Application of Artificial Neural Networks to the Simulation of Climate Elements, Drought Forecast by Two Indicators of SPI and PNPI, and Mapping of Drought Intensity; Case Study of Khorasan Razavi

Mohammad Hossein Jahangir*, Mahnaz Abolghasemi, Seyyedeh Mahsa Mousavi Reineh

Department of Renewable Energies and Environment, Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran.

PAPER INFO

Paper history:
Received 01 January 2020
Accepted in revised form 25 April 2020

Keywords:
Forecasting of Drought Intensity
Artificial Neural Network
Simulation
Multi-Layer Perceptron
Levenberg-Marquardt
Razavi Khorasan

ABSTRACT

Drought is considered as a destructive disaster that can have irreversible effects on different aspects of life. In this study, artificial neural network was used as a powerful means of modeling nonlinear and indefinite processes in order to simulate drought intensities at 7 synoptic stations of Khorasan Razavi from more than 35 years ago up to the year 2014. Input data were the calculations of the two indicators of PNPI and SPI by DIC software, and the output layer (drought intensity) was taken to the Matlab software and employed as the teaching data (from 25 years), experiment (from 5 years), and validation (from another 5 years). The 3-9-1 structure of the network of layers had the maximum accuracy with the error rate of less than 2 % and high correlation (more than 90 %). After trial and error for each station through sigmoid stimulation function in the Perceptron network, it was observed that the stations of Mashhad and Quchan had the minimum error and the maximum error was related to the station of Neyshabur. The results of comparisons and observations showed that the artificial neural network had high efficiency in simulation of the data. The obtained correlation amount of 0.999 for the base station represented the small error of the model in prediction. Drought forecasting was performed in this study by the trained algorithm in the artificial neural network without using the observation data. The results showed that rainfall, temperature, and speed models had a positive role in forecasting the provinces that would experience drought. Due to its lower amount of error, SPI indicator was selected for mapping, the findings of which showed that the highest drought intensity belonged to the near normal to normal wet lands.

1. INTRODUCTION

Forecasting the atmospheric processes in the planning and management of water resources is of high importance, especially in arid and semi-arid areas. In recent decades, most researchers have adopted multivariate regression and geostatistical models such as auto-regressive moving average and auto-regressive integrated moving average in the prediction and modeling of meteorological and hydrologic processes as well as characterization of flood and drought [1-3]. Decrease in atmospheric precipitation is the main reason of agricultural and hydrological drought, which in turn leads to increase in the evaporation of surface waters [4]. Most of the applied models in the literature enter the considered parameters into decision-making processes in a linear form, and they often cannot analyze complex climate and hydrology issues properly. Hence, it is necessary to introduce more efficient models to predict non-linear and complex phenomena [5]. For this reason, experts and scientists of hydrology and other related fields have intended to develop appropriate models for predicting when different hydrological events occur. Emergence of robust theories like phase algorithm and nervous network has brought a revolution in analyzing the behavior of dynamic systems in different scientific areas associated with the water issues. These methods are nowadays used for the prediction of meteorological parameters to attenuate the possibility of human error, increase precision, and reduce the limitations of massive amounts of data and computations. One of the common methods for the prediction of meteorological parameters is artificial neural networks, which are powerful and flexible tools independent of the dynamic model system. It is an intelligent method that has improved the recognition and prediction of important parameters with widespread application to different forecasting fields comprising complex processes. The major superiorities of the method are the ability to learn, extensible distribution of information, parallel processing, and durability [6]. Also, artificial neural networks have great modelling capabilities that can be applied to the forecasting of climate and hydrological issues, especially when the adopted network is able to extract the governing law of data, even confusing data [7].

