Rainfall analysis and forecasting in Mumbai metropolitan region using artificial neural networks (ANN)

Sangita Mishra S¹, K. Phadtare², Solapure M³

1- Assistant Professor, Department of Civil Engineering, AIKTC School of Engineering and Technology, New Panvel, Navi Mumbai, Maharashtra.
2,3- BE Students, Department of Civil Engineering, Saraswati College of Engineering, Kharghar, Navi Mumbai, Maharashtra.

dr.sushreemishra80@gmail.com
doi:10.6088/ijcsen.6011

ABSTRACT

Irregular rainfall pattern and high surface runoff are the main reasons of water logging and subsequent flood conditions in the Mumbai metro Politian region. Also, the urban heat island effect leads to increased rainfall, both in amounts and intensity in the city. Therefore, research on rainfall patterns and their variations as well as prediction of rainfall for short term as well as long term periods is necessary in Mumbai region. In the present study, an attempt has been made to perform statistical analysis of rainfall for various duration and return periods, and analysis of rainfall trend and intensity pattern for Mumbai region. The records of the observed annual maximum daily rainfall were analyzed statistically for different rainfall return periods (2, 3, 5, 10, 15, 20, 30, 50, 70 years) during the period 1994 to 2013 for Santacruz rain gauge station using Gumbel’s extreme value distribution model. The forecasted rainfall values obtained from Gumbel’s model is then validated using Artificial Neural Network model. It was observed from the study that the model architecture 4-2-1 performed best with a (R²) value of 0.99 and mean absolute error of 0.04 for rainfall forecasting in Mumbai metropolitan region.

Keywords: Rainfall analysis, Rainfall forecasting, Mumbai, ANN

1. Introduction

Mumbai is a densely populated city that is located on the coast. Being a tropical wet-dry region, this island city receives high intensity rainfall from monsoon from June to September months and the other months are mostly dry resulting in extreme climatic events like floods and droughts. The study of spatial variability of extreme rainfall events help to identify zones of high and low values of extreme rainfall events. Also, prediction of monsoon rainfall a few weeks or months advance is necessary to prepare contingency plans for extreme events like droughts and floods.

In recent decades, artificial neural networks (ANNs) have shown great ability in modeling and forecasting nonlinear and non-stationary time series in hydrology and water resource engineering due to their innate nonlinear property and flexibility for modeling. Some of the advantages of ANNs are (ASCE, 2000a). (1) They are able to recognize the relation between the input and output variables without explicit physical considerations. (2) They work well even when the training sets contain noise and measurement errors. (3) They are able to adapt to solutions over time to compensate for changing circumstances and (4) they possess other inherent information processing characteristics and once trained are easy to use.

In the present study, a detailed analysis of rainfall on monthly and annual basis was carried out to study the rainfall pattern in Mumbai metropolitan region. Long term forecasting of rainfall was done using the Gumbel’s extreme value distribution model. The forecasted values were then validated using ANN. Neural network models were developed using the recursive multi-step neural network approach (RMSNN). Different combinations of past rainfall were used for rainfall forecasting in this study.

2. Study area

The city of Mumbai (Greater Mumbai) consists of two administrative districts: the Island City District and the Suburban District. It extends between 18° and 19.20° N and between 72° and 73° E (Figure 1). Geographically, Greater Mumbai is an island separated from the mainland by the narrow Thane Creek and a relatively wider Harbour Bay. Mumbai has a tropical wet and dry climate, and may be best described as moderate temperatures with high level of humidity. Its coastal and tropical location ensure moderate temperature throughout the year, average of 27.2 °C and annual average precipitation of 242.2 cm (95.35 inches). The temperature of Mumbai city is on average about 30°C in summer and 18°C in winter. Mumbai has 4 distinct seasons: Winter (December–Feb); summer (March–May); Monsoon (June–Sep) and Post-Monsoon (Oct–Dec). Every year, Mumbai experiences heavy rainfall during the monsoon time as it is in the windward side of Western Ghat. Between June and September, the south west monsoon rains splurge the city. Occasionally, premonsoon showers are received in May and north-east monsoon showering may occur in October and November. The maximum annual rainfall ever recorded was 3,452 mm (136 in) for 1954. The highest rainfall recorded in a single day was 944 mm (37 in) on 26 July 2005. The average total annual rainfall is 2,146.6 mm (85 in) for the Island City, and 2,457 mm (97 in) for the suburbs. Due to incessant rains and simultaneous high tidal waves, the overall risk to flooding is more in coastal urban areas. Mumbai is prone to flooding and experiences severe flooding almost every year, for example, 2004 and 2007. Mumbai witnessed flooding each summer. But in 2005, the city witnessed the worst flooding in its recorded history.

