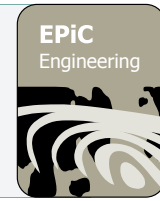




EPiC Series in Engineering

Volume 3, 2018, Pages 772–779

HIC 2018. 13th International
Conference on Hydroinformatics



Comparison of Accuracy of Artificial Neural Network (ANN) and Kriging Methods for *Estimating* Chlorine Concentration

Faezeh Ghalenoei¹, Hamed Mazandarani Zadeh^{2*} and Dragan Savic³

¹ Master Student of Water Sciences and Engineering Department, Faculty of Engineering and Technology, Imam Khomeini International University, Qazvin, Iran

² Assistant Professor of Water Sciences and Engineering Department, Faculty of Engineering and Technology, Imam Khomeini International University, Qazvin, Iran

³ Professor of Center for Water Systems; University of Exeter, Exeter, United Kingdom

mazandaranzadeh@eng.ikiu.ac.ir

Abstract

Groundwater is one of the major sources of fresh water. Maintenance and management of this vital resource is so important especially in arid and semi-arid regions. Reliable and accurate groundwater quality assessment is essential as a basic data for any groundwater management studies. The aim of this study is to compare the accuracy of two Artificial Neural Network (ANN) and Kriging methods in predicting chlorine in groundwater. In case of ANN, we created an appropriate emulator, which minimize the prediction error by changing the parameters of the neural network, including the number of layers. The best Kriging model is also obtained by changing the variogram function, such that the Gaussian variogram has the least error in interpolation of the amount of chlorine. To evaluate the accuracy of these two methods, the mean square error (MSE) and Coefficient of determination (R²) are used. The data set consists of the amount of chlorine, in a monthly basis, measured at 112 observation wells from 1999 to 2015 in aquifer Qazvin, Iran. MSE values for ANN and Kriging are 14.8 and 15.4, respectively,

which indicate that the ANN has a better performance and is more capable of predicting chlorine values in comparison with Kriging.

Key word: Artificial Neural Network, Groundwater Pollution, Kriging methods, Prediction, Quality Control

1. Introduction

Groundwater is one of the most valuable resources and it has become a dependable source of water in all climatic regions of the world (1). Management ground water is very important because the demand for groundwater has increased over the years and this has led to water scarcity in many parts of the world.

The qualitative data plays a key role in the management and modeling of groundwater (2), as the assessment of groundwater level changes. Therefore, to improve the management of groundwater resources, it is necessary to understand the performance and natural behavior of aquifer to investigate the effect of drainage and nutrition as well as the variations in the level of the water and eventually water quality changes. Qazvin plain is one of the highland plains among the plains of the country (3). Groundwater as the only reliable source is going away, because of long-term plans for exploitation of water resources from outside the basin, and decrease in water rights. The accurate aquifer status is one of the most critical issues in the management of the province's resources, the aquifer of Qazvin Plain was investigated and analyzed as a study area.

The purpose of the study is to compare Kriging and artificial neural network as interpolating techniques to find the best method by which estimating the spatial variations of the chlorine concentration (4).

2. Material and Methods

2.1. Case study

The study area Figure 1 located in the north-west of Iran with an area of 3918.79 km². The aquifer has a cold but dry climate with the annual mean temperature of 13°C and the annual precipitation is about 320mm. Monthly groundwater chlorine concentration was available for 112 observation wells from 1999 to 2015. Based on the monthly average ground water chlorine concentration.

