Projection of harvestable water from air humidity using artificial neural network (Case study: Chabahar Port)

Previsione della acqua raccoglibile dall’umidità dell’aria attraverso l’uso della Rete Neurale Artificiale (Caso di studio: Chabahar Port)

Chakavak Khajeh Amiri Khaledi

Department of Oceanography, Faculty of Marine Science, Chabahar Maritime University, Chabahar, Iran
E-mail address: C.Khajehamiri@cmu.ac.ir

Abstract. The optimum use of existing water resources as well as the efforts to achieve new water resources have been considered as two major solutions to the relative resolution of water scarcity. Through utilization of the information and meteorological data, it is possible to identify areas with potentials for water harvesting from air humidity. It also allows for collecting and converting them into fresh water using simple physical laws. Due to lack of atmospheric precipitations or inappropriate distribution of precipitations in Chabahar, located in the south of Sistan and Baluchistan province, Iran, water is a limiting factor for agricultural activities and even for the entire life. In this study, water was harvested from air humidity using a screen collector with dimensions of 1×1 m. The magnitude of water harvesting was monitored daily for a period of 365 days. The results revealed that approximately 20% of the water available in the air could be extracted in this area. Then, monthly meteorological data from Chabahar synoptic station between 1990 and 2011 was used to predict the harvestable water for the upcoming year using an artificial neural network. After determining the effective input variables in predicting the amount of harvestable water, the modeling was performed using Multi-Layer Perceptron Network (MLP) and General Feed Forward Network. The results indicated that the MLP network had a higher ability to predict the amount of harvestable water when compared to the GFF network (at the R² test stage it was 0.86 versus 0.44). The most suitable structure to predict harvestable water from the fog in Chabahar was the MLP Artificial Neural Network with the array of 12-1-25 and the Hyperbolic Tangent Stimulus Function with the Levenburg Marquette Training Law. Also, the values of the RMSE and MAE error rates were 2.19 and 1.81, respectively. Therefore, it is possible to predict the amount of harvestable water in the next 12 months which can be used in water resources management and productivity.

Keywords. Chabahar, air humidity, water extraction, prediction, neural network.
INTRODUCTION

Today, with the growth of population especially in developing countries, which are mostly located in arid and semi-arid regions, the need for water increased considerably. Iran is a country which is located in the arid belt zone of the world. Chabahar is the southernmost part of Sistan and Baluchistan province, where weather warming is its most significant climatic phenomenon due to its climatic characteristics and location in the ultra-tropical region. This region is one of the hottest and most humid regions of Iran, and its atmospheric precipitation is extremely little or it does not have good dispersal. There are no significant precipitations in this area for more than two thirds of the year, where most precipitations occur once or twice which is often as flood spring rains causing a considerable damage. However, given that this city is located near the Oman Sea, its relative humidity is largely high and even exceeds 85\% (IRIMO).

Therefore, due to the scarcity and inconsistency of the spatial and temporal distributions of precipitation, harvesting water from air humidity can be considered as a modern technology in the field of water resources. Large amounts of water can be collected via collectors for local and domestic uses, agriculture, or forestry (Davtalab et al., 2013). The lower cost of water harvesting from the fog compared to other water supply methods, as well as its simple and accessible technology and the high quality of harvested water and sustainability of water resources for many years are main advantages of this new technology.

Given the general understanding of meteorological conditions across the globe along with recognition of site topography, many parts of the world such as North America, the Middle East, North Africa, China and India can have the potential to utilize water harvesting program from the fog (Sekar and Randhir, 2007). To the best of our knowledge, only few studies have been conducted to evaluate water harvesting from the fog in Iran. Mousavi-Baigi and Shabanazadeh (2008) reported that up to 40 L of water per day was harvested from four different collectors in highlands of Khorasan-Razavi province of Iran.

