An Adaptive Neuro-Fuzzy Inference System (ANFIS) to Predict of Cadmium (Cd) Concentrations in the Filyos River, Turkey

Adem Yavuz Sonmez¹, Semih Kale², Rahmi Can Ozdemir¹, Ali Eslem Kadak¹

¹ Fisheries Faculty, Kastamonu University, Kastamonu, Turkey.
² Faculty of Marine Sciences and Technology, Çanakkale Onsekiz Mart University, Çanakkale, Turkey.

Abstract

Water quality is one of the main characteristics of a river system and prediction of water quality is the key factor in water resource management. Different physical, biological and chemical parameters including heavy metals can be used to assess river water quality. Evaluation of the water quality in the rivers is quite difficult and requires more time and effort because of the fact that many factors affect water quality. Traditional data processing methods are insufficient to solve this problem. Therefore, in this study, an adaptive neuro-fuzzy inference system (ANFIS) model was developed to predict the concentrations of cadmium (Cd) in the Filyos River, Turkey. For this purpose, water samples collected at 7 sampling locations in the river during December 2014-2015 were used to develop ANFIS model. The available data set was apportioned into two separate sections for training and testing the ANFIS model. Developed models aimed to use the least parameters to estimate Cd concentration. As a result, a relatively higher correlation (R²=0.91) was found between observed and modelled Cd concentrations. The results indicated that the ANFIS model gave reasonable estimates for the concentrations of Cd with a high degree accuracy and robustness. In conclusion, this paper suggests that ANFIS methodology produce very successful findings and has the ability to predict Cd concentration in water resources. The outcomes of this research provide more information, simulation, and prediction about heavy metal concentration in natural aquatic ecosystems. Therefore, ANFIS can be used in further researches on water quality monitoring.

Keywords: ANFIS, cadmium, heavy metal.

Introduction

Metals move to rivers from different sources, such as soils and rocks directly get exposed to surface waters, fallout of atmospheric particulate matter, decomposing dead organic matter, and from anthropological activities, including the release of refined and unrefined liquid wastes into the water resource (Olayinka & Alo, 2004, Sönmez et al., 2016). In particular, releasing of Cd into the aquatic environment from industrial activities is a serious threat for the aquatic ecosystems (Qasaimeh, Abdallah, & Bani Hani, 2012).

Heavy metal dynamics have nonlinear and highly complex relations with each other. Therefore, traditional data processing and modelling methods required several input parameters which are hard to reach and make it a time consuming and expensive process are insufficient to solve the problems related to water quality (Ranković, Radulović, Radiojević, Ostojić, & Čomić, 2012; Bayatzaheh Fard, Ghadimi, & Fattahi, 2017). The integration of different techniques and methods will contribute to the future of eco-environmental modelling (Chen, Morales-Chaves, Li, & Mynett, 2006). Soft computing techniques have been commonly used in many areas including environmental modelling and water resource engineering. Modelling complex nonlinear systems is one of the successful applications of artificial intelligence (AI) techniques such as fuzzy inference systems (FIS), artificial neural networks (ANNs), genetic algorithms and knowledge-based systems (Wang, Chau, Cheng, & Qiu, 2009).

Several authors studied various AI techniques in environmental modelling, water quality monitoring and assessment, predicting the concentrations of heavy metals and other quality parameters, and estimation and forecasting in climatic sciences (Soypak et al., 2003; Altunkaynak, Özger, & Çakmakçı, 2005; Kisi, 2005; Ocampo-Duque, Ferré-Huguet, Domingo, & Schuhmacher, 2006; Sengörur, Dogan, Koklu, & Samandar, 2006; Terzi, Keskin, & Taylan, 2006; Daiya, Singh, Gaur, Garg, & Kushwaha, 2007; İcaga, 2007; Elhatip & Kömür, 2012; Dogan, 2016).
Although heavy metal modelling has been studied by using various statistical and computational approaches, several aspects of their dynamics are still blurred (Akoto, Bruce, & Darko, 2008). Therefore, the purpose of this study was to create an adaptive neuro-fuzzy inference system model to predict the concentrations of Cd in the Filyos River, Turkey and to demonstrate the model application to identify nonlinear complex relationships between input and output variables.