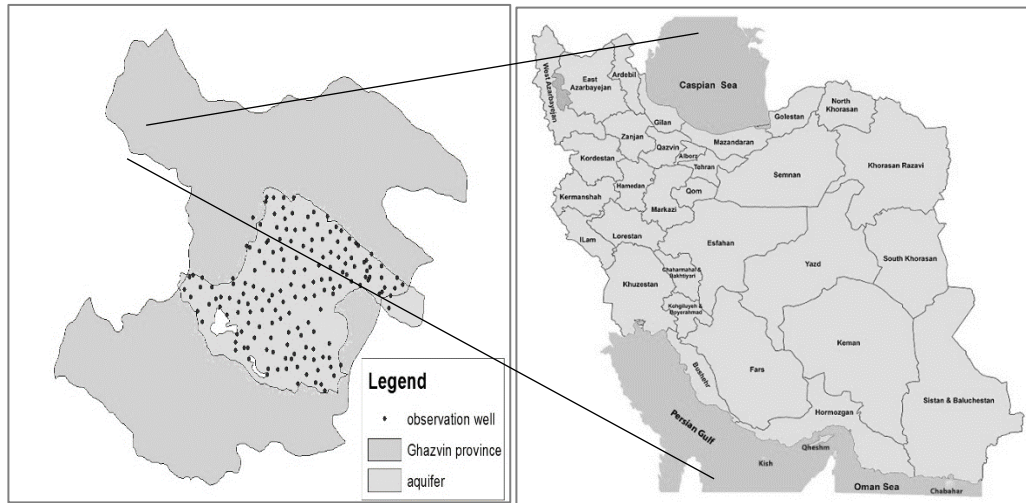


Figure 1: Case study

2.2. Artificial Neural Network (ANN)

An artificial neuron network (ANN) is a computational model based on the structure and functions of biological neural networks (5). The flow of information through the network affects the structure of the ANN, because a neural network changes - or learns, in a sense - based on that input and output. ANNs are considered nonlinear statistical data modeling tools where the complex relationships between inputs and outputs are modeled or patterns are found. An ANN has several advantages but one of the most recognized is the fact that it can actually learn from observing data sets. In this way, ANN is used as a random function approximation tool. ANNs have three layers that are interconnected (4). The first layer consists of input neurons. Those neurons send data on to the second layer, which in turn sends the output neurons to the third layer. Training an artificial neural network involves choosing from allowed models for which there are several associated algorithms. Geographic coordinates (include latitude and longitude) and chlorine concentration of observation wells was considered for input layer, which groundwater chlorine concentration predicts in the Output layer. In this study the number of hidden layers, the number of nodes and the transfer function obtained by trial and error. Also, the Leven berg-Marquardt algorithm is

selected as the learning algorithm. Figure 2 shows a typical feed forward neural network.

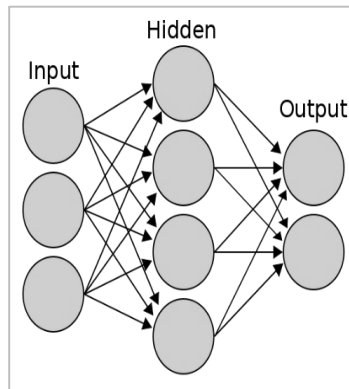


Figure 2: Typical feed forward neural network

2.3. Kriging

In statistics, originally in geostatistics, kriging or Gaussian process regression is a method of interpolation for which the interpolated values are modeled by a Gaussian process governed by prior covariances, as opposed to a piecewise-polynomial spline chosen to optimize smoothness of the fitted values (6). Under suitable assumptions on the priors, kriging gives the best linear unbiased prediction of the intermediate values. The first step is to calculate experimental semivariograms which express the spatial dependence between neighboring observations (7). The semivariograms, $\gamma(h)$, can be defined as one-half the variance of the difference between the attribute values at all points separated by h as follows (7):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(x_i) - Z(x_i + h))^2 \quad (1)$$

Where $Z(x_i)$ indicates the magnitude of variable, and $N(h)$ is the total number of pairs of attributes that are separated by a distance h . The experimental semivariograms were

fitted with various theoretical models like spherical, exponential, Gaussian, linear and power by the weighted least square method (6). The adequacy and validity of the developed semivariograms (8). Model was tested

satisfactorily by a technique called cross validation the most appropriate semivariograms was chosen based on the highest correlation coefficient (R^2) and lowest Mean Squares Error (MSE) of data test.

2.4. Evaluation criteria

2.4.1 Mean Square Error (MSE)

The mean square error is a model evaluation index that is often used with regression models. The forecast error is the difference between the correct value and the predicted value for a sample. The MSE measures the average of the square of the forecast error in a data set. The MSE can be calculated based on observed data as equation (2):

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2 \quad (2)$$

Where O_i is the observational data, P_i is the predicted data and N is the number of data available.

2.4.2 Coefficient of determination (R^2)

The R-square, is a statistical measure of how close the data are to the fitted regression line. The higher the value of R-squared, the higher the quality of the fitting is.

3. Results and Discussion

At first, the data were divided into two groups of training and test for both methods. The training data was used to calibrate the models to predict the amount of chlorine to achieve the minimum error and the test data used to evaluate the trained models.

