Forecasting Droughts using Artificial Neural Networks

Gayan Illeperuma and Upul Sonnadara
Department of Physics, University of Colombo, Colombo 3, Sri Lanka

Abstract
The use of artificial neural networks as a tool to forecast droughts in Sri Lanka is presented. Predictions were made using the Standardized Precipitation Index (SPI) as the drought monitoring index. Monthly rainfall recorded at 13 climatological stations covering both the wet and dry zones over a long time period have been used as the input to train and test the neural networks. The analysis covers rainfall data recorded over the time span 1870 to 1980.

The SPI was computed by fitting a probability density function to the frequency distribution of the monthly precipitation records of each station. The developed neural network model was tested for SPIs with time windows of 1-6 months. For SPI-3 computed with a 3 month time window, the average correlation coefficient was found to be 0.90 with the lowest being 0.84 for Nuwara Eliya and highest being 0.94 for Batticaloa and Jaffna.

In general, the accuracy of the predictions was higher for the stations in the dry zone compared to the stations in the wet zone. The model predictions were superior for the period from May to July (which is the first part of the South-West Monsoon season) compared to the rest of the year. The accuracy of the predictions increased with the length of the window used in computing the SPI values. The results of this work shows that neural network models trained on SPI can be used to forecast water scarcity.

Introduction
The ability to predict extreme weather events including droughts is immensely important in order to mitigate their effects. Between 1967 and 1991, droughts affected 50% of the 2.8 billion people who suffered from natural disasters and of the 3.5 million people who lost their lives due to natural disasters 35% were due to droughts. Over 50% of the world’s most populated areas are highly vulnerable to droughts. Droughts trigger a complex set of impacts that range across many sectors of the economy and reaches well beyond the physical area affected by them.

Both statistical and neural network models based on ground level parameters such as surface rainfall, crop condition or water availability are often used to forecast droughts. These models are dependent on the geographical location and cannot be readily used in a different location. Unlike statistical models, neural network models do not require one to understand the underlying relationships between input parameters. Even when the input parameters are changed, the model can be simply retrained without any redesigning. The work done by Cannon et. al. clearly illustrate the superiority of artificial neural networks over linear stochastic methods to forecast droughts. Thus, in this work, a model based on artificial neural networks was developed to forecast droughts in Sri Lanka.

In the past, many different network architectures were tested by number of researchers to predict climate parameters. When developing a neural network model for forecasting, the network is developed through three stages: (i) The training stage where the network is trained to predict future data, based on past and present data. (ii) The testing stage where the network is tested to stop training, to keep on training, or to change the architecture. (iii) The evaluation stage where the network ceases training and is used to forecast future data and to calculate the accuracy or forecasting ability. In the

* Corresponding author: upul@phys.cmb.ac.lk
present study, a multilayer feed forward back propagation neural network was realized. Instead of rainfall amounts, the network was trained to predict a standard drought monitoring index.

Most of the recent drought prediction models which are based on statistical or artificial neural networks use the Standardized Precipitation Index (SPI) as the drought monitoring index\textsuperscript{1,2,9}. Using SPI to predict droughts has increased the ability to compare results obtained in different regions of the world. The work done by Mishra et. al. and Hayes et. al. demonstrate the ability of predicting SPI values using artificial neural networks\textsuperscript{1,9}. For a detailed review of popular drought monitoring indices and comparison between them (SPI – Standard Precipitation Index, EDI - effective drought index, DI - deciles index, PN - percent of normal and CZI – China Z index) the reader is referred to the work done by Morid et. al.\textsuperscript{10}.

**Experimental**

The rainfall data used in this work was obtained from the Meteorology Department of Sri Lanka. The data sample consists of monthly rainfall data recorded at 13 climatological stations covering both the wet and dry zones. For each station, over 100 years of data during the period from 1870 to 1980 was utilized as the input to train and test the neural network.

The SPI is computed by fitting a probability density function to the frequency distribution of precipitation summed over the time scale of interest. This is performed separately for each month. Basically SPI is a centre zeroed index. The most common way to calculate the index is to fit a statistical distribution (Gamma distribution) and transform it to the standard normal distribution\textsuperscript{9}. The gamma distribution is defined by its frequency or probability density function defined as

\[
g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad \text{for } x > 0
\]

where \( \alpha \) is a shape factor, \( \beta \) is a scale factor and \( x \) is the amount of precipitation. \( \Gamma(\alpha) \) is the gamma function defined as

\[
\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} \, dy
\]

Fitting the distribution to the data requires \( \alpha \) and \( \beta \) to be estimated. The cumulative probability distribution is then transformed to the standard normal distribution with a mean zero and a variance of one, which is the SPI\textsuperscript{9}.

As shown in Table 1, large positive SPI values indicate extremely wet situations while large negative SPI values indicate extremely dry situations. Since SPI is normalized unlike direct rainfall amounts, it can be used to track climate changes in any terrain, irrespective of its average rainfall. This gives the advantage of comparing climate changes in different places and time scales. Another advantage of using SPI is its limited output range. A standard normal distribution has little contribution outside \( \pm3 \) which is ideal to use with neural networks.