**Material and Methods**

**Study Area and Sampling**

Filyos River is geographically located at the southern west coast of the Black Sea in Turkey. It flows through the West Black Sea River Basin and run into Black Sea from Filyos town in Zonguldak city (Figure 1). It is totally 312 km in length. It also has two main branches: Devrek Stream and Yenice Stream and 13300 km² drainage area (Kucukali, 2008). Çaycuma and Gökçebey districts situate near the Filyos River and their populations have been increasing day by day. Although the economy of the region is dependent on agriculture and forestry, it has been changed with the construction of some industrial factories such as cement and paper factories (Seker et al., 2005). There are substantial differences in the richness of terrestrial ecosystem attributable to the topographical features of the region and the delta of the river is a natural environment with untouched areas (Kucukali, 2014).

The water samples were collected twice at a time by monthly from seven stations at the Filyos River in Turkey between December 2014 and December 2015 (Figure 1). The coordinates of sampling stations were obtained using a handheld GPS device (Table 1). Samples were collected with the Nansen bottle and preserved in the polyethylene (PE) bottles after filtering with 0.45 μm pore size membrane filter. Both the Nansen bottle and PE bottles were washed with...
water to be collected in the sampling stations (Alam, Tanaka, Stagnitti, Allinson, & Maekawa, 2001). Collected water samples were stored in a cooler at +4°C. Heavy metal (Fe, Cu, Mn, Zn, Ni, Cr and Cd) concentrations were determined by Inductively Coupled Plasma Optical Emission Spectrometry (ICP/OES).

Data Preparation and Input Selection

Normalization of water quality parameters is significant for avoiding larger numbers in a uniform range and to be able to scale the data in a same range. Normalization process has an important effect on data preparation to speed up the training time in ANFIS. There are many statistical normalization techniques such as Z-score, min-max, median, sigmoid, and statistical column normalization (Jayalakshmi & Santhakumaran, 2011). In this study, all input data were normalized using Z-score normalization technique for which formula is given in Eq (1). Z-score normalization technique uses standard deviation and mean for each feature through a set of training data for normalizing each input feature.

\[
S_i = \frac{(x_i - \mu_i)}{\sigma_i}
\]

where, \(x_i\) is input data, \(\sigma_i\) is standard deviation of \(x_i\), \(\mu_i\) is mean of \(x_i\), \(S_i\) is normalized \(x\) variable as input to ANFIS.

It is important to determine whether the variables have a relationship with one another while modeling Cd values for the Filyos River. For this reason, a coefficient of correlation with Cd was prepared which is tabulated in Table 2.

Adaptive Neuro-Fuzzy Inference System (ANFIS)

A fuzzy system based on the logical rules of premises and conclusions cannot be analyzed with traditional probability theories. Developing fuzzy “if-then” rules is the starting point in constructing a fuzzy system. An effective tool for this purpose is a method that can convert data to the required fuzzy rules. Alternatively, artificial neural networks are capable of generating appropriate relations between input and output variables through learning capabilities based on different training patterns. A combined system of artificial neural networks and fuzzy inference accomplished with using numerical data to predict output can generate an influential tool, adaptive neuro-fuzzy inference system (ANFIS).

ANFIS is a type of neural network focused on Takagi-Sugeno fuzzy inference system. It is an AI technique currently using in hydrological processes (Bisht & Jangid, 2011). ANFIS is firstly introduced by Jang (1993) and it is based on the first-order Sugeno fuzzy model. ANFIS commonly uses either back-propagation or a combination of back-propagation and least square estimation for prediction of membership function parameter (Jang, Sun, & Mizutani, 1997). The most significant purpose of integrating neural networks with fuzzy systems is to exercise learning ability of neural network while the learning capability is a benefit in terms of fuzzy system; on the other hand, there are more benefits for a combined system from the aspect of a neural network.

