

Application of a Neural Network Technique for Prediction of the Water Quality Index in the Dong Nai River, Vietnam

Nguyen Hien Than¹, Che Dinh Ly¹, Pham Van Tat² and Nguyen Ngoc Thanh³

1. Department of Natural Resources Management, Institute of Environment and Resources, Vietnam National University, HCMC 700000, Vietnam

2. Faculty of Science and Technology, Hoa Sen University, HCMC 700000, Vietnam

3. Institute of Environmental Science, Engineering and Management, Industrial University of Ho Chi Minh City 700000, Vietnam

Abstract: Recent trends in environmental management of water resource have enlarged the demand for predicting techniques that can provide reliable, efficient and accurate water quality. In this case study, the authors applied the Artificial Neural Networks (ANN) to estimate the water quality index on the Dong Nai River flowing through Dong Nai and Binh Duong provinces. The information and data including 10 water quality parameters of the Dong Nai River at 23 monitoring stations were collected during the recorded time period from 2010 to 2014 to build water quality prediction models. The results of the study demonstrated that the Water Quality Index (WQI) forecasted with GRNN was very significant and had high correlation coefficient ($R^2 = 0.974$ and $p = 0.0$) compared to the real values of the WQI. Moreover, the ANN models provided better predicted values than the multiple regression models did.

Key words: Artificial Neural Networks, water quality forecast, water quality prediction.

1. Introduction

As far as water environment is concerned, water quality is a multivariate phenomenon. Today, one of the challenges of water quality control is to minimize the cost and make more intelligent system assisting tools in water quality assessment [1]. The ANN is one of the useful tools to solve almost all those differences that have just applied in the environmental fields over the recent years.

The concept of artificial neurons was first known in 1943 [2]. The overall structure of artificial neural networks is biologically inspired by the human brain [3, 4]. Artificial neural network systems are able to perform arithmetic operations like the nervous system [5]. The ANN approach has an array of advantages over traditional phenomenological paradigms, since they require known input data set without any assumptions

[6]. The ANN has the power to approximate any non-linear relationship that exists between a set of inputs and their corresponding set of outputs [7]. Artificial neural networks not only have the ability of computing, processing, predicting and classifying data, but also have the advantages of nonlinearity, input-output mapping, adaptability, generalization and fault tolerance [8]. Applications of ANNs in the areas of water engineering, ecological sciences, and environmental sciences have been reported since the beginning of the 1990s [9, 10]. In water resources management, the ANN has proved to be a powerful tool for water quality assessment and classification [11]. It's also used to predicting various ecological processes and phenomenon related to water resources [5, 12, 13]. In recent years, there have been a board range of applications of the ANN to predict water quality [4].

Juahir, et al. [1] investigated the water quality index estimation of the Langat river basin based on independent variables such as DO, COD, BOD, SS,

Corresponding author: Nguyen Hien Than, Ph.D. student, main research field: environmental quality assessment.

N-NO₃⁻ and pH. Kanani et al. [14] applied a multi-layer perceptron neural network to predict total dissolved solid, as well as dissolved oxygen and predicting biochemical oxygen demand using pH, TS, T-Alk, T-Hard, Cl⁻, PO₄³⁻, K⁺, Na⁺, N-NH₄⁺, N-NO₃⁻ and COD [6]. Similarly, Abyaneh [15] also employed the ANN to forecast DO and BOD. Almost all variables of these studies were parameters to be observed in the laboratory analysis and only focused on estimating separate pollutants of water body instead of the WQI.

As authors know, river water quality evaluation using traditional water samples for laboratory analyses is accurate but very expensive and time consuming [8]. Although the ANNs applied for the domain of hydrology and water resources research, few studies have been conducted on the prediction of water quality parameters using meteorological data and field-measured parameters. Almost all independent variables in previous studies were the parameters of water quality to be analyze in the laboratory. This study spent much time to determine samples and lead to heavy budgets for the government. If we cut down on the cost of environmental monitoring and shorten time to evaluate the state of water quality, we'll save a giant pool of money and understand promptly the water body.