<table>
<thead>
<tr>
<th>SPI values</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>+2.00 and above</td>
<td>Extremely wet</td>
</tr>
<tr>
<td>+1.50 to +1.99</td>
<td>Very wet</td>
</tr>
<tr>
<td>+1.00 to +1.49</td>
<td>Moderately wet</td>
</tr>
<tr>
<td>-0.99 to +0.99</td>
<td>Near normal</td>
</tr>
<tr>
<td>-1.00 to –1.49</td>
<td>Moderately dry</td>
</tr>
<tr>
<td>-1.50 to –1.99</td>
<td>Very dry</td>
</tr>
<tr>
<td>-2.00 and less</td>
<td>Extremely dry</td>
</tr>
</tbody>
</table>
Droughts are hard to detect and describe. Unlike rainfall they are not easily measured. Thus, SPI was used only as a drought quantification parameter\(^1\). The first step was to convert all rainfall data into SPI values. The data were divided into two parts, the data from 1870 to 1950 was used as the training set and the data from 1951 to 1980 was used as the testing set to check the forecasting accuracy. A feed forward artificial neural network with back propagation for weight adjustments was used in this work. The developed network consisted of three layers, namely, input, hidden and output with 30, 8 and 1 neurons respectively. Training epoch of 500, a minimum gradient of \(10^{-6}\) and a momentum of 0.9 was set. Batch training was used with Levenberg-Marquardt optimization weight updating (mathworks-2008). The input vector consists of the previous 30 SPI values. Transfer functions were pure linear and tan sigmoid.

**Figure 1: Basic architecture of ANN developed for predicting droughts**

### Results and discussion

In this work, drought conditions were predicted by attempting to forecast future SPI values. SPI values were calculated for a range of time windows varying from 1 month to 6 months. For example, to calculate SPI-3 for March; rainfall of January, February and March was used. The evaluation of the accuracy of model forecasts was carried out using root mean square error (RMSE), mean average error (MAE), percentage error (PE) and correlation coefficient (R). The estimated accuracies by comparing actual values and predicted values for the Colombo station for different time windows (SPI-1 to SPI-6) are shown in table 2. Similar results were observed for all other stations.

<table>
<thead>
<tr>
<th>SPI</th>
<th>Root Mean Square Error (RMSE)</th>
<th>Mean Average Error (MAE)</th>
<th>Percentage Error (PE)</th>
<th>Correlation Coefficient (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPI-6</td>
<td>0.005</td>
<td>+0.002</td>
<td>0.029</td>
<td>0.999</td>
</tr>
<tr>
<td>SPI-4</td>
<td>0.028</td>
<td>-0.001</td>
<td>0.071</td>
<td>0.999</td>
</tr>
<tr>
<td>SPI-3</td>
<td>0.074</td>
<td>-0.007</td>
<td>0.634</td>
<td>0.994</td>
</tr>
<tr>
<td>SPI-2</td>
<td>0.193</td>
<td>-0.026</td>
<td>0.615</td>
<td>0.963</td>
</tr>
<tr>
<td>SPI-1</td>
<td>0.444</td>
<td>-0.037</td>
<td>1.349</td>
<td>0.823</td>
</tr>
</tbody>
</table>

It can be seen that all SPI predictions produce very high accuracies. It is also seen that when the time window is increased, the correlation between the model prediction and actual values also increases. This observation can be explained by the way the SPI time series is calculated. Unlike
rainfall series, SPI follows the standard normal distribution. This conversion removes any sudden peaks leaving a smooth slowly varying curve which is easier to predict using neural network models. Increasing accuracy for longer time windows is also expected since averaging over many past values tends to make the variation in the SPI time series quite smooth.

The observed and one month ahead forecasted SPI values for three time windows are shown in Figure 2. The neural network model works extremely well with the SPI-6. The deviations between observed and predicted values are somewhat apparent in SPI-3. Figure 2 shows that although the model could predict the general trend, predictions deviate at extreme values for SPI-1.
As discussed, the neural network model was successful in predicting 1 month ahead lead time for SPI. The model could be used to predict further into the future by multi-step approach by using the predicted values as inputs. When this was done for SPI-3, after a lead time of 4-5 months, the accuracy becomes quite poor.

In Figure 3, the geographic variation of model predictions is shown. Compared to stations in the wet zone (average R=0.87), model predictions are superior for the stations in the dry zone (average R=0.93). Badulla, which is in the eastern slope of the high lands shows an average performance while Nuwara Eliya which is on the western slopes of the highland shows the poorest performance. From all stations:
stations, Baticaloa and Jaffna show the best performance with R=0.94. Figure 3 also shows the correlation plots for SPI-6 and SPI-1 for the Baticaloa station.

![Figure 3: Correlation plots for SPI-6 and SPI-1 for the Baticaloa station.](image)

In Figure 4, annual variation of the predictions for SPI-3 for the Galle station which received rainfall predominantly from the South-West monsoon is shown. High prediction accuracy is seen for the early part of the South-West monsoon period (May – July). For SPI-6, accuracy was high throughout the year. A similar trend was observed for most of the stations.

**Conclusions**

In this work, an artificial neural network model was used to make 1 month lead time predictions for the Standard Precipitation Index (SPI) which is interpreted as a drought quantification parameter. Monthly rainfall recorded at 13 different meteorological observatories for 110 years (1870 – 1980) was used as the data set to train and to the test neural network model.

The SPI was computed by fitting a probability density function to the frequency distribution of the monthly precipitation records of each station. The model shows superior predictions for SPI-6 compared to SPI-1. For SPI values computed with a 3 month time window, the average correlation coefficient was found to be 0.90. As expected, when the lead time was increased, model predictions were degraded at each time step into the future.

The accuracy of the predictions was higher for the stations in the dry zone (R=0.93) compared to the stations in the wet zone (R=0.87). The model predictions were superior for the monsoon season compared to the inter-monsoon seasons. The highest accuracy was seen during the period from May to July (which is the first part of the South-West Monsoon season). The developed neural network model can be a very useful tool for water resource planners to take necessary steps in advance when there is water scarcities which may eventually develop into drought conditions.

**Acknowledgements:** Financial assistance by the National Research Council (grant number NRC 06-18) is acknowledged.

**References**