3. Methodology

Analysis of rainfall pattern and trend is done using the daily rainfall data from (1993-2013) collected from IMD regional office, Santacruz, Mumbai.
3.1 Rainfall pattern analysis

From the data obtained from IMD (Indian Meteorological Department, Santacruz), Rainfall pattern and behavior has been analyzed on annual and monthly basis and various graphs has been plotted for years (1994-2013). Trend analysis of a time series consists of the magnitude of trend and its statistical significance. Obviously, different workers have used different methodologies for trend detection. In general, the magnitude of trend in a time series is determined either using regression analysis (parametric test) or using Sen.’s estimator method (non-parametric method).

Both these methods assume a linear trend in the time series (Jain and Kumar, 2012), (Sen et al., 2013). Regression analysis is conducted with time as the independent variable and rainfall/temperature as the dependent variable. The regression analysis can be carried out directly on the time series or on the anomalies (i.e. deviation from mean).

A linear equation,

\[ y = mt + c \]

where ,
\[ c \] : the intercept trend
\[ m \] : the slope,

The linear trend value represented by the slope of the simple least-square regression line provided the rate of rise/fall in the variable.

3.2 Rainfall forecasting

As it is already discussed that rainfall forecasting is crucial for determination of drought or flood events and is helpful for developing early warning measures. In this study, rainfall
forecasting is carried out using Gumbel’s extreme value distribution model. Then the model is validated using artificial neural networks.

### 3.2.1 Gumbel’s model

Gumbel introduced the concept of extreme value distribution and developed a model for prediction of hydrologic event such as flood peaks, maximum rainfall, etc (Das, 2004). Gumbel found that the probability of occurrence of an extreme event, equal to or larger than a value is given by the following equation:

\[ P(X \geq x) = 1 - e^{-e^{(y)}} \]

Where,
- \( P \) - the probability of occurrence
- \( X \) - the event the hydrologic series
- \( x \) - the desired value of the event
- \( y \) - the reduced variate

Gumbel’s equations for field

The variate \( X_i \) can also be expressed as

\[ X_i = \bar{X} + K \sigma' \]

\( \sigma' \) - standard deviation of sample

\[ \sigma' = \sqrt{\frac{3\sigma^2}{N-1}} \]

\( K \) - frequency factor

\[ K = \frac{y_i - \bar{y}}{\bar{\sigma}} \]

\( y_i \) - the reduced variate for a given return period \( T \)

Where,
- \( y_n \) - the reduced mean as a function of sample size \( N \)
- \( \sigma \) - reduced standard deviation as a function of sample size \( N \)

### Return period

Return period or recurrence interval is the average interval of time within which any extreme event of given magnitude will be equalled or exceeded at least once

\[ T = \frac{N+4}{R} \]

Where,
- \( N \) - Total number of years of record
- \( R \) - Rank of observed rainfall values arranged in descending order.

### 3.2.2 Artificial neural network (ANN) model

Artificial Neural Network (ANN) methodology allows us to develop nonlinear systems receiving large quantities of inputs that development based on only instances of input-output relations. A major application field of ANNs is forecasting. ANN is especially best fit for discovering accurate solutions in an environment characterized by nonlinear, irrelevant, complex, partial or noisy data.
A Neural network essentially has three layers which are the input layer, the output layer and the hidden layer in between. Neural network is made up of interconnected neurons and each neuron has the relationship with neighboring neurons. The relationship between the neurons is measured by weight. The ANNs learning is the process of finding the optimal weight matrices in a systematic manner, in order to achieve the desired value of target outputs. The performance measure used to train ANNs is minimizing the network error function. Figure 3.5.3 represents a typical neural network with 4 input layers, 1 hidden layer and 1 output layer.

