A Comparison of Performance of Artificial Neural Networks for Prediction of Heavy Metals Concentration in Groundwater Resources of Toyserkan Plain

Meysam Alizamir,1,* and Soheil Sobhanardakani2

1Young Researchers and Elite Club, Hamedan Branch, Islamic Azad University, Hamedan, IR Iran
2Department of the Environment, Hamedan Branch, Islamic Azad University, Hamedan, IR Iran

*Corresponding author: Meysam Alizamir, Young Researchers and Elite Club, Hamedan Branch, Islamic Azad University, Hamedan, IR Iran. Tel: +98-9125750213, E-mail: meysamalizamir@gmail.com

Received 2017 April 16; Revised 2017 May 15; Accepted 2017 June 05.

Abstract

Nowadays, about 50% the world's population is living in dry and semi dry regions and has utilized groundwater as a source of drinking water. Therefore, forecasting of pollutant content in these regions is vital. This study was conducted to compare the performance of artificial neural networks (ANNs) for prediction of As, Zn, and Pb content in groundwater resources of Toyserkan Plain. In this study, two types of artificial neural networks (ANNs), namely multi-layer perceptron (MLP) and Radial Basis Function (RBF) approaches, were examined using the observations of As, Zn, and Pb concentrations in groundwater resources of Toyserkan plain, Western Iran. Two statistical indicators, the coefficient of determination (R²) and root mean squared error (RMSE) were employed to evaluate the performances of various models. The results indicated that the best performance could be obtained by MLP, in terms of different statistical indicators during training and validation periods.

Keywords: Artificial Neural Networks, Heavy Metals, Groundwater, Multi-Layer Perceptron, Radial Basis Function, Toyserkan Plain

1. Introduction

Nowadays, heavy metals are present in the environment as a consequence of natural and anthropogenic activities, such as atmospheric emissions, weathering, volcanic eruptions, urbanization, industry, mining, agricultural activities, domestic activities, fuel combustion, and exhaust gases from automobiles. Heavy metals could cause serious adverse health effects in humans; therefore, they are known as the most dangerous pollutant (1-3). Bioaccumulation, non-biodegradable and long biological half-lives are the most important characteristics of heavy metals. Therefore, heavy metals become toxic and dangerous when they are not metabolized by the body and accumulate in body tissues of living organisms (4, 5).

Water is very important for mankind existence and economical development. Nowadays, about 50% the world’s populations are utilizing groundwater as a source of drinking water and other requirements. Therefore, the contamination of these resources by toxic heavy metals leads to serious problems (3). It has been proven that consumption of food, especially drinking contaminated water is one of the major sources of exposure of humans to toxic heavy metals (6). In this regard, discharge of heavy metals in the environment could contribute to increasing global concern and therefore monitoring and forecasting of their contents in the environment and especially in groundwater resources are important for food safety and public health protection.

Arsenic is a naturally occurring element, the major source of which is weathering of sedimentary and igneous rocks. Also, anthropogenic activities, including smelting of non-ferrous metals, production of energy from fossil fuel, mining activities, and applications of agricultural inputs, especially arsenical pesticides or herbicides, are other major sources of this element. Arsenic in groundwater resources was identified as a widespread intense problem all over the world (7-11). Arsenic could cause various types of adverse health effects, including black foot disease, cancers of the kidney, liver, prostate, bladder, lung, colon, skin, and the other serious diseases (12-14).

Zinc is an essential structural and functional element for several cellular processes, which often catalyze reactions, binding to substrates by favoring various reactions through the mediation of redox reactions, via reversible changes in the metal ions oxidation state. An excessive in-
take of Zn may play a major role in cancer etiology, and harms physiological activities, such as breathing (16-20).

Lead is well-known for its toxic properties and has serious adverse effects on human health, especially development of abnormalities in children. It has been proved about 85% of daily doses of this element is absorbed into the human body through food consumption. Therefore, food is the main source of non-occupationally human exposure to Pb (21-24).

The artificial neural networks (ANNs) has been successfully applied in environmental sciences during the recent years. In the environmental forecasting context, recent experiments have reported that ANNs may offer a promising alternative for estimating heavy metals’ concentration (25-30). Keskin et al. (2015) (31) employed different artificial neural networks for prediction of water pollution sources in Sivas, Karabuk, and Bartin areas of Turkey. Mandal et al. (2015) (32) successfully used ANN approach for modeling of As (III) removal process. Podder and Majumder (2015) (33) utilized ANN for forecasting phycoremediation efficiency of both As (III) and As (V) ions. Alizamir and Sobhanardakani (2016) (11) and Alizamir et al. (2017) (34) investigated the accuracy of ANN techniques in modeling heavy metals concentration in groundwater resources of Asadabad plain.