The prediction of water demand for urban, agricultural, and industrial uses is the main factor in planning and managing water resources. Today, the use of computer models and software has become commonplace for prediction, and managers can easily make decisions by entering available data and analyzing outputs. One of the software commonly used is artificial neural networks (ANNs). The rapid expansion of the use of these networks as an empirical and effective model in various sciences including meteorology and climatology, suggets the need to these valuable models. ANNs are an effective tool for modeling nonlinear systems. since these networks do not require a mathematical equation for complex phenomena of interest (Kumar et al., 2002).

Accordingly, due to water scarcity in Chabahar on the one hand and the presence of heat and humidity on the other hand, and for optimal use of new water resources, the purpose of this study was to assess the potential use of harvesting water from the air humidity and also predict the amount of harvestable water using an ANN in the Chabahar region for 12 months.

Parole chiave. Chabahar, umidità dell’aria, estrazione di acqua, previsione, rete neurale.
MATERIALS AND METHODS

Area of study

Sistan and Baluchestan province is the largest province of Iran. The Oman Sea in south of Sistan and Baluchestan covers all southern borders of the province. This huge moisture source can affect most part of the province’s southern regions, especially Chabahar coastal area. Chabahar is located at 60°37’ E longitude and 25°17’N latitude, with a height of 8 meters from the free water level of the southernmost city of Sistan and Baluchestan province. The location of the Chabahar station is shown in (Fig. 1). The long-term study (2010-1991) of Chabahar climate parameters showed average annual temperature of 26 °C, average minimum temperature of 23.1°C, average maximum temperature of 29.4°C, average annual precipitation of 125.4 mm, average annual evaporation of 6.9 mm and average annual humidity of 74%.

Data

In order to predict the amount of harvestable water from air humidity in Chabahar, the required meteorological data and statistics including monthly data of absolute maximum temperature, absolute minimum temperature, average maximum temperature, average air temperature, absolute maximum relative humidity, absolute minimum relative humidity, average maximum relative humidity, average minimum relative humidity, average relative humidity, total precipitation, total evaporation, average evaporation, average sunshine, and average QFF pressure were obtained from the Chabahar synoptic station during the years 1990 to 2011.

Nonparametric tests were performed on the weather data of Chabahar synoptic station using HYFRAN-PLUS software (2008) and the results showed that the data were 95% homogeneous, random and independent and do not have trend.
The feasibility of water harvesting from air humidity in the area using statistical data

The results of theoretical calculations in the Chabahar area (Fig. 2) showed that because of average relative humidity of above 70%, except for the 3 cold months of the year (December, January, and February) and the relatively low temperature range, prone to any water harvesting design from humidity.

Calculation of the amount of harvestable water using statistical data

In order to investigate the existent water potential of the Chabahar station air, the absolute humidity parameter with unit of g / m³ should be used. This parameter is calculated according to equation 1:

\[ m = \frac{216.98}{T} \times e \]  

(1)

Where T is the temperature in degrees Kelvin and e is the vapor pressure in hPa. The results of this equation represent the amount of water vapor contained in one cubic meter of air.

As the meteorological data was read and recorded for periods of 3 hours, then the calculation unit of harvestable water must be calculated for 3 hours and with specific wind speed. The maximum amount of harvestable water was calculated by using equation 2:

\[ \text{(Humidity index)} \times \text{(wind speed)} \times \text{(one hour)} \times \text{(time index)} \times \text{(harvesting index)} = \text{amount of harvestable water} \]

(2)

Water harvesting

For the practical calculations of the amount of water harvesting from air humidity, a screen collector with dimensions of 1×1 m was designed and implemented (Fig. 3) (Mahmoudi et al., 2016). The amount of water harvesting from this collector was daily monitored for a period of 365 days (from September 2011 to September 2012).