The aim of this study is to explore variables of surface water quality collected to meteorological data and field-measured parameters, as well as finding the best ANN model to predict water quality of the Dong Nai river correctly. In addition, this study also compares the ANNs with the multiple regression models to verify their predicting ability in the condition of the long term patterns. This research is useful to control and manage surface water environment in Vietnam.

2. Material and Methods

2.1 Study Area and Monitoring Data of the Dong Nai River

The Dong Nai river is the largest national river

basin and plays an important role for the economic center of the country in southern Vietnam [16]. Dong Nai River system involves the Dong Nai mainstream and 4 major rivers: the La Nga river, the Be river, the Saigon and the Vam Co river. The Dong Nai river flows through 5/8 provinces of the Southern Key Economic Zone in Vietnam with a total length of 513/628 km, whereas flowing through Dong Nai province is the longest one, about 294/628 km, 46% of its total length [17].

The scope of the investigation was part of the Dong Nai river running through the territory of Dong Nai and Binh Duong provinces. These are two provinces which are the fast economic growth in Vietnam's Southern Key Economic Zone, leading to high environmental pressure on the water quality of Dong Nai river. All monitoring data were collected from Department of Natural Resources and Environment in Binh Duong and Dong Nai provinces and the meteorological data were collected in the statistical Yearbooks 2015. There were 23 monitoring stations, 4 stations in Binh Duong and 19 stations in Dong Nai for a time period of five years from 2010 to 2014 (Fig. 1). The collected data were checked outliers before inputting to prediction models. A series of the data set in 2010-2013 was used to build predicting models and the data set of 2014 was exploited to forecast water quality.

2.2 The Artificial Neural Network Prediction Method

In this paper, the authors employed two artificial neural networks to forecast the water quality including the Multi-layer Perceptron (MLP) and the Generalized Regression Neural Network (GRNN).

In the MLP model, a feed forward MLP network consists of an input layer and an output layer with one hidden layer; between each layer there are a certain number of artificial neurons [18]. In this study, input layer is 10 parameters to be measured on the Dong Nai river. The experimental data of the ANNs is implemented in the hidden layer. The number of

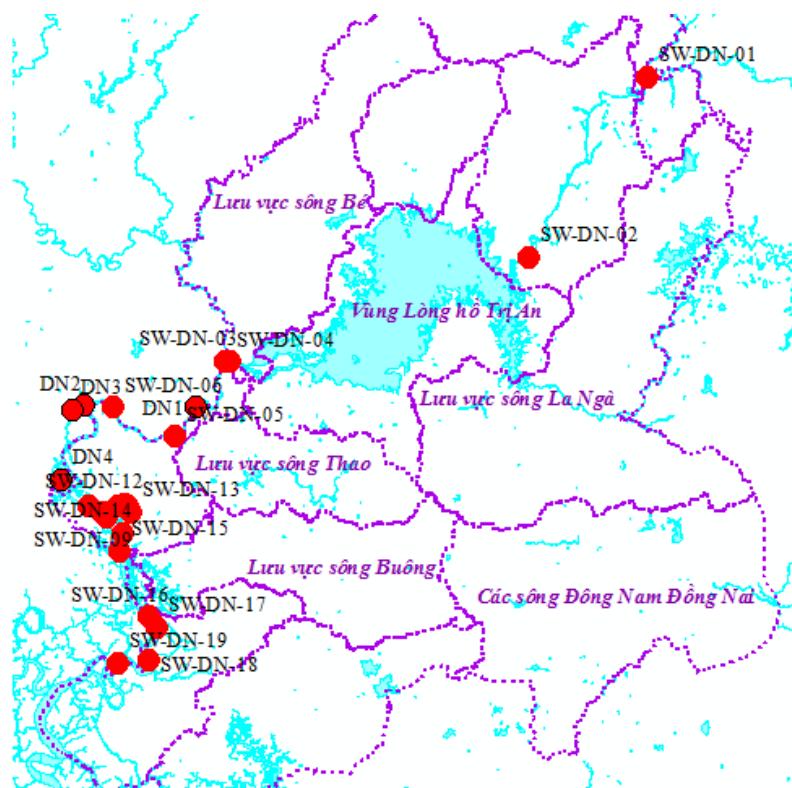


Fig. 1 Location of the monitoring stations on the Dong Nai river.

neurons in hidden layer can be determined for a specific case study. Each layer is composed of neurons, which are interconnected with each other by synaptic weights. In each neuron, a specific mathematical function called the activation function accepts input from previous layers and generates output for the next layer. Based on this process, the process of the water quality assessment is expressed in Fig. 2.