![Figure 2: ANN architecture with four input and one output for mean annual rainfall data.](image)

In order to perform the weather forecasting using neural network and comparing the performance of neural network model with different learning rates and by setting different number of neurons in hidden layer, backpropagation algorithm and supervised learning is used in this study.

The following Steps are followed in the process:
1. Data preprocessing.
2. Defining the ANN model.
3. Training of ANN.
4. Testing of data.

The relationship between the output \( y(t) \) and the inputs \( r(t-1), r(t-2)\) and \( r(t-3)\) can be represented by,

The following mathematical equation:

\[
y(t) = s_1 \left( \sum_{j=1}^{J} w_j s_2 \left( \sum_{i=1}^{I} w_i r(i-t) \right) \right)
\]

Where,
- \( y(t) \) is an output from the network,
r(t-i) is the inputs to network.
Wj and Wi are the connection weights.
S1 and S2 are activation function.

The most commonly used function is a logistic sigmoid function given by equation:

\[ s(y) = \frac{1}{1 + e^{-x}} \]

For validation of Gumbel’s distribution model, annual rainfall data from 1994-2014 is used in this study. Also the forecasted rainfall value for 2015 which is predicted using Gumbel’s model is also used as input data. Further rainfall moving average series i.e. RH1, RH2, RH3, RH4 etc were calculated and used as input layers in the neural network model to get the desired output.

4. Results and discussion

3.3 Rainfall pattern and trend analysis

As stated earlier, the rainfall pattern analysis was done using daily rainfall data for the period (1994-2013). The daily rainfall was then summed into weekly, monthly and yearly rainfall. Wide variation of annual rainfall was observed with the maximum value of 5112.5 mm and minimum of 1155.7 mm. This erratic rainfall distribution gives rise to extreme climatic conditions like floods and droughts in Mumbai region. A detailed statistical analysis of annual rainfall is shown in table 1. As the rainfall in Mumbai occurs mainly between June-September, a further detailed analysis of rainfall is done only for monsoon months. (Figure 2). From the figure it is clear that there is much variation of rainfall between months and also in the same month for the entire period under observation. Hence the rainfall trend was studied for individual months for the period under observation.

<table>
<thead>
<tr>
<th>Description</th>
<th>Santacruz (1994-2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of years of observation</td>
<td>19</td>
</tr>
<tr>
<td>Minimum rainfall (mm)</td>
<td>1155.7</td>
</tr>
<tr>
<td>Maximum rainfall (mm)</td>
<td>5112.5</td>
</tr>
<tr>
<td>Mean rainfall (mm)</td>
<td>2497.15</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>857.2626</td>
</tr>
<tr>
<td>Median</td>
<td>2180.7</td>
</tr>
<tr>
<td>Coefficient of variation</td>
<td>0.37</td>
</tr>
</tbody>
</table>

A positive rainfall trend is observed in June, July and August months (Figure 4). However, the rainfall trend was nearly linear in September month for all the years except for the year 2005 (Figure 5). Mumbai experienced an acute and disastrous flood hazard in this year because of this unusually high rainfall. The maximum daily rainfall in this year was 944 mm in the September month. To estimate the probability of occurrence of maximum one day rainfall in future, further analysis of rainfall was carried out.
Rainfall analysis and forecasting in Mumbai metropolitan region using artificial neural networks (ANN)
Sangita Mishra S et al.,

International Journal of Civil and Structural Engineering
Volume 6 Issue 2 2015

4.2 Rainfall forecasting

4.2.1 Gumbel’s Extreme value distribution model

Analysis of rainfall trend and pattern and analysis of probable maximum precipitation indicated erratic and uneven distribution of rainfall in Mumbai region over the years. So there is a need to estimate the quantity of rainfall that may occur after 1 year and after many years. Hence the rainfall amount is estimated for different return periods using Gumbel’s extreme value model. From the analysis it was observed that the rainfall amount may be 2219.67 mm.
in the year 2015. The analysis indicated a positive trend in the quantity of rainfall in future. (Table 2, Figure 6)