Prediction of the amount of harvestable water using artificial neural network

The traditional statistical methods for modeling complex and nonlinear systems are often unmanageable, especially if the relationship between output and meas-
ured characteristics of the model is not clear (Namdar Khojasteh et al., 2011). But today, with the advent of science and the invention of smart methods, the necessity of its substitution (traditional methods) is posed. One of these smart methods is artificial neural networks (ANN). The use of ANN technique in dissolving engineering issues began in the late 1980s (Flood and Karam, 1994 a,b). The basic concepts of ANN and its application in hydrology are described in the report by the American Society of Civil Engineers (ASCE, 2003). High ability of ANN to predict and simulate water resource issues has been shown previously.

Artificial neural networks like the natural neural networks are composed of components called nerve cells. As in the natural neural network, a number of cells are responsible for receiving the stimulus effect, some for information processing, and a number for transmitting the response to the stimulus to the desired member. In ANN a number of cells also are responsible for receiving problem data, some for information processing, and some also for providing the answer to the problem. In all ANN, there is an input layer, an output layer, and some hidden layers. In the mathematical modeling of the neuron, a set of data is used as the input of the neuron (which may be the outputs of the other neurons) (Noori et al., 2013).

The calculation method in the neural networks is that the inputs to the neurons (x1 to x2) are multiplied by weights (w1 to w2), and the sum of the results of each input after applying in a function which is called the transfer function, is applied and the output of the neuron is determined. Equation (3) represents its mathematical model:

\[ net_j = \sum_{i=1}^{n} w_{ij} x_j \]  

In some cases, the steady-state value in each neuron namely Biase weight is also added to the above-mentioned equation and the equation 3 is given by the equation 4 (Fattahi et al., 2008):

\[ net_j = \sum_{i=1}^{n} w_{ij} x_j + b \]  

In this research, Hyperbolic Tangent Transfer Functions was used in the hidden layer. This function is most commonly used in simulations. Neural networks include various types of structures that are divided into different types based on direction of data entry and process (ASCE, 2003). For this reason, after review, two conventional types of artificial neural networks were used. These models included the Multi-Layer Perceptron-MLP model and the Generalized Feed Forward-GFF, which have high ability in predicting different climatic parameters (Azadeh et al., 2009; Behrang et al., 2010; Hung et al., 2008; Senthil Kumar et al., 2005).

In this research, to predict the amount of harvestable water among different training methods, the method of back propagation error with the Leungberg-Marquard algorithm was used because of faster convergence in network training. The basis of the method of back propagation error is based on the law of error-correction learning, which consists of two main paths of forward and backward. In the forward path, the input vector is applied to the network and its effects propagate through the middle layers to the output layer, and the output vector produces the real network response. (Ghabaei Sough et al., 2010).

NeuroSolution software, version 6, was used to investigate the possibility of predicting the amount of harvestable water using ANN. This software has the potential of designing, learning and evaluating ANN, and includes different networks with different learning rules due to using various stimulus functions among the existent stimulus functions in the software box. Also, in order to increase the accuracy and speed of the implementation of ANN, normalized data in the range of [0, 1] should be used. Since NeuroSolution software has the ability to normalize the data, the implementation of this step was done automatically by software.

**Neural network architecture and its performance evaluation criteria**

The choice of architecture in neural network calculations is a trial and error method in which the optimal network can be determined using different varieties of hidden layers and related neurons.

In order to evaluate the performance of the neural networks, the three factors namely the coefficient of explanation \( R^2 \), the root mean square error (RMSE) and the mean absolute magnitude error (MAE) were used. \( R^2 \) is a dimensionless criterion and its best value is equal to one. It is calculated based on the following equation (5):

\[ R^2 = \frac{\sum_{i=1}^{k} X_i Y_i}{\sqrt{\sum_{i=1}^{k} X_i^2 \sum_{i=1}^{k} Y_i^2}} \]  

The root mean square error (RMSE) and mean absolute magnitude error (MAE) also represent the error rate...
of the model. The best values for RMSE and MAE are zero and are calculated according to equations 6 and 7, respectively

\[ RMSE = \sqrt{\frac{1}{k} \sum_{k=1}^{k} (X_k - Y_k)^2} \]  
\[ MAE = \frac{1}{k} \sum_{k=1}^{k} |X_k - Y_k| \]

Where \( X_k \) = the observed values, \( Y_k \) = estimated values, and \( K \) = the number of data. Whatever RMSE and MAE are closer to zero and \( R^2 \) is closer to one, indicates that the outputs are more accurate and the observed and predicted values are closer to each other (Fattahi et al., 2008).