Generalized Regression Neural Networks (GRNN) were invented by Specht (1991). GRNN nodes require two main functions to calculate the difference between all pairs of input pattern vectors and estimate the probability density function of the input variables and weighting the calculated distance of any point by the probability of other points occurring in that area which yields a predicted output value [1, 9].

In this study, before the learning process both input and output variables were normalized within the range of 0-1 as Eq. (1):

$$X_{\text{normalized}} = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (1)$$

where, x is the measured value of pollutant, x_{\min} and

x_{\max} are the minimum and maximum values in the data set of this parameter, respectively.

Besides, the authors also used the multiple regression models to compare their results with the ANN models.

In this model, the dependent variable or response variable are called the independent variable [2]. Two physical models embracing the multiple linear model and the non-linear model are two traditional models with a wide range of application in the forecasting issues. The authors observed these models to predict water quality of the Dong Nai river. These regression models were established in SPSS software for windows.

2.3 The Water Quality Index (WQI)

In the ANN models, the WQI was employed with an output player for the ANN model. The WQI was applied in this study according to the Decision No.879/QĐ-TCMT dated 1 July 2011 of the Vietnam Environment Administration [19].

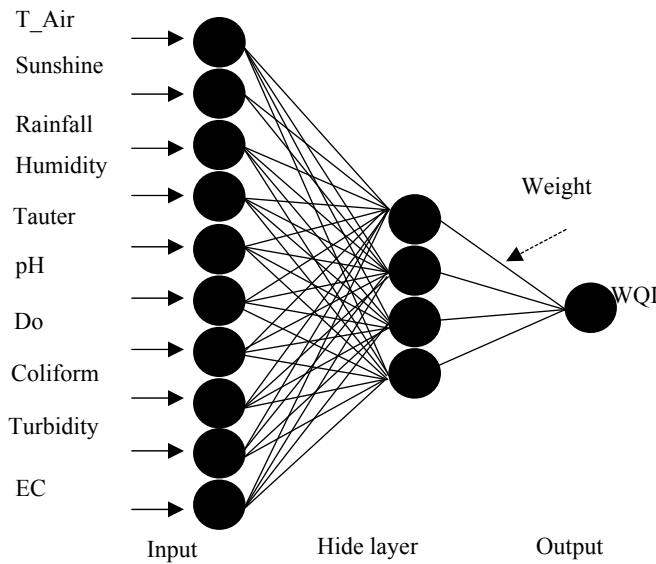


Fig. 2 The multilayer perceptron networks architecture.

2.4. Statistical Analysis

In order to validate the result of prediction model, the authors employed a variety of statistical analyzing techniques to compare the results of the study such as anova one way, correlation regression analysis and modal test.

The error was measured by Root-mean-square Error (RMSE) in Eq. (2). The RMSE is a frequently used measure of the differences between values (sample and population values) predicted by a model or an estimator and the values actually observed. The lower the RMSE of forecasting models are, the better forecasting models are.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{WQI}_{act} - \text{WQI}_{pre})_i^2}{n}} \quad (2)$$

Where, WQI_{act} is the real value of the water quality index; WQI_{pre} is predicted value of the water quality index; n is the number of case of the monitoring data.