Takagi-Sugeno type fuzzy inference system is used in ANFIS where every rule’s output can be a constant term or can be a linear combination of input variables addition to a constant term. The weighted average of every rule’s output is the final output. The basic architecture of ANFIS which has two inputs (\(x, y\)) and one output (\(z\)) is presented in Figure 2.

The rule base of ANFIS contains two Takagi-Sugeno type if-then rules as given below:

Rule 1: If \(x\) is \(A_1\) and \(y\) is \(B_1\); then \(z = p_1 x + q_1 y + r_1\)

Rule 2: If \(x\) is \(A_2\) and \(y\) is \(B_2\); then \(z = p_2 x + q_2 y + r_2\)

where \(A_1, A_2, B_1\) and \(B_2\) are nonlinear parameters while \(p_1, p_2, q_1, q_2, r_1\) and \(r_2\) are linear parameters.

Layer 1 is the fuzzification layer in which \(x\) is the input of \(A_1\) and \(B_1\) nodes and \(y\) is the input of \(A_2\) and \(B_2\) nodes. \(A_1, A_2, B_1\) and \(B_2\) are used in the fuzzy theory for allocating the membership functions as linguistic labels. The membership relationship between the input and output functions of this layer can be shown as follows:

---

**Table 1. Coordinates of the sampling stations**

<table>
<thead>
<tr>
<th>Station code</th>
<th>Latitude (N)</th>
<th>Longitude (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>41.552125</td>
<td>32.081409</td>
</tr>
<tr>
<td>S2</td>
<td>41.299621</td>
<td>32.096858</td>
</tr>
<tr>
<td>S3</td>
<td>41.261114</td>
<td>32.157202</td>
</tr>
<tr>
<td>S4</td>
<td>41.211915</td>
<td>32.267908</td>
</tr>
<tr>
<td>S5</td>
<td>41.167057</td>
<td>32.579248</td>
</tr>
<tr>
<td>S6</td>
<td>41.233736</td>
<td>33.303101</td>
</tr>
<tr>
<td>S7</td>
<td>41.037858</td>
<td>33.294250</td>
</tr>
</tbody>
</table>
where $\mu_{Ai}$ and $\mu_{Bi}$ indicate the membership functions and $O_{i1}$ and $O_{i2}$ indicate the output functions.

Layer 2 is the product layer which includes two fixed nodes labeled with $\Pi$. The outputs of this layer are $w_1$ and $w_2$. These outputs are the weight functions of layer 3 and the product of the input signal can be shown as follows:

$$\begin{align*}
O_{i1} &= \mu_{Ai}(x), \ i = 1, 2 \\
O_{i2} &= \mu_{Bi}(y), \ i = 1, 2
\end{align*}$$

(2)

Layer 3 is the normalized layer, which includes two fixed nodes labeled with $N$. $\overline{w_1}$ and $\overline{w_2}$ are the outputs of this layer. Normalizing the weight function is the task of this layer in the next process:

$$O_{i3} = \overline{w_i} = \frac{w_i}{w_1 + w_2}, \ i = 1, 2$$

(3)

Layer 4 is the defuzzification layer which includes two adaptive nodes. The relationship between the inputs and output of this layer can be shown as follows:

$$O_{4i} = \overline{w_i}(p_i x + q_i y + r_i), \ i = 1, 2$$

(5)

where $p_i$, $q_i$, and $r_i$ are the linear parameters of the node and $O_{4i}$ is the output of this layer.