Many research studies have recently been conducted dealing with the prediction of river flow, rainfall-runoff modeling, and estimation of hydrologic parameters using neural networks and, in most cases, they have been proven effective in predicting and stimulating hydrologic parameters [8-11]. Some studies have aimed to forecast drought cycle and simulate climate elements using dynamic structures and develop models through artificial neural networks, e.g., the research conducted by Krisbo & Mooma (2003) in Spain [5] and those on Kanchos river basin in Mexico [12], Kanabatis river basin in the east of India and Kamilo (2008), and Altotachos river basin. In Iran, different research studies have also addressed draught by adopting artificial neural networks and large-scale climate signals [13]. The obtained results of the neural models have shown that during the warm-phase
ENSO and negative-phase NAO, the country experiences wet conditions and, during the cold-phase ENSO and positive NAO, drought occurs. Afkhami et al. [14] intended to forecast drought in the synopsis station of Yazd using two regression models of RN and TLRN from dynamic structures of artificial neural network. The results of their study showed that TLRN was more efficient than other models in the simulation of the drought phenomenon. Farahmand et al. [15] evaluated the annual amount of rainfall using meteorological parameters with the aid of the artificial neural network. The results showed that artificial neural networks had high efficiency in forecasting the annual rainfall amount with acceptable precision. Also, the best perceptron model was three-layer neural network, which comprised three neurons in the medial layer. Ultimately, effective parameters for forecasting rainfall in the next month were determined heating degree day with the base of 21°C and average maximum moisture. In another study [16], artificial neural network was utilized for the simulation of climate elements and forecasting of the drought cycles in 20 meteorological stations of Esfahan. The results showed that in the patterns, maximum temperature, water flow, and rainfall had a positive role in forecasting drought in Esfahan province, and applying artificial neural networks could lead to the forecast of drought cycle in the province with above 95 % of precision. Nasri [17] utilized artificial neural network in risk forecast and management and found it a robust model in the prediction of flood and drought. Rezaeian-Zadeh & Tabari [18] showed precision of this method with the minimum amount of error in forecasting the conditions of different climate zones. Afkhami et al. [19] applied artificial neural networks to 13 climatology stations and one synoptic station in Yazd zone. The obtained results indicated significant flexibility of the artificial neural networks, which made the method an appropriate tool for modeling with log data.

In this research, first, precision of the artificial neural network in the simulation and forecasting of drought in the climate of Khorasan Razavi is evaluated and compared with other results. Then, after ensuring enough precision and low error of the method in simulation, drought forecasting will be done for the years lacking statistics and with regard to some climate parameters. Finally, mapping will be carried out for the province by selecting the indicator with the highest precision.

2. MATERIALS AND METHODS

Artificial neural network was adopted in order to simulate the climate elements and determine draught intensity by using SPI and PNPI indicators in synoptic stations of Khorasan Razavi and meteorological information from 1980 to 2014 (35 years) on annual time scale. After determining input data (in three layers of average annual rainfall, average annual temperature, and average annual wind speed), target data (obtained by the calculation of the two indicators of PNPI and SPI), and output layer (draught intensity) by the software, 75 % of the information was used as the teaching data, 15 % for testing, and the rest 15 % for validation. A summary of the neural network structure and neurons is given in Table 2 in terms of RMSE (correlation coefficient). Finally, mapping of the zones was performed in terms of the selected indicators.

2.1. Studied zone

Khorasan Razavi province with 14864/118 km² of area is the fourth largest province in Iran located in the north east of the country. The population of the province is 5999529 and it is located at 36.321247°N 59.532639°E. The climate of this province is arid and semi-arid. The highest point of this province is located on Mount Binalud in the northern areas of Neyshabur with 3211 meters of elevation, and the lowest point of the province is Sarakhs with 300 meters of elevation on the borders of Iran and Turkmenistan. The average annual rainfall of Khorasan Razavi is 102 millimeters [20]. The position of the province and information about some of the stations are brought in Table 1 and Figure 1.

![Figure 1. Position of Khorasan Razavi province in Iran and some stations.](image-url)
2.2. The mathematical theory of neural networks

The mathematical theory of neural networks is in fact a simplified model of information processing pattern in human brain. It enables preforming processes and finding the favorite non-linear combination for the relation between input and output of each system. Also, through learning process, this network is trained with the available data to perform the future forecasting. Neurons in the neural network are very simplified representations of biological neurons with lower abilities (Figure 2). In fact, artificial neural network is a mathematical model with the ability of modeling and development of non-linear relations for interpolation.

![Figure 2. A schematic of the Multi-Layer Perceptron (MLP) neural network [Source: MSc thesis of the author, 2016].](image)

Generally, every neural network consists of three layers. The first layer, namely the input layer, includes a few neurons, which in this study is devoted to average annual rainfall, average annual temperature, and average annual wind speed. The hidden layers include a few variable neurons, the optimal number of which is determined by examination and repetition for making the minimum error in terms of, e.g., RME. Efficiency of the neural network highly depends on appropriate selection of the number of neurons in the hidden layer. The last layer, i.e., output layer, is used to properly output the functions for increasing the speed of the network.