3. Results and Discussion

3.1 Input Variables for Predicting Models

In order to choose the parameters getting involved in forecasting models, input variables were collected

based on the correlation analysis and the cluster analysis. The predictor variables were screened and selected based on the priority target, where parameters can be measured in the field, such as statistical data availability and significant impact on water quality. The initial predictors included 4 meteorological variables and 12 pollutants of water quality. It can be seen from the Fig. 3 that the pollutants of the Dong Nai river were divided into two main groups. Cluster 1 had 2 sub-group including DO and average air temperature, water temperature, humidity, sunshine hours. The authors chose all these parameters as independent variable in cluster because they were appropriate with the priority target. Similarly, the cluster 2 also had two sub-group encompassing rainfall and coliform, TSS, turbidity, BOD_5 , NO_3^- , NO_2^- , electrical conductivity (EC), NH_4^+ , COD and pH. As can be seen, it was also divided into 4 sub-sections. According to authors' preference, rainfall, coliform, turbidity, EC and pH were collected as predictors in cluster 2. Finally, ten parameters were selected enclosing x_1 = air temperature, x_2 = sunshine hours, x_3 = rainfall, x_4 = humidity, x_5 = water temperature, x_6 = pH, x_7 = DO, x_8 = turbidity, x_9 = coliform and x_{10} = electrical conductivity that were used for the input

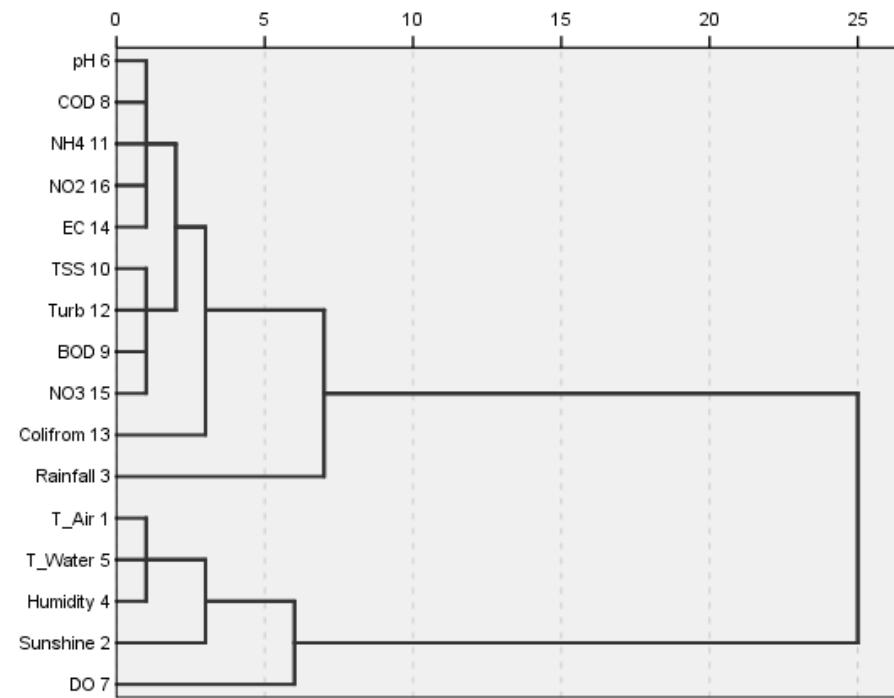


Fig. 3 Dendrogram showing clustering of water quality parameters on the Dong Nai river.

layer and WQI (Y) was a dependent variable for the output layer in the ANN models. Only coliform parameter was a variable to be measured in the laboratorial analysis.

3.2 Predicting Models of Water Quality

The results of the study built four forecasting models of the water quality.

The multiple linear regression model was derived from the enter method; the equation of this model is as follows:

$$WQI_{\text{Linear}} = 0.624 - 0.015x_1 + 0.079x_2 + 0.001x_3 - 0.019x_4 + 0.079x_5 - 0.480x_6 + 0.406x_7 - 0.687x_8 - 1.088x_9 + 0.078x_{10}$$

The results of the multiple linear regression analysis obtained R-Square = 0.698, adjusted R-Square = 0.697, p_value = 0, F = 524.438 and Durbin-Watson = 1.438, which meant that this model was statistically significant.

