
<td>0.67</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.14</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>-0.29</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>-0.29</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4** Eigenvalue of factors f1, f2 and f3

<table>
<thead>
<tr>
<th></th>
<th>f1</th>
<th>f2</th>
<th>f3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>8.1924</td>
<td>0.2100</td>
<td>0.0044</td>
</tr>
<tr>
<td>Variability</td>
<td>91.0269</td>
<td>2.3334</td>
<td>0.0486</td>
</tr>
<tr>
<td>Cumulative</td>
<td>91.0269</td>
<td>93.3603</td>
<td>93.4089</td>
</tr>
</tbody>
</table>
Figure 5. The traincgb training algorithm gave the highest coefficient of correlation, Nash–Sutcliffe efficiency, index of agreement and coefficient of determination, i.e., 0.89, 0.77, 0.94 and 0.89, respectively. Thus, traincgb training algorithm has been used for developing FFNN and WNN models in the Bagmati River basin.

For further analysis, traincgb training function has been taken for training the neural network. During training, the

<table>
<thead>
<tr>
<th>Air temperature at pressure level (mb)</th>
<th>Geopotential height at pressure level (mb)</th>
<th>Relative humidity at pressure level (mb)</th>
<th>Specific humidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.3883</td>
<td>0.9392</td>
<td>0.6516</td>
</tr>
<tr>
<td>20</td>
<td>0.9115</td>
<td>0.9715</td>
<td>0.8967</td>
</tr>
<tr>
<td>30</td>
<td>0.8341</td>
<td>0.9799</td>
<td>0.9595</td>
</tr>
<tr>
<td>50</td>
<td>0.4484</td>
<td>0.9934</td>
<td>0.95</td>
</tr>
<tr>
<td>70</td>
<td>−0.0663</td>
<td>0.9968</td>
<td>0.9675</td>
</tr>
<tr>
<td>100</td>
<td>−0.2064</td>
<td>0.9864</td>
<td>0.9518</td>
</tr>
<tr>
<td>150</td>
<td>0.2151</td>
<td>0.9718</td>
<td>0.9203</td>
</tr>
<tr>
<td>200</td>
<td>0.7087</td>
<td>0.9626</td>
<td>0.9162</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.9369</td>
<td>0.9724</td>
<td>0.9477</td>
</tr>
<tr>
<td>300</td>
<td>0.951</td>
<td>0.97</td>
<td>0.9589</td>
</tr>
<tr>
<td>400</td>
<td>0.963</td>
<td>0.9549</td>
<td>0.9773</td>
</tr>
<tr>
<td>500</td>
<td>0.9764</td>
<td>0.8937</td>
<td>0.9961</td>
</tr>
<tr>
<td>600</td>
<td>0.9901</td>
<td>0.6281</td>
<td>0.9913</td>
</tr>
<tr>
<td>700</td>
<td>0.9892</td>
<td>−0.3298</td>
<td>0.9918</td>
</tr>
<tr>
<td>850</td>
<td>0.9485</td>
<td>−0.884</td>
<td>0.9777</td>
</tr>
<tr>
<td>925</td>
<td>0.9253</td>
<td>−0.911</td>
<td>0.9749</td>
</tr>
<tr>
<td>1,000</td>
<td>0.9005</td>
<td>−0.9141</td>
<td></td>
</tr>
</tbody>
</table>
number of cycles has been increased in steps up to 4,000 and changing the internal parameter, i.e., max fail, minimum gradient, error, delta values of the network. It has been found that the convergence was not static for these numbers of cycles. Then, the network was trained with the different transfer functions (i.e., the increase or decrease values of error tolerance and varied values of the learning parameter). The learning parameter and the error tolerance have been fixed with low root mean square error, high values of coefficient of correlation, Nash–Sutcliffe efficiency and index of agreement. The neurons in the hidden layer have been increased from minimum to the highest coefficient of correlation. The number of neurons and internal parameters which give the highest coefficient of correlation has been selected for input combination.

