PREDICTION OF DISSOLVED OXYGEN VALUES IN AKARCAY BASIN USING ARTIFICIAL NEURAL NETWORKS

PREPARED BY BEKİR YILMAZ

ANKARA – 2015

EXPERTISE THESIS ABSTRACT

In this study, Akarcay Basin is chosen for prediction of dissolved oxygen values by using Artificial Neural Networks (ANNs). The basin covers the geographic area where a water body extends from the source to the place where it is poured, providing water to this water source. It is located in the west of Turkey, in the Mediterranean, Aegean and Central Anatolia regions, The basin is also approximately 130 km long and 20 km wide. However, the basin is geographically located between 30^{o}-32^{o} east longitudes and 38^{o}-39^{o} north latitudes.

Figure 1. Akarcay Basin

Akarçay River which gives the basin it’s own name, has a west to east flow direction form. There are also small streams in the basin, that in turn participate in Akarçay, which is being poured into Eber and Akşehir lakes to the east of the basin.

Dissolved Oxygen

The dissolved oxygen concentration determined in water, refers to the degree of water contamination. It also demonstrates the concentration of organic matter in the water and the extent, which water can clean itself. The percentage of dissolved oxygen in water is inversely proportional to temperature and atmospheric pressure. At high pressure, high amounts of oxygen are dissolved within water. However, low amounts of oxygen are dissolved at high temperatures. Therefore, these parameters are closely related. Thus the problem demonstrates a cross relation between physicochemical parameters due to the nature of the water chemistry. This brings to table for a solution with using neural networks for dissolved oxygen values.

Sampling Points

In this study 8 sampling points in the basin are used for the model. The dataset, contains a total of 25024 rows of data for 34 parameters, including Dissolved Oxygen. The data is composed of 4 samples every 3 months in a year.

Table 1. Sampling Points

Name	**Sampling Point ID**	**Longitude**	**Latitude**
Eber Gölü – Kocakandiralik	11-18-01-013	31.145760749625023E	38.62998785107423N
Eber Gölü - Yali Mevkii	11-18-01-012	31.123124727415796E	38.62712178395024N
Eber Gölü - Ortasi - Donbay	11-18-01-011	31.144234497360323E	38.64774207963218N
Bolvadin Köprüsü	11-18-00-007	31.04793848659954E	38.66029414315788N
Seker Fabrikasi Sonrasi	11-18-00-004	30.79281097828344E	38.6858962060904N
Afyon AAT Sonrasi	11-18-00-038	30.584768801245332E	38.77206312016254N
Arapli Deresi	11-18-00-037	30.424874945918997E	38.847392030049534N
Eber Gölü - Eber Regülatörü	11-18-02-016	31.230391059590833E	38.61077187571954N

The dataset is a discrete time series. Before putting the data to the ANN model, it was normalized by using min-max normalisation method between the values of 0.1-0.9. In this method, the largest and smallest values within a group of data are considered. All other data is normalized according to these values. The aim here is to normalize the smallest value to 0.1 and the maximum value to 0.9 and make it spread all other data to this range. Thus, it is aimed to minimize the effect of peak values on the model.

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

Equation 1. Min-max Normalisation

In this study, Sequence-Parallel NARX model had been used because of the long-term dependency and time series varying and depending on other parameters due to the nature of the water chemistry. The model - NARX (Nonlinear Autoregressive Exogenous) networks- are two-layer feed, loop and dynamic networks.

Within the scope of the study, it has been tried to be estimated by using different datasets and models at 8 sampling points in Akarçay Basin, using the measured values of Dissolved Oxygen value, month, year, season and other various parameters. Then these models were tested and compared for their predictive performance on the respective data sets. In order to find the most suitable NARX model, an artificial neural network application was developed in MATLAB environment.
In NARX-based ANNs, the sigmoid activation function is used in the hidden layer and the linear activation function is used in the output layer. Levenberg-Marquardt (LMA) algorithm is chosen as the learning algorithm.

In order to find the simplest and most efficient model, firstly, it was tried to estimate the dissolved oxygen concentration by using various configurations of basic NARX Network based structures having 34, 9 and 2 neurons in the input layer at the sampling point of the Afyon Waste Water Treatment Plant. The best results were obtained in the NARX configuration with 9 neurons in the input layer and 16 neurons in the hidden layer, 4: 4: 40 with input and feedback delay. The model was then run at other sampling points and the success rate of estimating dissolved oxygen was found to be 75%.

As a result, after a bunch of experiments, it was seen that the models with 16 neurons in the hidden layer showed more successful performance than the other models.

Table 2. Successful NARX Model Prediction Accuracy

PS	İstasyon	GKNS	MSE	İterasyon	Eğitim R^2	Validasyon R^2	Test R^2	Genel R^2
9	Afyon AAT Sonrası	16	6,3543463	4	0,7387013	0,8025744	0,7926454	0,7588552
9	Araplı Deresi	16	4,6284336	4	0,7722549	0,7773032	0,807465	0,7773427
9	Bolvadin Köprüsü	16	19,346285	4	0,970648	0,773941	0,7618822	0,8569918
9	Eber Gölü Kocakandıralık	16	20,125312	3	0,961912	0,9380015	0,9707762	0,7782892
9	Eber Gölü Ortası Donbay	16	8,3159556	4	0,8387508	0,5496063	0,7966091	0,8051873
9	Eber Gölü Yalı Mevkii	16	3,3037464	4	0,8503067	0,7901705	0,7836038	0,8153727
9	Eber Regülatörü	16	8,2644293	3	0,7602356	0,7486304	0,7761003	0,7577016
9	Şeker Fabrikası Sonrası	16	3,7874824	4	0,7444892	0,8049507	0,7723141	0,7743452

The dissolved oxygen value used in the study is not only estimated by taking the values of other parameters, but also by taking the past dissolved oxygen values in the same season as time series, a hybrid approach has been put forward in solving the problem.

The model and approach put forward in this study will be supported with strong rule-based statistical methods for the prediction of water quality in the current situation and it will be possible to base on artificial intelligence based decision support systems that can learn over time and produce better results. However, the hybrid approach, including the time series feature, can be used in conjunction with the parameters associated with other water quality parameters.

The model developed in this study can be used in assessing the accuracy and validity of data from field work as well as predictive ability in DSSs (Decision Support Systems). Thus, at the sampling points studied, it can help to automate a stronger control mechanism. In the case that there is no intervention within the basin, future measures on the basis of parameters can be foreseen.

Predictive models can be created not only for dissolved oxygen value, but also for other water quality parameters using ANNs based on the basic NARX model. It is also possible to produce more accurate results by making modifications on the presented models without changing the default properties of the used NARX model. In addition to the Levenberg-Marquardt Algorithm, alternative learning algorithms could be tried instead of sigmoid and linear activation functions in hidden and output layers as well.