Rainfall intensity-duration-frequency (IDF) model using an artificial neural network approach

Resat Acar*, Samet Çelik and Serkan Senocak

Atatürk University, Engineering Faculty, Civil Engineering Department, 25240 Erzurum, Turkey

Received 05 February 2007; revised 08 January 2008; accepted 09 January 2008

Multilayer perceptron (MLP) artificial neural network (ANN) model was trained and used to model rainfall intensity-duration-frequency (IDF) relationship for short duration rainfall (SDR), in a catchment located in a semiarid climate in Turkey. Trained ANN model has been found more suitable to predict SDR than classical statistical model.

Keywords: Artificial neural network, Intensity-duration-frequency (IDF), Modelling, Nonlinear regression

Introduction

Artificial neural networks (ANNs) have been successfully used for rainfall prediction\(^1\)-\(^3\), rainfall intensity-duration-frequency process\(^4\)-\(^6\), and for predicting short duration rainfall values for 5, 10, 15, 30 and 60 minutes for each return periods (2, 5, 10, 25, 50 and 100 years). In addition, ANNs are applied for prediction of evaporation\(^7\), rainfall-runoff\(^8\)-\(^11\), flood disaster\(^12\), and for river flow time series prediction\(^13\). In these hydrological applications, a multilayer feed-forward backpropagation algorithm is used\(^14\).

This study presents rainfall intensity-duration-frequency (IDF) model using ANN for 0zmir province located in semiarid climate in Turkey.

Methodology

ANN Architecture

In ANN architecture (Fig. 1), each neuron \(j\) receives incoming signals from every neuron \(i\) in the previous layer. Associated with each incoming signal \((X_i)\) is a weight \((W_{ij})\). The effective incoming signal \((S_j)\) to neuron \(j\) is the weighted sum of all incoming signals and \(b_j\) is neuron threshold value.

\[
S_j = \sum_{i=1}^{n} X_i W_{ij} + b_j \quad \ldots(1)
\]

Effective incoming signal, \(S_j\), is passed through a nonlinear activation function (commonly used in the logistic sigmoid function) to produce outgoing signal \((y_j)\) of a neuron \(j\). This transfer function is continuously differentiable, monotonic, symmetric, bounded between 0 and 1\(^{15}\).

In this study, both statistical and graphical criteria were adopted to select the desired optimal network model. Statistical criteria consist of average squared of error (ASE), coefficient of determination \((R^2)\), and mean absolute relative error (MARE) given as

\[
ASE = \frac{\sum_{i=1}^{N} (I_{t_i} - \hat{I}_{t_i})^2}{N} \quad \ldots(2)
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (I_{t_i} - \hat{I}_{t_i})^2}{\sum_{i=1}^{N} (I_{t_i} - \bar{I}_{t_i})^2} \quad \ldots(3)
\]

\[
MARE = \frac{\sum_{i=1}^{N} |\hat{I}_{t_i} - I_{t_i}| / I_{t_i}}{N} \quad \ldots(4)
\]

where \(I_{t_i}\), actual normalized rainfall intensity; \(\hat{I}_{t_i}\), predicted normalized rainfall intensity; \(\bar{I}_{t_i}\), mean of \(I_{t_i}\).
Fig. 1 — Architecture of neural network model

Symbols:
- $a^1 = f^1(W^1 p + b^1)$
- $a^2(I) = f^2(W^2 a^1 + b^2)$
- $n^1$
- $n^2$

Legends:
- $f^1$: Hidden activation function (Logsig)
- $f^2$: Output activation function (linear)
- $W^1$: Weight matrix in the first layer
- $W^2$: Weight matrix in the second layer
- $b^1$: Bias vector in the first layer
- $b^2$: Bias vector in the second layer
- $p$: Input vector
- $a^1$: Output vector in the first layer
- $a^2$: Output vector in the second layer
- $n^1$: Net input in the first layer
- $n^2$: Net input in the second layer
\( R^2 \) statistical measures linear correlation between actual and predicted rainfall values.

ASE and MARE are applied to quantify the error between observed and predicted values. Optimal value for \( R^2 \), ASE and MARE are 1.0, 0 and 0 respectively. Graphical performance indicator gives better results when the data pairs are closing to 45 degree line. Good superposition between desired and calculated rainfall values in the training and testing phases are obtained. Input variables as well as output variables are first normalized linearly in the range of 0 and 1. Normalization is done using Eq. (5)

\[
\bar{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{5}
\]

where \( \bar{X} \), standard value of input, \( X_{\min} \) and \( X_{\max} \), minimum and maximum of actual observed values; and \( X \), observed value. By standardizing variables and recasting them in dimensionless units, an arbitrary effect of similarity between variable units is removed.