According to the knowledge of the authors of this study, there is not any published work in the literature related to comparison of two different neural networks (MLP and RBF), in forecasting of heavy metals concentration in groundwater resources of Toyserkan plain.

Based on the geological structure of Toyserkan Township, the rocks and minerals of this region contain heavy metals, such as As, Zn and Pb. Also rapid agricultural growth in this area, could lead to discharge of toxic heavy metals into groundwater resources through overutilization of agricultural inputs, such as chemical and organic fertilizers, especially phosphorus fertilizers, zinc sulfate, and heavy metal based pesticides (3, 35). Therefore, this study was conducted to assess the application of multilayer perceptron (MLP) models for predicting heavy metals (As, Zn, and Pb) concentration in the groundwater resources of Toyserkan Plain, and to determine the accuracy of MLPs compared with radial basis function (RBF) approaches.

2. Methods

2.1. Study Area

Toyserkan Plain is located in Toyserkan Township of Hamedan province, western part of Iran. The area of this region is about 800 km². Water requirement of most residents of this township is supplied from groundwater resources, including 1243 wells, 400 springs, and 220 aqueducts (3, 35).

2.2. Sample Collection

In the current study, based on the Cochran’s sample size formula, in total 60 groundwater samples were collected from 20 different open and tube wells located in agricultural and residential areas of plain during the period from September to November. Figure 1 shows the sampling stations in the study area.

Figure 1. Map of Sampling Stations

2.3. Sample Preparation and Analysis

The groundwater samples were taken in acid washed 200-mL polyethylene bottles to avoid unpredictable changes in characteristic, as per standard procedures. Samples were then filtered using Whatman 42 filter paper, preserved with 6N of suprapur nitric acid 65% (Merck, Germany) and kept at a temperature of 4°C for further analysis (3, 36). Finally, contents of heavy metals (As, Zn, and Pb) in groundwater samples were determined using inductively coupled plasma-optical emission spectrometers at wavelengths (nm) of 188.98 for As, 206.20 for Zn, and 220.35 for Pb (Varian, 710-ES, Australia).

2.4. Artificial Neural Networks

2.4.1. Multilayer Perceptron Neural Network

Artificial neural network as a nonlinear technique could solve problems, which are not suitable for conventional methods. One of the most commonly used neural network structure is the multi-layer perceptrons (MLPs) network with one or more hidden layers. The MLP employed in the current study had three layers, including an
input layer, a hidden layer, and an output layer. Detailed information about MLP are available in the literature (37). In the current study, the sigmoid function was used as the hidden layer’s activation function as follows:

\[
f(x) = \frac{2}{1 + e^{-2x}}
\]

(1)

For optimizing the neural network problem, a training algorithm is needed. Since there are several types of algorithms for training a network, it is essential to find an algorithm, which provides the best outputs. Recently, Levenberg-Marquardt training algorithm is utilized due to better performance and speed of learning (11, 38). Figure 2 shows the three-layer MLP neural network for this study, having one hidden layer with several nodes between the input and output layers. The code of ANN modelling was written using the MATLAB software.

### 2.4.2. Radial Basis Function Neural Network

The Radial Basis Function (RBF) is utilized in a wide range of prediction problems (39, 40). The RBF architecture possess input, hidden, and output layers. Each layer has some neurons. The hidden layer uses RBF function \((\phi (x,c))\), which is dependent on the distance from the origin to collect the input layer neurons. Therefore, variations of the RBF function are based on radial distance, \(r = ||x - c||\), where \(x\) denotes the input variable and \(c\) is the center of function (41). Equation 1 shows the RBF neural network output:

\[
f(x) = \sum_{i=1}^{N} c_i \phi(||x - x_i||)
\]

(2)

### 2.5. Models Performance Evaluation

The performance of artificial intelligence methods in training and testing periods is evaluated via two common statistical indicators, such as determination coefficient (R\(^2\)) and root mean square error (RMSE), which are expressed as follows:

\[
R^2 = \frac{\sum_{i=1}^{n} (O_i - \bar{O}_i) (P_i - \bar{P}_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O}_i)^2} \sum_{i=1}^{n} (P_i - \bar{P}_i)^2
\]

(3)

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}}
\]

(4)

Where \(O_i\) and \(P_i\) respectively account for the observed and predicted values at time \(i\), while the terms \(P_i\) and \(O_i\) imply the mean of the observed and predicted values; and \(n\) shows the number of data points.