Model inputs and training courses and verification

Selection of model inputs is an important step in designing ANN. The most important factor in choosing the inputs of the model is the physics dominating the process of the research. Therefore, various inputs included maximum absolute temperature \( T_{\text{max,abs}}(^\circ \text{C}) \), absolute minimum temperature \( T_{\text{min,abs}}(^\circ \text{C}) \), average maximum temperature \( T_{\text{max,mean}}(^\circ \text{C}) \), average minimum temperature \( T_{\text{min,mean}}(^\circ \text{C}) \), average air temperature \( T_{\text{mean}}(^\circ \text{C}) \), absolute maximum relative humidity \( H_{\text{max,abs}}(\%) \), absolute minimum relative humidity \( H_{\text{min,mean}}(\%) \), average relative humidity \( H_{\text{mean}}(\%) \), Total precipitation \( R(mm) \), total evaporation \( E(mm) \), mean evaporation \( E_{\text{mean}}(mm) \), mean sunlight hour \( H_{\text{sun}}(s) \), average pressure \( P_{\text{OFF}}(\text{Hpa}) \) and the amount of harvestable water \( P(\text{m}^3/\text{day}) \) were considered monthly. The data set given to the network is divided into two general categories: a training set and a test set. This monthly data was generated between 1990 and 2011 and was introduced to the neural network model.

Finally, all available data were randomly divided into two groups as training (70%) and calibration (30%) groups. This categorization is based on the usual practice and there is no specific rule in this regard. However, various studies have shown that for training a better ANN, the number of training data should be more than the test stage (Diamantopoulou et al., 2005).

RESULTS AND DISCUSSION

The results of the field experiment (collecting water using screen collector) indicated that the amount of water available was 0.0086 m³/day in June and 0.0011 m³/day in February (Fig. 4). This indicated that the maximum amount of water was harvested in the warm season with highest relative humidity. On the other hand, the minimum amount of water was collected in the cold season with lowest relative humidity.

Since the whole water in the atmosphere cannot be harvested, harvesting indices of 10%, 20%, and 50% were used in Eq. 2. The comparison of the harvestable water from the screen collector with the values obtained from the theoretical indicators indicated no significant difference between the theoretical data of 20% and the real collector values (Fig. 5). Therefore, it can be concluded that only 20% of the existent atmospheric humidity can be harvested in southeastern Iran, and conse-
Projection of harvestable water from air humidity using artificial neural network (Case study: Chabahar Port)

Projection of harvestable water from air humidity using artificial neural network (Case study: Chabahar Port)

Table 1 presents some of the best results from the 12-month forecast, at the testing stage using the MLP and GFF ANN model, with different inputs and in different states of the number of hidden layers and hidden layer neurons.

The investigation of performance of the models indicated that in the test phase, the MLP had $R^2 = 0.86$ while RMSE and MAE were 2.19 and 1.81, respectively which was ideally able to predict the amount of harvestable water for 12 months (Table 1, Figs. 6 and 7).

According to Table 1, the selected model with the lowest standard error and the highest coefficient of determination has been MLP ANN with 12 neurons in the input layer, 1 hidden layer, 25 neurons for the hidden layer (12-1-25 array), whose stimulus function has been the Hyperbolic Tangent by Lewenburg Marquette Training Law.

On the other hand, the best result obtained from the GFF network at the test stage was $R^2 = 0.64$ whose RMSE and MAE values were 4.41 and 4.38, respectively.