The multiple non-linear regression model also obtained good outcomes with R = 0.8032, p = 0 following the equation:

$$WQI = \exp(0.76893x_1 + 0.55925x_2 - 0.74576x_3 + 1.06642x_4 + 0.37844x_5 - 0.8336x_6 + 0.44273x_7 - 1.88901x_8 - 7.99988x_9 + 0.00551x_{10}) / (1 + 0.63999(4.76955x_1 + 0.91513x_2 - 3.66871x_3 + 4.18864x_4 + 0.85399x_5 - 1.44709x_6 + 0.80426x_7 - 3.8963x_8 - 9.43955x_9 + 0.43951x_{10}))$$

Besides, the two ANN models were established including 70% training data (1,593 samples) and 30% (683 samples) testing data. Network training process is carried out to achieve the best model when either a threshold minimum square error value is reached or the test set square error begins to rise. Testing dataset was randomly selected to evaluate the results of training process.

The results of the GRNN and the MLFN were displayed in Table 1:

Table 1 The results of establishing prediction models.

Prediction models	R ² training	RMSE training	RMSE testing
Linear regression	0.6970	0.1763	
Nonlinear regression	0.8032	0.1255	
GRNN	0.9950	0.0191	0.0438
MLFN (10-4-1)	0.9602	0.0212	0.0492

It can be seen from the Table 1 that the ANN models had better water quality prediction than multiple regression models. Especially, the GRNN (10-1) was the best performance because it received the lowest RMSE of 0.0191 in training and 0.0438 in testing. MLFN (10-4-1) was the second best model in four models, 0.0212 and 0.0492 of RMSE in training and testing respectively. Although the multiple linear regression and the non-linear regression obtained high correlational coefficient, a large number of the water quality indexes were not appropriate for the real value. The multiple linear regression models failed to capture the long term patterns while ANN models were successful in predicting those patterns. Moreover, it also see that training error of the GRNN obtained lower than the MLPN did. The scatter plot of actual and predicted value between training process and testing process of these models indicated that

predicted value of the GRNN was better than that of MLPN because scatter level of the GRNN model was more symmetrical than the MLPN shown in Fig. 4 and Fig. 5.

The results of water quality prediction of the Dong Nai river obtained great results. Testing regression models of the WQI running from the GRNN and calculated values were high correlation: R square = 0.9738, adjusted R square = 0.974, F = 18,424 and P = 0.0. The difference between the two results was a not statistically significant difference: p = 0.66 (accepted H_0) and $F = 12:19 < F_{\text{Crit}} = 3.85$. Obviously, the results of the water quality index between the ANN and real water quality index were similar to each other displayed in Fig. 6. Thus, the ANN models were good results both in term of statistical and real condition. Especially, the GRNN model was the best predicting model in four models observed.

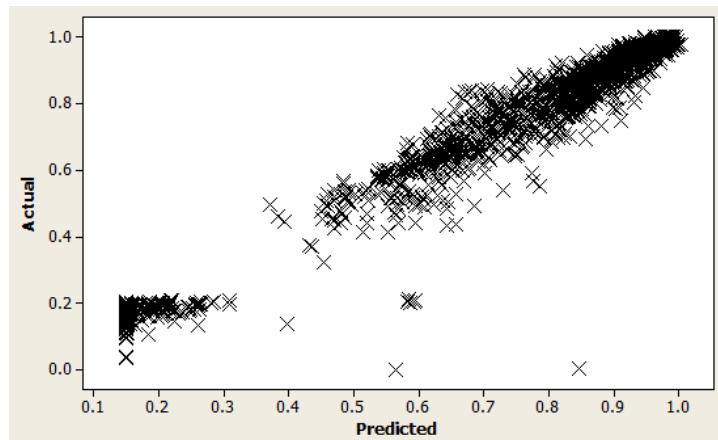


Fig. 4 Actual vs. Predicted value of the MLPN (training).