### Study of Catchment and Database Formulation

Izmir, located in Western Anatolian Region of Turkey, receives only 700 mm/year. Flooding in the Western Anatolian Region is probably the most severe hydrometeorological hazards in Turkey. Thus climate forecasting is very important for this region due to large interannual variability in precipitation and flooding. Entire rainfall data for 1938-2000 is represented in short duration rainfall values and predicted values for 5, 10, 15, 30 and 60 minutes of duration (Fig. 2) for each return periods (\( T = 1 \) to 100 years) by using standard statistical distributions like Generalized Extreme Values (GEV), Gumbel, Normal, Two-parameter Lognormal, Three-parameter Lognormal, Gamma, Pearson type III and Log-pearson type III distributions (Appendix A).

\( \chi^2 \) goodness-of-fit test was used to choose the best statistical distribution among all distributions. In ANN model, input neurons are represented by rainfall duration

![Fig. 2 — Comparison between actual and ANN predicted rainfall values](image)
and return period, and output neurons by expected intensities of rainfall value.

**Results and Discussion**

In this study, 20 years of data (for 5, 10, 15, 30 and 60 min) were used for testing ANN model, while remaining data of 59 years was used for model training/verification. Training phase of ANN model was terminated when ASE for testing databases was minimal. ANN was trained at 700 iterations with 4 hidden nodes. The comparison between predicted and actual rainfall values at training and testing phases showed excellent agreement with $R^2$ as 0.99896 and 0.99939 respectively (Table 1). Statistical parameters of predicted and actual values of rainfall for the entire database are practically identical (Table 1). To evaluate ANN performance, nonlinear regression (NLR) technique was applied with the same data sets used in ANN model. Apparently, ANN approach predicted effectively better than NLR ($R^2$, 0.99682).

**Conclusions**

ANN has been found capable to model rainfall intensity-duration-frequency (IDF) relationship for short duration rainfall (SDR) in the semiarid regions. ANN method has been found more suitable to predict SDR than classical NLR model.

**References**

Appendix A

<table>
<thead>
<tr>
<th>Statistical distributions</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized extreme value (GEV)</td>
<td>[ f(x) = \frac{1}{\alpha} \left[ 1 - k \left( \frac{x-u}{\alpha} \right) \right]^{1/k - 1} - \left[ 1 - k \left( \frac{x-u}{\alpha} \right) \right]^{1/k} ]</td>
</tr>
<tr>
<td>Extreme value type I (Gumbel)</td>
<td>[ f(x) = \frac{1}{\alpha} \exp \left[ - \frac{x-u}{\alpha} \right] - \exp \left[ - \frac{x-u}{\alpha} \right] ]</td>
</tr>
<tr>
<td>Normal</td>
<td>[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} ]</td>
</tr>
<tr>
<td>Two-parameter lognormal</td>
<td>[ f(x) = \frac{1}{x \sigma \sqrt{2\pi}} \exp \left[ - \frac{(\ln x - \mu_y)^2}{2\sigma_y^2} \right] ]</td>
</tr>
<tr>
<td>Three-parameter lognormal</td>
<td>[ f(x) = \frac{1}{(x-a) \sigma \sqrt{2\pi}} \exp \left[ - \frac{(\ln (x-a) - \mu_y)^2}{2\sigma_y^2} \right] ]</td>
</tr>
<tr>
<td>Two-parameter Gamma</td>
<td>[ f(x) = \frac{1}{\alpha^\beta \Gamma(\beta)} x^{\beta-1} e^{-\frac{x}{\alpha}} ]</td>
</tr>
<tr>
<td>Pearson III</td>
<td>[ f(x) = \frac{1}{\alpha \Gamma(\beta)} \left( \frac{x-\gamma}{\alpha} \right)^{\beta-1} e^{\frac{x-\gamma}{\alpha}} ]</td>
</tr>
<tr>
<td>Log-Pearson III</td>
<td>[ f(x) = \frac{1}{\alpha \cdot \Gamma(\beta)} \left[ \frac{\ln(x) - \gamma}{\alpha} \right]^{\beta-1} e^{\frac{\ln(x) - \gamma}{\alpha}} ]</td>
</tr>
</tbody>
</table>