Layer 5 is the output layer which includes a fixed node labeled with $\Sigma$. The output of this layer is comprised all the input components, which denotes the cleaning rates results. The output can be defined as follow:

$$O_{5i} = \sum_i \overline{w_i} f_i = \frac{\Sigma \overline{w_i} f_i}{\Sigma \overline{w_i}}, \ i = 1, 2$$

(6)

Matlab software was used to conduct the data computation for ANFIS. ANFIS training includes the gradient descent method and the least squares method. ANFIS training algorithms embedded in the Matlab fuzzy inference toolbox simplify data processing using the training and forecasting functions. Main computation process consists of four stages. Data input is the first stage. Assigning fuzzy sets is the second stage. Some fuzzy sets should be assigned to

**Table 2. Basic statistics of the measured variables in the Filyos River**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Mean</th>
<th>SE</th>
<th>SD</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Correlation with Cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iron (Fe)</td>
<td>$\mu g \ L^{-1}$</td>
<td>0.2248</td>
<td>0.0019</td>
<td>0.0176</td>
<td>0.1554</td>
<td>0.2810</td>
<td>0.1730</td>
</tr>
<tr>
<td>Copper (Cu)</td>
<td>$\mu g \ L^{-1}$</td>
<td>0.0811</td>
<td>0.0001</td>
<td>0.0008</td>
<td>0.0778</td>
<td>0.0830</td>
<td>0.4780</td>
</tr>
<tr>
<td>Manganese (Mn)</td>
<td>$\mu g \ L^{-1}$</td>
<td>0.0927</td>
<td>0.0012</td>
<td>0.0107</td>
<td>0.0504</td>
<td>0.0995</td>
<td>0.0000</td>
</tr>
<tr>
<td>Zinc (Zn)</td>
<td>$\mu g \ L^{-1}$</td>
<td>1.1559</td>
<td>0.0157</td>
<td>0.1430</td>
<td>0.6769</td>
<td>1.2291</td>
<td>0.0000</td>
</tr>
<tr>
<td>Nickel (Ni)</td>
<td>$\mu g \ L^{-1}$</td>
<td>0.0481</td>
<td>0.0002</td>
<td>0.0021</td>
<td>0.0457</td>
<td>0.0529</td>
<td>0.0000</td>
</tr>
<tr>
<td>Chromium (Cr)</td>
<td>$\mu g \ L^{-1}$</td>
<td>0.0341</td>
<td>0.0002</td>
<td>0.0014</td>
<td>0.0313</td>
<td>0.0359</td>
<td>0.0000</td>
</tr>
<tr>
<td>Cadmium (Cd)</td>
<td>$\mu g \ L^{-1}$</td>
<td>0.0433</td>
<td>0.0002</td>
<td>0.0014</td>
<td>0.0416</td>
<td>0.0464</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

SE: standard error; SD: standard deviation.
each type of input and output data for the data input. The system will automatically determine the membership functions for them consistent with range of data and the fuzzy sets in the process of data processing. Using the training function for learning of input data is the third step. ANFIS will learn the data by performing the training data function and collect training errors after the analysis. The last stage is prediction of the output.

Dataset was randomly separated into two equal parts as training and test data. Sub-clustering partitioning technique with Gaussian membership functions was used for generating FIS to fuzzify the input data. The training stage contained an iterative procedure which targeted to calculate optimum values by minimizing the sum of squared differences between training data values and model predictions. Hybrid learning algorithm was selected to train fuzzy inference system. Training process continued until errors maintained the stability.

Results and Discussion

The concentrations of heavy metals including Fe, Cu, Mn, Zn, Ni, Cr and Cd were measured at seven stations in Filyos River. Basic statistics of the dataset is given in Table 2. ANFIS method in Matlab software was used to develop a model to predict Cd concentration. In ANFIS modelling, observed dataset was divided into three stages for training, testing and validation. Gaussian membership function (gaussmf) was considered for input and linear membership function was considered for output parameters. Hybrid algorithm was used to define the optimum number of parameters to describe the FIS.