3.1 The results of the Kriging method

In the first step, the fitting of the Kriging model was investigated for various variogram functions. Using the trial and error method, the Gaussian function was selected as the best function, and the Kriging model was calibrated using training data. To evaluate the calibrated model, the amount of chlorine was calculated for the test. The forecasted data obtained as the results of the Kriging method (simulate data), are compared with the measured data (main data). The regression line is computed regarding the relationship between simulate data and main data, as depicted in Figure 3.

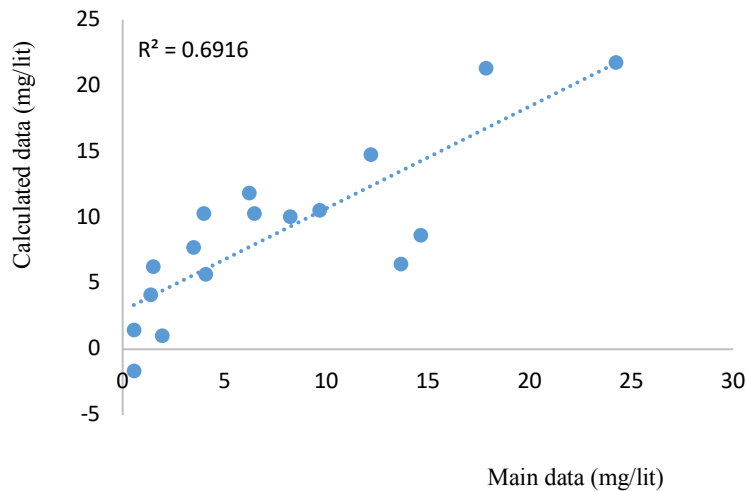


Figure 3: Regression line between measured and calculated data using kriging for testing data series

3.2 Artificial neural network method results

In this method, 70% of the data was used to train the neural network and the rest was considered for network evaluation. Initially, the network was implemented for different transfer functions in the hidden layer, output layer and two training algorithms. The results showed that the neural network with the sigmoid tangent transfer function in the hidden layers, with 15 neurons in the first layer and 10 neurons in the second hidden layer and the linear function in the output layer have the best estimate.

The evaluation criteria for both data sets in the Kriging model and the neural network is presented in Table 1.

Method	R ²	MSE
Kriging	0.6916	42.386
Artificial neural network	0.8383	24.481

Table 1: The statistical results are calculated for both data sets in the Kriging model and the neural network

4. Conclusion

The purpose of this study is to compare Kriging and artificial neural network as interpolating techniques to find the best method by which estimating the spatial variations of the chlorine concentration. To evaluate the accuracy of these two methods, the mean square error (MSE) and coefficient of determination (R²) are used. Regarding the values obtained for the two statistics (MSE and R²), it was concluded that the neural network has more accuracy than kriging in estimating the amount of chlorine in groundwater. Using this research, it is possible to determine the optimal location of adding new wells to improve the groundwater maps.

References

1. H. Mazandarani Zadeh, P. Daneshkar Arasteh, & M. Hashemi Tammeh. (2015). Optimal design of observation well network for ground water water table estimation using genetic algorithm.
2. M. Karamooz, R. Hashemi Aliya, A. Moridi, A. Ahmadi. (2006). Quantitative and qualitative simulation of surface and underground waters - The case study of Isfahan Lanjanat plain.
3. M. Hashemi Tameh, H. Mazandarani Zadeh, P. Daneshkar Arasteh. (2016). A Novel Solution to Define the Optimum Number and Location of New Wells to Improve Groundwater Level Map.
4. H. Mazandarani Zadeh, P. Daneshkar Arasteh, M. Hashemi Tameh. (2015). Offering accurate method for forecasting groundwater level.

5. M. Kholghi, SM. Hosseini . (2006). Estimation of aquifer transmissivity using kriging, artificial neural network, and neuro-fuzzy models.
6. N. Saberi . (2010). Study of location variations of groundwater quality in Gorgan city for drinking water using geo- statistics.
7. Fisher, Jason C . (2013). Optimization of Water-Level Monitoring Networks in the Eastern Snake River Plain Aquifer Using a Kriging-Based Genetic Algorithm Method.
8. Ch. Jing. .(2013). Optimization Design of Groundwater Level Monitoring Network Based on the Geo-Statistics.