### 3. Results and Discussion

Descriptive statistics of elements content (µg L\(^{-1}\)) in groundwater resources of Toyserkan Plain in the spring season are presented in Table 1. Data in Table 1 show that among the analyzed groundwater samples, As was detected in amounts ranging from 0.08 µg L\(^{-1}\) to 7.48 µg L\(^{-1}\) with an average level of 3.67 ± 2.23 of µg L\(^{-1}\); Zn was detected in amounts ranging from 0.12 µg L\(^{-1}\) to 15.64 µg L\(^{-1}\) with an average level of 3.84 ± 4.23 of µg L\(^{-1}\); and Pb was detected in amounts ranging from 0.09 µg L\(^{-1}\) to 5.50 µg L\(^{-1}\) with an average level of 1.66 ± 1.50 µg L\(^{-1}\).

<table>
<thead>
<tr>
<th>Element</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>As</td>
<td>0.08</td>
<td>7.48</td>
<td>3.67 ± 2.23</td>
</tr>
<tr>
<td>Zn</td>
<td>0.12</td>
<td>15.64</td>
<td>3.84 ± 4.23</td>
</tr>
<tr>
<td>Pb</td>
<td>0.09</td>
<td>5.50</td>
<td>1.66 ± 1.50</td>
</tr>
</tbody>
</table>

Based on independent one-sample t-test comparing the heavy metal contents in groundwater samples of Toyserkan Plain with maximum permissible limits (µg L\(^{-1}\)) (100.0, 2000.0, and 100.0 for As, Zn, and Pb, respectively) established by the world health organization (WHO) (35), it was shown that the mean contents of all analyzed elements were lower than MPL.

In this study, for the same basis comparison, the same training and testing sets are utilized for the 2 models. Two standard statistical performance evaluations are employed to evaluate the performances of models. For all heavy metals concentrations, all the ANN models were first trained using the data in the training sets (using the first 75% of the data) to obtain the optimized set learning coefficients, and then tested (using the 25% of the data).

The ANN models for heavy metals concentration forecasting were developed using the MATLAB R2014 software program. The MLP and RBF models were trained and tested based on the same dataset for each metal. The number of neurons in the hidden layer was selected using a trial-and-error procedure (42). The Levenberg-Marquardt algorithm (43) was employed to train the MLP model.

Table 2 presents the results of MLP and RBF models in terms of various performance statistics. It could be observed from Table 2 that the MLP model had better performance during both training and testing, and it outperformed RBF in terms of standard statistical indicators. For As concentration in the training phase, the MLP model obtained the best R\(^2\) and RMSE of 0.9464 and 0.5834, respectively and the RBF model obtained R\(^2\) and RMSE of 0.9377.
and 0.7313, respectively. Analyzing the results during testing, it could be observed that the MLP model outperformed the RBF model. For Zn concentration, in the training phase, the MLP model obtained the best $R^2$ and RMSE of 0.9821 and 0.9946, respectively and the RBF model obtained $R^2$ and RMSE of 0.9702 and 1.0427, respectively. Analyzing the results during testing, it could be observed that the MLP model outperformed the RBF model. Similarly, for Pb concentration, in the training phase, the MLP model obtained the best $R^2$ and RMSE of 0.9748 and 0.2666, respectively and the RBF model obtained $R^2$ and RMSE of 0.9524 and 0.3684, respectively. Analyzing the results during testing, it could be observed that the MLP model outperforms the RBF model.

The performances of models developed in this paper during the training and testing phases for all heavy metals concentration are shown in Figures 3 to 8. It could be observed that the MLP model had less scattered estimates and that the values were denser in the neighborhood of the straight line compared to the RBF model. In both the model training phase and the model testing phase, the MLP simulated data showed greater agreement with the observed data than the RBF simulated data. Overall, it could be concluded that the MLP model for all heavy metals provided more accurate forecasting results than the RBF model for forecasting of heavy metals concentration in the Toyserkan plain.