4.2.2 Artificial neural network (ANN) model

The neural network models were developed using the feed forward training with the standard back propagation algorithm to forecast rainfall values. Past combination of rainfall values were used as input parameters as explained earlier. The available data were split into three parts, 60% of the total data set was used for training, 20% for testing and 20% of the total dataset were used for validation.

<table>
<thead>
<tr>
<th>Return Period in years</th>
<th>Predicted Rainfall in mm ($X_t$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2219.67</td>
</tr>
<tr>
<td>3</td>
<td>2663.489</td>
</tr>
<tr>
<td>5</td>
<td>3157.497</td>
</tr>
<tr>
<td>10</td>
<td>3986.129</td>
</tr>
<tr>
<td>15</td>
<td>4127.02</td>
</tr>
<tr>
<td>20</td>
<td>4375.348</td>
</tr>
<tr>
<td>30</td>
<td>4874.64</td>
</tr>
<tr>
<td>50</td>
<td>5145.86</td>
</tr>
</tbody>
</table>

The forecasted rainfall using Gumbel’s model for 2 years return period from 2013 was found to be 2219.67mm which is being validated using ANN which is a standard model used by Indian meteorological department (IMD) with different number of trails as shown in table 3 and the results obtained are in the form of graphs having actual v/s forecasted plot which are also provided below. From all these trials it is found that the model architecture having 4
neurons in input, two hidden neurons and one output neuron (4-2-1) worked best with correlation coefficient value as 0.98 and regression coefficient value as 0.99 least absolute relative error of 0.04 (Figure 7,8).

**Figure 7:** Actual V/S Forecasted Rainfall for model architecture (4-2-1)

**Figure 8:** Scatter plot for actual V/S forecasted rainfall for model (4-2-1)

**Table 3:** Different model architectures

<table>
<thead>
<tr>
<th>Trial Number</th>
<th>Model Architecture</th>
<th>Regression Coefficient ($R^2$)</th>
<th>Correlation Coefficient (R)</th>
<th>Absolute Relative Error (ARE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4-2-1</td>
<td>0.99</td>
<td>0.98</td>
<td>0.04</td>
</tr>
</tbody>
</table>
5. Conclusion

Due to the high population growth and looming climate change, Mumbai is experiencing wide variation in spatial-temporal distribution of rainfall. Hence, rainfall trend and pattern analysis was done for Mumbai metropolitan region for a duration of 20 years (1994-2013). It was observed that aberrant variation in rainfall either causes flood or shortage of water supply. Hence rainfall forecasting is necessary to predict the rainfall well in advance to prepare contingency plans. This study incorporated Gumbel’s model to forecast rainfall for different return periods and then the forecasted rainfall is validated using ANN. Results obtained from detailed rainfall analysis such as rainfall estimates for different return periods using the above methods could be used for hydrologic modeling and for early warning systems.

6. References


<table>
<thead>
<tr>
<th></th>
<th>Neurons</th>
<th>Correlation</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4-10-1</td>
<td>0.94</td>
<td>0.97</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>4-4-1</td>
<td>0.93</td>
<td>0.97</td>
<td>0.08</td>
</tr>
<tr>
<td>4</td>
<td>3-8-1</td>
<td>0.95</td>
<td>0.90</td>
<td>0.08</td>
</tr>
<tr>
<td>5</td>
<td>3-5-1</td>
<td>0.93</td>
<td>0.87</td>
<td>0.09</td>
</tr>
<tr>
<td>6</td>
<td>3-3-1</td>
<td>0.94</td>
<td>0.84</td>
<td>0.08</td>
</tr>
</tbody>
</table>

(Note: 4-2-1 means 4 neurons in input, 2 neurons in hidden layer and 1 neuron in output layer)