2.3. Architecture of the multi-layer neural network

In many extremely complicated mathematical problems that require complicated non-linear equations, a multilayer perceptron network can be simply used by defining weights and appropriate functions. Depending on the type of problem in neurons, various activation functions can be used. Demonstrating proximate hypothesis, Hernick and others suggested that a novel neural network with a sigmoidal hidden layer and a linear output layer could predict each complicated mapping with a proximate degree of accuracy. This hypothesis involves the minimum number of hidden layers and, hence, it significantly decreases complexity of the model [21]. In recent years, several general raters have been proposed, e.g., Multi-Layered Perceptron (MLP) in which enough neurons can be considered. MLP is able to estimate an appropriate proximate for each function and use the available information in large numeral sets. It should be noted that, in general, one cannot suggest any appropriate number of neurons for the medial layer and this choice should be made by trial and error. In this research, three-layer network perceptron is considered by the training algorithm after emission of error.

2.4. Design stages and running of the neural network

To run the neural network, the following steps should be taken:

- Collecting and pre-processing the required data for the intended neural network.
- Determination of the appropriate type and structure of neural network and developing a network with high efficiency.
- Testing the network by a portion of the collected data (stage of training).
- Examination of the trained network by the rest of the data (examination level).
- Saving the network if the result of examination is acceptable; otherwise, repeating steps 2 to 4.

General characteristics of the studied neural network in forecasting the variation of draught are given in the following. Since certain laws are not available, several structures are required for training and designing the neural network in this study. Choosing data for meta-measurement different from those for training or learning of the artificial neural network, simultaneously, can help us take steps more confidently than when the developed models are not pre-trained [22, 23]. In this study, Multi-Layer Perceptron (MLP) network is used with Back Propagation (BP) and Levenberg-Marquardt (LM) learning tech. The effects of changes in parameters are surveyed in many repetitions and the coefficients with better results for training the network in the modelling are introduced. In order to simulate the climate elements and determine draught intensity, the two indicators of PNPI and SPI are employed.

<table>
<thead>
<tr>
<th>Station</th>
<th>Geographical location</th>
<th>Elevation (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mashhad</td>
<td>16°N 36°E 59°38′</td>
<td>999.2</td>
</tr>
<tr>
<td>Sarakhs</td>
<td>N 32° 36°E 61°10′</td>
<td>235.0</td>
</tr>
<tr>
<td>Quchan</td>
<td>N 4° 37°E 30° 58′</td>
<td>1287.0</td>
</tr>
<tr>
<td>Sabzevar</td>
<td>N 12° 36°E 39° 57′</td>
<td>972.0</td>
</tr>
<tr>
<td>Turbat-jam</td>
<td>N 16° 35°E 13° 59′</td>
<td>1450.0</td>
</tr>
<tr>
<td>Kashmir</td>
<td>N 12° 35°E 28° 58′</td>
<td>1109.7</td>
</tr>
<tr>
<td>Neyshabur</td>
<td>N 16° 36°E 48° 58′</td>
<td>1213.0</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of the studied stations in the province [Source: weather.com].

21
For the synoptic situations of Khorasan Razavi province, meteorological information from 1980 to 2014 (35 years) was employed on an annual time scale. After determining input data (in three layers of average annual rainfall, average annual temperature, and average annual wind speed), target data (obtained by calculating the two indicators of PNPI and SPI through DIC software), and output layer (draught intensity) and normalizing them, information from 25 years was utilized for training, 5 years for examination, and the rest 5 years for validation. Finally, mapping of the present situation was carried out for zone evaluation measurements of the neural network.

2.5. Neural network performance evaluation criteria

To evaluate and compare the obtained results by the applied models and methods in this research, two measures, namely root mean square error in Equation (1) and correlation coefficient (random error $(R^2)$) in Equation (2), are employed as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\text{obv}_i - \text{pr}_i)^2}{N}}$$ (1)

$$R^2 = \frac{\sum_{i=1}^{N} (\text{obv}_i - \bar{\text{obv}})^2}{\sum_{i=1}^{N} (\text{pr}_i - \text{obv})^2}$$ (2)

In the above relations: $\text{obv}$ is the observed values, $\bar{\text{obv}}$ is average observed value, $\text{pr}$ is pre-estimated values of network models and $N$ shows the number of total data of each stage of training data and examination data. The closer the value of RMSE to 0 and the numeral value to 1, the closer the observed and predicted data and the more accurate the answers at each stage. These measurements are made for important patterns at the test and training stages of the neural network.