Training data was consisted of 75% of dataset while 25% of dataset was assigned for testing in ANFIS. Average error value was found as 0.00019036 for both training and testing with 100 epochs (Figure 3). The relationships between input and output variables including three rules are illustrated in Figure 4.

ANFIS model was developed with a high correlation coefficient for training, testing and validation between 0.868 and 0.978. The higher the R-squared ANFIS models were filtered using different membership functions for different combinations of the variables. Performance of the models was evaluated for training and testing. Table 3 displays the mean absolute deviation (MAD), mean squared error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), Nash-Sutcliffe efficiency (E), and coefficient of determination ($R^2$).

Observed and modelled values of Cd concentrations are illustrated in Figure 5.
Figure 4. Rules of ANFIS model illustrating the relationships between input and output variables.

Table 3. The performance parameters of ANFIS model for testing, training, and whole stage

<table>
<thead>
<tr>
<th>Stage</th>
<th>MAD</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAPE</th>
<th>E</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>0,00013468</td>
<td>0,00000004</td>
<td>0,00019313</td>
<td>0,30969120</td>
<td>0,980557449</td>
<td>0,978</td>
</tr>
<tr>
<td>Testing</td>
<td>0,00052381</td>
<td>0,00000058</td>
<td>0,00076345</td>
<td>1,16788005</td>
<td>0,675131021</td>
<td>0,868</td>
</tr>
<tr>
<td>Whole</td>
<td>0,00023313</td>
<td>0,00000018</td>
<td>0,00041873</td>
<td>0,52682332</td>
<td>0,90772321</td>
<td>0,913</td>
</tr>
</tbody>
</table>

Figure 5. Comparison of measured and modelled Cd concentrations in the Filyos River water.
demonstrates the comparison of observed and predicted Cd concentrations and Figure 6 illustrates the scatter plot for the model of predicted and observed Cd concentration in the Filyos River. The respective correlation coefficient of 0.913 was found between observed and modelled Cd values.

Kemper and Sommer (2002) estimated the concentrations of heavy metal in soils from using back-propagation neural network (BPNN) and multiple linear regression (MLR). Almasri and Kaluarachchi (2005) forecasted the distribution of nitrate in groundwater using the modular neural networks. Rooki, Doulati Ardejani, Aryafar, and Bani Asadi (2011) anticipated heavy metals in acid mine drainage using BPNN, general regression neural network (GRNN), and MLR in Shur River, Iran. Yılmaz Öztürk et al. (2014) applied fuzzy logic assessment method for heavy metal pollution in Apa Dam Lake and indicated that fuzzy logic methods suggested more precise evaluation than traditional classification methods. Qasaimeh et al. (2012) used ANFIS for heavy metal sorption in aquatic environments and found a higher correlation of 0.98 between observed data and modelled results. Yesilnacar and Sahinkaya (2012) used ANN for prediction of sulfate and sodium adsorption ratio in an unconfined aquifer in Turkey. Valente, Ferreira, Grande, de la Torre, and Borrego (2013) performed FIS to predict the metal concentrations of an acid mine drainage. Sönmez et al. (2013b) conducted a comparative analysis of water quality assessment methods for heavy metal pollution in Karasu Stream. The authors used four assessment methods to evaluate heavy metal pollution in the stream and highlighted that fuzzy mathematical models were more reasonable. Sönmez et al. (2013a) used fuzzy logic evaluation method to assess the water quality in terms of heavy metal pollution in Karasu Stream and they recommended fuzzy logic evaluation method as a tool to predict heavy metal concentrations and water quality assessment. El Badaoui, Abdallaoui, Mansouri, and Lancelot (2013) used ANN of multilayer perceptron type to predict of heavy metal concentrations in Moroccan aquatic sediments. Chang et al. (2014) applied the neuro-fuzzy networks to evaluate the concentration of arsenic in Huang gang Creek, Taiwan. Keskin, Diğenci, and Kaçaroğlu (2015) applied ANN to estimate water pollution in Turkey. Ghadimi (2015) used ANN to anticipate the heavy metal concentrations in the groundwater of Arak, Iran. Akpomie et al. (2016) used statistical tools to predict heavy metal concentrations and developed a mathematical model for computer modelling of heavy metal concentrations. Bayatzadeh Fard et al. (2017) used AI techniques to predict heavy metal concentrations in the groundwater resources of Lakan lead-zinc mine, in Khomein, Iran. The authors developed different models namely ANN, hybrid ANN with biogeography-based optimization (ANN-BBO), and multi-output adaptive neural fuzzy inference system (MANFIS) to estimate the heavy metal concentrations. They indicated that MANFIS model was a reliable method to predict heavy metal concentrations with a high degree of robustness and accuracy.