Figure 6 | Calibration and validation of (a) FFNN and (b) WNN models at (1) Benibad, (2) Dhenge Bridge and (3) Kamtaul stations.
WNN model

Again, the monthly data of selected predictors and observed precipitation from 1981 to 2013 have been used for the development of the model. The monthly data of selected predictors have been decomposed by wavelet using one-dimensional Daubechies discrete wavelets, up to the second level. The decomposition level of wavelet has been selected by \( \log_{10}(N) \), where \( N \) = total number of observation data. In this study, total number of observation data was 396 and the value of \( \log_{10}(N) \) is 2.59. Hence, decomposition of the signal has been done up to the second level. In the WNN model, 24 input variables (8 D1, 8 D2 and 8 A2) are generated after decomposition of eight potential predictors where D1, D2 and A2 are detail coefficient at level 1, 2 and approximation at level 2, respectively. Minmax threshold has been used to denoise the decomposed signals. These denoised signal data from the years 1981 to 2003 have been fed to the neural network model for calibration, and trained the model with traincgb training algorithm. During training, the number of cycles has been increased in steps up to 4,000 and changing the internal parameter, i.e., max fail, minimum gradient, error, delta values of the network. Data from 2004 to 2013 have been used for the validation of the WNN model.

Calibration and validation of the monthly FFNN and WNN models of Benibad, Dhenge Bridge and Kamtaul stations

The observed data of rainfall from the years 1981 to 2003 have been used for the calibration and from 2004 to 2013 have been used for the validation of FFNN and WNN models. It has been noticed that the best convergence was achieved in FFNN model using traincgb training algorithm. The weights for this best-trained structure were saved to evaluate the trained network. The potential input vector was used for the training and validation of the neural network. Figure 6 shows the computed precipitation for the calibration period (1981 to 2003) and validation period

<table>
<thead>
<tr>
<th>Stations</th>
<th>Calibration</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Validation</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NSE</td>
<td>R</td>
<td>D</td>
<td>RMSE</td>
<td>NSE</td>
<td>R</td>
<td>D</td>
<td>RMSE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benibad</td>
<td>FFNN</td>
<td>0.76</td>
<td>0.88</td>
<td>0.94</td>
<td>2.21</td>
<td>0.74</td>
<td>0.87</td>
<td>0.94</td>
<td>2.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WNN</td>
<td>0.88</td>
<td>0.94</td>
<td>0.974</td>
<td>1.61</td>
<td>0.88</td>
<td>0.938</td>
<td>0.97</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>Dheng Bridge</td>
<td>FFNN</td>
<td>0.69</td>
<td>0.86</td>
<td>0.926</td>
<td>2.67</td>
<td>0.68</td>
<td>0.857</td>
<td>0.92</td>
<td>2.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WNN</td>
<td>0.89</td>
<td>0.952</td>
<td>0.97</td>
<td>1.67</td>
<td>0.89</td>
<td>0.95</td>
<td>0.97</td>
<td>1.69</td>
<td></td>
</tr>
<tr>
<td>Kamtaul</td>
<td>FFNN</td>
<td>0.82</td>
<td>0.89</td>
<td>0.94</td>
<td>2.27</td>
<td>0.81</td>
<td>0.88</td>
<td>0.94</td>
<td>2.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WNN</td>
<td>0.88</td>
<td>0.945</td>
<td>0.97</td>
<td>1.66</td>
<td>0.87</td>
<td>0.93</td>
<td>0.96</td>
<td>1.78</td>
<td></td>
</tr>
</tbody>
</table>

Table 6 | Summary of FFNN and WNN results for the calibration and the validation periods of meteorological stations

Table 7 | Computed statistics for observed and predicted rainfall data
Figure 7 | Box plots of the observed and the downscaled precipitation of FFNN and WNN in (a, b) Benibad, (c, d) Dheng Bridge and (e, f) Kamtaul stations.
(2004 to 2013) of FFNN and WNN models, respectively. The results of the WNN model matches better than the results of the FFNN model with the observed data. Computed peaks from the WNN model are more or less the same and there is no lag for the occurrence of the peaks.