3. RESULTS AND DISCUSSION

3.1. Identifying important patterns in training and testing

To forecast the course of draught intensity through the artificial network, statistics from synoptic stations of Khorasan Razavi province in a 35-year period (1980-2014) were used. The diagrams obtained by PNPI and SPI methods indicate that draught intensity has experienced a significant increase during the included years. This increasing trend is also predicted to continue in the following years (Figures 3 & 4).

Figures 5-8 display the obtained results by the neural network model using the two indicators of PNPI and SPI and compare them with real data (observed data) for different patterns from Mashhad station, proving high precision of the model and network (more than 99 %). Figure 5 shows the correlation value for PNPI indicator with four classes of data (training, testing, validation, and total) from Mashhad station.

The diagrams given in the below figures show high accuracy of the results of ANN.
Figure 5. The correlation coefficient for PNPI index with four categories of data (training, testing, validation, and total) from Mashhad station.

Figure 6. Comparison between PNPI and the observed results at Mashhad station.

Figure 7. The correlation coefficient for SPI index with four categories of data (training, testing, validation, and total) from Mashhad station.
3.2. Determination of the appropriate indicator for forecasting draught intensity

One of the major concerns of draught studies is identifying the proper indicator that can evaluate intensity, stability, and magnitude of draught in a zone. Each indicator certainly uses different factors on each time scale. An appropriate draught indicator should be inclusive and reflect the short-term conditions of drought so that it reacts to different types of draught, e.g., agricultural, meteorological, and hydrological. Also, it should not be limited to any specific season and must be able to specify the draught regardless of whether it is, e.g., summer or winter; the same can be said about whether the indicator is dealing with a wet-climate region or a dry zone. In this research, RMSE and R² are employed as measures to forecast draught intensity. The closer the RMSE value to zero and the value of R² to 1, the better the indicator for forecasting draught. Both indicators, i.e., SPI and PNPI, have very good correlation coefficients (R²=0.99) and error of the indicators in different stations were from 0.0070 to 0.31, representing high precision of both indicators. However, since the error value for SPI is less than that for PNPI, SPI is used for mapping and other calculations.

3.3. Mapping

For mapping, the value of SPI was calculated for the available systems in DIC software and the required slope was developed in GIS software. Then, the map for the distribution of draught on surface zone was prepared by the Kriging method (k), which is the best known and most widespread way of modeling draught. This way of interpolation is appropriate for the data collected through a locally defined procedure. As shown in Figure 9, the highest intensity of draught belongs to the nearly normal to normal wet levels.

To predict draught intensity in Khorasan Razavi province, artificial neural network was employed using the two indicators of normal rainfall (SPI) and percentage of normal rainfall (PNPI). After collecting data from synoptic stations, they were statistically examined and normalized for entering into the artificial neural network. At the next stage, the appropriate structure for the neural network, stimulation function, number of neurons, and the number of hidden, input, and output layers were selected. Then, three average annual temperatures and three average wind speeds were considered for the prediction of draught intensity by the given two indicators. High correlation between the observed data and the predicted ones was observed. A summary of the considered structure for the neural network, neurons, and other information related to different stations as well as the general pattern of the stages (training, testing, and validation) is given in Table 2 in terms of error (RMSE) and correlation coefficient (R²).
According to the obtained results for correlation coefficient and error with three input layers for Mashhad station and two for others, one output layer and nine hidden layers provided the maximum precision and minimum error (less than 2%) with high correlation (more than 4%). Evaluation of each station with sigmoidal stimulation function in the perceptron network shows that Mashhad and Quchan stations have the minimum error and Neyshabur station the highest error. Figures 6 and 8 show the simulated results of the model for the two indicators: (a) the comparison between SPI and the observed results at Mashhad station and also (b) the comparison between PNPI and the observed results at Mashhad station.

The curves and calculations in the drawn diagrams prove that artificial neural network has high efficiency in the simulation of the studied data. The correlation degree of 0.999 for the base station (Mashhad) indicates minimum error of the model in forecasting. Since the number of the studied stations was large, for the sake of conciseness, the results for correlation and the comparisons are given only for Mashhad station. Information on the two indicators for different stations was large, for the sake of conciseness, the results for Mashhad station are brought in Table 2, and Figure 10 shows the total correlation pattern for the simulated and observed values at Mashhad station.