A relatively higher correlation (R²=0.91) was achieved between observed and modelled Cd concentrations (Table 3). The prediction ability of ANFIS model is nearly similar to (Soyupak et al., 2003; Zhao, Nan, Cui, & Guo, 2007; Akpomie et al., 2016). Akpomie et al. (2016) found a correlation coefficient of 0.86 for Cd concentration. Similarly, Soyupak et al. (2003) found the coefficient of determination as 0.95 and Zhao et al. (2007) found it as 0.94 between observed and modeled DO values. Singh et al. (2009) found the coefficient of correlation for observed and modelled DO values as 0.85, 0.77, and, 0.85 for the training, testing and validation, respectively. Figure 7 shows the surface view plots of the anticipated fuzzy rules. These 3D plots illustrate

![Figure 6](image_url)  
**Figure 6.** Scatter plot for the model of predicted and observed Cd concentrations in the Filyos River.
Figure 7. Surface view of the relationships between input and output variables.
the relationships between input and output variables.

A relatively higher correlation between observed and modeled values of Cd concentrations could be related to the experimenting with comparatively smaller dataset for input variables in the study. Several authors found a higher correlation coefficient in different aquatic environments for different variables (Soyupak et al., 2003; Sengor et al., 2006; Zhao et al., 2007; Dogan et al., 2009). On the other hand, Nemati et al. (2015) found relatively poor results from ANFIS model between observed and modelled values of dissolved oxygen. Singh et al. (2009) found relatively lower correlation between observed and modelled values of DO. Singh et al. (2009) stated that the relatively lower correlation might be by the reason of longtime sampling duration, sampling locations, distributed large areas, and non-homogenous characteristic of the water quality variables.

ANFIS takes advantages of the outstanding learning algorithms of ANN and superb prediction functions of FIS. It can make available the calculation without mathematical modeling and provide a good solution for the problem of the non-linear prediction. ANFIS has an adaptive background and use training data to generate a fuzzy inference system. Therefore, ANFIS was chosen in the study. The obtained results of the study can be compared with related researches that discussed in the paper. The results demonstrated that suggested ANFIS models are capable to predict the Cd concentration in rivers using restricted dataset.

**Conclusion**

In this study, ANFIS models were developed to predict the Cd concentration in Filyos River. The results of the study indicated that ANFIS methodology produced very successful findings and had the ability to predict Cd concentration in water resources. It was determined that the ANFIS model was a reliable method to predict heavy metal concentrations in water resources with an acceptable degree of robustness and accuracy. Use of adaptive neuro-fuzzy inference system is significant for unclear system which has no experience with data behavior. The outcomes of this research provide more information, simulation, and prediction about heavy metal concentration in natural aquatic ecosystems. Therefore, ANFIS can be used in further researches on water quality monitoring.

**Acknowledgement**

This study was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK Project No: 114O145). An earlier version of this paper was presented at International Congress on Environmental Researches and Technology (ICERAT 2017) in Belgrade, Serbia.

**References**


A neural network for prediction in southeastern Turkey.