**Evaluation of models using statistical parameters**

The statistical parameters (NSE, R, D and RMSE) have also been computed to examine the performance of the FFNN and WNN models in the calibration and validation for Benibad, Dhenge Bridge and Kamtaul stations and are presented in Table 6. The values R, NSE and D were higher and RMSE was lower in the WNN model than FFNN model. Thus, the WNN model shows better results than the FFNN model at all three stations during calibration and validation of the model, which signifies that the performance of the WNN model is better than the FFNN model. The statistics of the observed and generated precipitations have also been computed and are shown in Table 7. The

**Figure 8** | REC curve of FFNN and WNN models for (a) Benibad, (b) Dhenge Bridge and (c) Kamtaul stations.
mean, standard deviation and skewness of generated precipitation at three downscaling locations were qualitatively well captured by the WNN approach.

The results of FFNN and WNN are also shown in Figure 7 as box plots in order to compare the minimum, maximum and the median value of the observed and

![Figure 9](image-url) Comparison of observed and the downscaled monthly precipitation values of FFNN and WNN models at Benibad, Dheng Bridge and Kamtaul stations.
downscaled precipitation data. The statistics of downscaled precipitation in the months of June, July, August and September of the WNN model were comparatively the same as the observed statistics. However, it was not the same for the FFNN model. Thus, downscaled results of the WNN model are better than FFNN model.

The performances of FFNN and WNN models for the three stations were also evaluated using regression error characteristic (REC) curve and are shown in Figure 8. Area over the curve (AOU) in WNN was smaller than FFNN for the selected stations, which shows that the WNN model performs better than the FFNN model.

Figure 9 shows the comparison of observed and downscaled monthly precipitation values of FFNN and WNN models at Benibad, Dhenge Bridge and Kamtaul stations using scatter diagram. The result shows that precipitation

![Figure 8: Regression Error Characteristic (REC) curve for FFNN and WNN models.](image-url)

Area over the curve (AOU) in WNN was smaller than FFNN for the selected stations, which shows that the WNN model performs better than the FFNN model.

![Figure 9: Comparison of observed and downscaled monthly precipitation values.](image-url)

Figure 9 shows the comparison of observed and downscaled monthly precipitation values of FFNN and WNN models at Benibad, Dhenge Bridge and Kamtaul stations using scatter diagram. The result shows that precipitation
data of FFNN model were further from the regression line. However, the maximum number of precipitation data of the WNN model lies within the prediction interval and it was very close to the regression line with better regression.
coefficient than FFNN model. As a result, it was found that the performance of the WNN model was better than the FFNN model.

Projection of rainfall using RCP 4.5 and RCP 8.5 scenario

Figure 10(a)–10(c) show the observed monthly precipitation from 2009 to 2013 and projected monthly precipitation from 2014 to 2036 using the WNN model with RCP 4.5 and RCP 8.5 scenarios for Benibad, Dhenge Bridge and Kamtaul stations, respectively. The results of the RCP 4.5 and RCP 8.5 scenarios were almost the same and both results match well. The patterns of observed and predicted precipitations using RCP 4.5 and RCP 8.5 scenarios of the CMIP5 climate model have been found to be similar.

Projection of precipitation using CMCC-CMS RCP 4.5 and MPI-ESM-MR RCP 4.5 scenario

The projected precipitation from 2014 to 2036 using the WNN model with CMCC-CMS RCP 4.5 and MPI-ESM-MR RCP 4.5 scenario are presented in Figure 11. From Figure 11, it can be seen that the pattern of observed and predicted precipitation using two CMIP5 climate models are of a similar nature. Statistics of scenario forecasts obtained from the statistical downscaling model are represented in Table 8, which shows a decrease in monthly precipitation from 66.62 to 84.21% at Benibad and 6.40 to 22.27% at Kamtaul. There is an increase in precipitation by 9.48% in the years 2014–2024 and decrease in precipitation from 4.53 to 21.74% at Dhenge in different projections. The projection shows statistically significant changes in precipitation during monsoon season for future periods and is shown in Figure 12.