After determining the values of error and correlation coefficient, the next step is validation of the simulation and another words, simulated results are well near real observations. The next step requires predicting values of the two indicators for the following year in a limited period by means of the neural network using the algorithm trained by the previous data. This forecast depends on the precision and error of the network for the previous data and as long as the obtained results represent a low amount of error, they are acceptable. MLP classifier makes use of the following algorithm for calculating the inputs receiving an individual knot:

$$net_j = \sum_i w_{ij} I_i$$

where $net_j$ is the input parameter that receives the individual neuron j, $W_{ij}$ shows the weights of neurons i and j, and $I_i$ stands for the output of neuron i belonging to the sender, input or hidden layer. The output value derived from neuron j is computed through Equation (4).

$$f(net_j) = (1 + \exp(-net_j))$$

The results of comparison between draught forecasting by the neural network and real draught statistics indicate no significant difference and imply that the amount of error is acceptable. Accordingly, artificial neural network can be considered as an efficient model for forecasting and simulation with high precision.

4. CONCLUSIONS

Natural disasters like flood and drought can lead to huge damages, e.g., agricultural, biological, etc., as well as social and economic effects, e.g., immigration of rural people from villages to the margins of the cities. Hence, the return period and intensity of such disasters should be identified to prevent the resulting damages and bad outcomes. In recent years, artificial neural network has been employed frequently to predict and model draught in different regions of the world. The obtained results in this study indicated that artificial neural network was more efficient than linear models in forecasting the cycle and intensity of drought. Correlation coefficient and error were investigated with three input layers for Mashhad station and two for others, one output layer, and nine hidden layers. Trial and error with sigmoidal stimulation pattern for the simulated and observed values at Mashhad station.

Table 2. Statistics for the general pattern of the model for the studied stations.

<table>
<thead>
<tr>
<th>Station</th>
<th>Record data</th>
<th>Error</th>
<th>The correlation coefficient for total data</th>
<th>Layer structure</th>
<th>Trial and error</th>
<th>Type of algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>SPI Index</td>
<td>PNPI Index</td>
<td>SPI Index</td>
<td>PNPI Index</td>
</tr>
<tr>
<td>Mashhad</td>
<td>35</td>
<td>0.00070</td>
<td>0.027</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
</tr>
<tr>
<td>Sarakhs</td>
<td>30</td>
<td>0.003</td>
<td>0.03</td>
<td>0.99</td>
<td>1</td>
<td>1-9-2</td>
</tr>
<tr>
<td>Quchan</td>
<td>30</td>
<td>0.0008</td>
<td>0.01</td>
<td>0.99</td>
<td>0.99</td>
<td>1-9-2</td>
</tr>
<tr>
<td>Sabzevar</td>
<td>35</td>
<td>0.0004</td>
<td>0.3</td>
<td>0.99</td>
<td>0.99</td>
<td>1-9-2</td>
</tr>
<tr>
<td>Turbat-jam</td>
<td>35</td>
<td>0.005</td>
<td>0.02</td>
<td>0.99</td>
<td>0.99</td>
<td>1-9-2</td>
</tr>
<tr>
<td>Kashmar</td>
<td>28</td>
<td>0.004</td>
<td>0.31</td>
<td>0.99</td>
<td>0.99</td>
<td>1-9-2</td>
</tr>
<tr>
<td>Neyshabur</td>
<td>24</td>
<td>0.06</td>
<td>0.001</td>
<td>0.99</td>
<td>1</td>
<td>1-9-2</td>
</tr>
</tbody>
</table>

The obtained results in this study indicated that artificial neural network was more efficient than linear models in forecasting the cycle and intensity of drought. Correlation coefficient and error were investigated with three input layers for Mashhad station and two for others, one output layer, and nine hidden layers. Trial and error with sigmoidal stimulation pattern for the simulated and observed values at Mashhad station.
function in perceptron network for each station showed that Mashhad and Quchan stations had the lowest error and Neyshabur station the highest error. Also, the findings of network training for the two indicators of PNPI and SPI by input data from 35 years (from 1980 to 2014) on average annual rainfall and average annual temperature were evaluated. The results of evaluation showed maximum precision of the model with an error amount less than 2% and high correlation (more than 90%). A correlation diagram was drawn between the real and forecasted values, which showed agreement to a great extent (more than 99%). Overall, artificial neural network performed efficiently in forecasting drought in Khorsan Razavi province. According to the findings of mapping through Kriging method by the SPI indicator, the maximum drought intensity belonged to the near normal to normal wet zones.

As the final conclusion of this study, it can be stated that artificial neural network can have a broad extent of applications, for both seasonal and short-term forecasting, to water resource management, biological studies, extraction, draught management, and climate change studies.

5. ACKNOWLEDGEMENT

The authors would like to thank the journal editor and Faculty of New Sciences and Technologies, University of Tehran.

REFERENCES