Comparison of the model efficiency statistic ($R^2$) between the MLP model and the RBF model, presented in Table 2, revealed that the MLP model had outperformed the RBF model in both the training phase as well as in the testing phase. In the testing phase, the RBF model had lower $R^2$ than the MLP model for all heavy metals.

Nor et al. (2014) (44) applied planar electromagnetic sensor array and artificial neural network for nitrate and sulfate estimation in water sources. The results demonstrated that artificial neural network achieved a high degree of accuracy to estimate the water contamination and it is a robust low-cost approach for water source monitoring.

![Figure 2. A Three-Layer Multilayer Perceptrons Neural Network Model](image)

#### Table 2. Comparative Performance of Artificial Neural Networks for As, Zn and Pb Concentration

<table>
<thead>
<tr>
<th>Element</th>
<th>Methods</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$RMSE$</td>
<td>$R^2$</td>
<td>$RMSE$</td>
</tr>
<tr>
<td>As</td>
<td>MLP</td>
<td>0.5834</td>
<td>0.9464</td>
</tr>
<tr>
<td></td>
<td>RBF</td>
<td>0.7313</td>
<td>0.9377</td>
</tr>
<tr>
<td>Zn</td>
<td>MLP</td>
<td>0.9946</td>
<td>0.9821</td>
</tr>
<tr>
<td></td>
<td>RBF</td>
<td>1.0427</td>
<td>0.9702</td>
</tr>
<tr>
<td>Pb</td>
<td>MLP</td>
<td>0.2666</td>
<td>0.9748</td>
</tr>
<tr>
<td></td>
<td>RBF</td>
<td>0.3684</td>
<td>0.9524</td>
</tr>
</tbody>
</table>

y = 0.9718x + 0.2696  
R² = 0.9464

MLP Prediction of As Concentration - Training Phase

Figure 1. Observed and Simulated As Concentration by the Multi-Layer Perceptron Model During the Training and Testing Phases

y = 1.2913x - 1.2779  
R² = 0.9302

Figure 2. Observed and Simulated As Concentration by the Radial Basis Function Model During the Training and Testing Phases

y = 0.9625x - 0.2793  
R² = 0.9377

Figure 3. Observed and Simulated As Concentration by the Multi-Layer Perceptron Model During the Training and Testing Phases

y = 0.6591x + 1.146  
R² = 0.9199

Figure 4. Observed and Simulated As Concentration by the Radial Basis Function Model During the Training and Testing Phases

y = 0.8674x - 0.0358  
R² = 0.9821

Figure 5. Observed and Simulated Zn Concentration by Multi-Layer Perceptron Model During the Training and Testing Phases
The equations and graphs presented in the text indicate the following:\n\[ y = 0.8753x - 0.0301 \quad R^2 = 0.9702 \]
\[ y = 0.6853x + 0.4722 \quad R^2 = 0.8953 \]
\[ y = 0.9668x + 0.16 \quad R^2 = 0.9748 \]
\[ y = 0.7426x + 0.255 \quad R^2 = 0.9626 \]

The figures illustrate the observed and simulated Zn concentration by Radial Basis Function (RBF) model during the training and testing phases, and the observed and simulated Pb concentration by Multi-Layer Perceptron (MLP) model during the training and testing phases.

**Figure 6.** Observed and Simulated Zn Concentration by Radial Basis Function Model During the Training and Testing Phases

**Figure 7.** Observed and Simulated Pb Concentration by Multi-Layer Perceptron Model During the Training and Testing Phases

**Figure 8.** Observed and Simulated Pb Concentration by Radial Basis Function Model During the Training and Testing Phases
4. Conclusions

In the current study, the long-term changes in trends of heavy metals (As, Zn, and Pb) concentration in groundwater resources of Toyskerkan Plain were estimated by 2 different artificial neural network approaches. The data from actual field observed data in the Toyskerkan plain were employed to develop models investigated in this study. The results were analyzed visually by using scatter plots and also in terms of parameters, such as RMSE and R². The results obtained in this study indicated that the ANNs methods were powerful tools to model heavy metals concentration. This study found that the MLP model was substantially more accurate than the RBF model. The accurate forecasting results for As, Zn, and Pb heavy metals concentration in the Toyskerkan plain revealed that the MLP method is a potentially very useful approach for heavy metals concentration forecasting.

References