Also, the best data for the 12-month prediction of the amount of harvestable water were as follows:

$$P = f(T_{\text{max}}^\text{mean}, T_{\text{min}}^\text{mean}, H_{\text{max}}^\text{mean}, H_{\text{min}}^\text{mean}, H_\text{mean}, R, P_{\text{QFF}})$$

The performance of this network in the 12-month prediction of absolute humidity in Chabahar is represented in Fig. 6 indicating the predicted data and observed data relative to the line. Fig. 6 displays the relationship between the output of the neural network and the actual values of the harvestable water in the form of first-order equation and the standard deviation of the first-order bisector line. Note that the closer the data are to the one to one graph, the greater the model’s ability to estimate the harvestable water will be. Fig. 6 also revealed that the results of ANN had a small dispersion.

The simulated and observed values of the 12-month prediction of the best structure of the ANN are demonstrated in Fig. 7. The graphical results of this figure indicated no significant difference between the observed and simulated values during the study period. This result was already confirmed based on the error scaling criteria (Table 1).

CONCLUSION

Since the study area suffers from a lack of adequate water, especially in rural areas, which sometimes even have difficulty with drinking water, some solutions should be developed to manage the existing resources.
The agricultural sector as well as optimal water management (precipitation, air humidity, etc.) in the future, can help efficiently. Predicting and calculating available water (MLP 12-1-25 2.19 1.81 0.86) in the air can be extracted in Chabahar region. After determining the effective input variables in predicting the amount of harvestable water using ANN, the modeling was performed using MLP and GFF. Between the two MLP and GFF networks, the MLP network presented greater ability than the GFF network to predict the amount of harvestable water.

The most suitable structure for predicting the harvestable water from fog in the Chabahar area was MLP ANN with 12 neurons in the input layer, 1 hidden layer, 25 neurons for the hidden layer, (12-1-25 array) and the Hyperbolic Tangent Stimulus Function with the Levenberg Marquardt Training Law. In general, it can be stated that ANN is a powerful model with high capability which can be viewed positively in predicting hydraulic problems especially when this network is able to extract the rule dominating data.

Since the amount of harvestable water is a nonlinear and complex phenomenon and many meteorological parameters are involved in its estimation, this research was conducted to predict it and to introduce an accurate estimate of the amount of harvestable water with the highest accuracy and lowest error. Notably, the prediction of harvestable water for 12 months, especially in warm and humid areas such as Chabahar, can be very valuable.

<table>
<thead>
<tr>
<th>Model</th>
<th>Network types</th>
<th>number of input</th>
<th>number of hidden layers</th>
<th>Hidden layers neurons</th>
<th>RMSE</th>
<th>MAE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P = F(T_{max_{abs}}, T_{min_{abs}}, H_{max_{abs}}, H_{mean}, R_{QFF}, E_{mean}, H_{sun})$</td>
<td>MLP 13</td>
<td>1</td>
<td>23</td>
<td>MLP</td>
<td>2.69</td>
<td>1.84</td>
<td>0.79</td>
</tr>
<tr>
<td>$P = F(T_{max_{abs}}, T_{min_{abs}}, T_{max_{over}, T_{min_{over}}, T_{mean}}, H_{max_{abs}}, H_{min_{abs}}, H_{mean}, R_{QFF}, E_{mean}, H_{sun})$</td>
<td>MLP 11</td>
<td>1</td>
<td>30</td>
<td>MLP</td>
<td>2.73</td>
<td>1.95</td>
<td>0.81</td>
</tr>
<tr>
<td>$P = F(T_{max_{abs}}, T_{min_{abs}}, T_{max_{over}, T_{min_{over}}, T_{mean}}, H_{max_{abs}}, H_{min_{abs}}, H_{mean}, R_{QFF})$</td>
<td>MLP 12</td>
<td>1</td>
<td>25</td>
<td>MLP</td>
<td>2.19</td>
<td>1.81</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Tab. 1. Comparison of different networks for 12-month prediction of the amount of harvestable water at the testing stage.
REFERENCES