CONCLUSIONS

In this study, future projections of precipitation have been carried out using statistical downscaling at Benibad, Dhenge Bridge and Kamtaul stations in the Bagmati River basin. For this purpose, potential predictors based on good correlation are selected by comparing two statistical methods, i.e., Pearson product moment correlation and factor analysis. Factor analysis gives better correlation value than the Pearson product moment correlation. Selected potential predictors are air temperature at pressure level 600 mb and 700 mb, relative humidity at pressure level 500 mb and 700 mb, specific humidity at pressure level 600 mb and 700 mb and geopotential height at pressure level 50 mb and 70 mb. These potential predictors are used as inputs for developing the FFNN and denoised WNN models for prediction of precipitation. Different training algorithms have been compared during the development of FFNN and it is found that traincgb training algorithm gives the highest R, 0.89, NSE, 0.77 and index of agreement, 0.94. Based on the results, traincgb training algorithm has been used to develop two models, FFNN and denoised WNN models for prediction of precipitation. Both the models have been calibrated using potential predictors from the year 1981 to 2003 and validated from the year 2004 to 2013 for all three selected stations. Precipitation has been downscaled using RCP 4.5 and RCP 8.5 scenario.

Table 8 | Statistics of scenario forecasts obtained from statistical downscaling model and examining statistically significant changes in monsoon season precipitation for future periods

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Period</th>
<th>Benibad</th>
<th>Dhenge</th>
<th>Kamtaul</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>1981–2013</td>
<td>35.87</td>
<td>37.69</td>
<td>35.79</td>
</tr>
<tr>
<td>MPI-ESM RCP 4.5</td>
<td>2014–2024</td>
<td>−68.01</td>
<td>−21.74</td>
<td>−16.63</td>
</tr>
<tr>
<td></td>
<td>2025–2036</td>
<td>−73.50</td>
<td>−18.12</td>
<td>−6.40</td>
</tr>
<tr>
<td>MPI-ESM RCP 8.5</td>
<td>2014–2024</td>
<td>−74.91</td>
<td>9.48</td>
<td>−20.81</td>
</tr>
<tr>
<td></td>
<td>2025–2036</td>
<td>−77.25</td>
<td>−4.53</td>
<td>−12.83</td>
</tr>
<tr>
<td>CMCC RCP 4.5</td>
<td>2014–2024</td>
<td>−66.62</td>
<td>2.30</td>
<td>−20.29</td>
</tr>
<tr>
<td></td>
<td>2025–2036</td>
<td>−84.21</td>
<td>12.41</td>
<td>−22.27</td>
</tr>
</tbody>
</table>
of the CMIP5 climate model. The performances of FFNN and WNN models have been compared using statistical parameters (mean, median and standard deviation), boxplot, REC curves, and scatter diagram with predictive interval. The above performance indicators show the performance of the WNN model for the prediction from 2014 to 2036 was better. The future monthly precipitation for the period of 2014 to 2036 is projected with CMCC-CMS RCP 4.5, MPI-ESM-MR RCP 4.5 and RCP 8.5. It is found from the projected precipitation that there is a decrease in the precipitation at the three rain gauge stations in Bagmati River basin.

REFERENCES


Ratner, B. 2009 The correlation coefficient: its values range between +1/−1, or do they? Journal of Targeting, Measurement and Analysis for Marketing 17, 139–142.


Rodríguez-Robleiro, M., Ayon, J., Cascante, G., Pandey, M., Alyousef, R. & Topper, T. 2019 Application of correlation analysis techniques to surface wave testing for the evaluation of reinforced concrete structural elements. NDT & E International 102, 68–76.


Veitch, D. 2005 Wavelet Neural Networks and Their Application in the Study of Dynamical Systems. MSc dissertation, Department of Mathematics, University of York, UK.


First received 30 October 2018; accepted in revised form 24 June 2019. Available online 26 July 2019