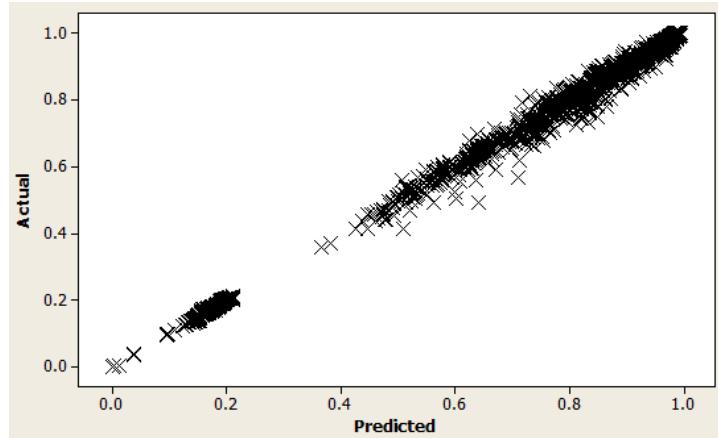


Fig. 5 Actual vs. Predicted value of the GRNN (training).

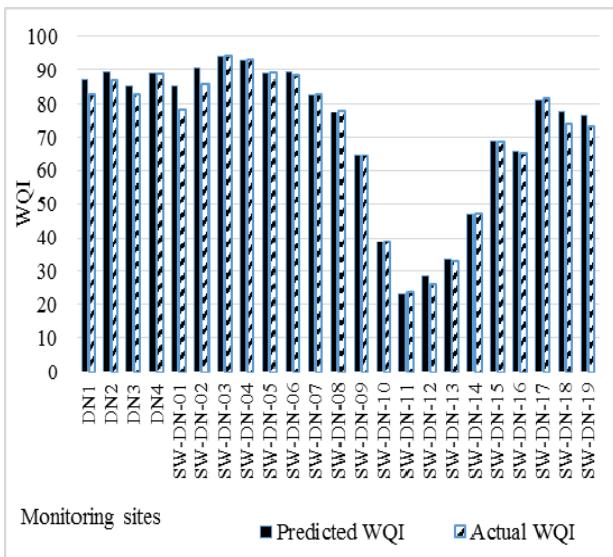


Fig. 6 The predicted WQI of the Dong Nai river in 2014.

4. Conclusion

Meteorological data and pollutants of field-measured parameters encompassing DO, air and water temperature, humidity, sunshine hours, rainfall, coliform, turbidity, EC and pH were able to forecast better the Dong Nai river water quality. The use of these variables is much significant in research and control of water quality. This not only contributes to reduce environmental fund, but also save time to evaluate the water quality, especially as our economic conditions at present is poor.

In this study, the GRNN (10-1) model and the MLPN (10-4-1) model were better predictions of water quality of the Dong Nai river than the multiple regression models. In particular, the GRNN obtained more outstanding results than the MLPN. The ANN models demonstrated tremendous potential as forecasting tools which are reasonable, suitable and reliable.

Although the multiple regression models were statistically significant, they contained higher errors than the ANN models did. The ANN models solved excellently a long term patterns of data and provided good forecasting results in phenomenal non-linear of series time that the multiple regression models couldn't overcome.

Acknowledgment

The authors would like to thank Toshiba Vietnam for scholarship support.

References

- [1] Juahir, H. Z., Sharifuddin, M., Toriman, M. E., Mokhtar, M., and Man, H. C. 2004. "Application of Artificial Neural Network Models for Predicting Water Quality Index." *Journal Kajuruteraan Awam* 16 (2): 42-55.
- [2] Sarani, N., Soltani, J., Sarani, S., and Moasher, A. 2012. "Comparison of Artificial Neural Network and Multivariate Linear Regression Model to Predict Sodium Adsorption Ratio (SAR) (Case Study: Sistan River, Iran)." *International Journal of Biological, Ecological and Environmental Sciences* 1: 2277-4394.
- [3] Patki, V. K., Shrihari, S., and Manu, B. 2013. "Water Quality Prediction in Distribution System Using Cascade Feed Forward Neural Network." *International Journal of Advanced Technology in Civil Engineering* 2 (1): 84-91.
- [4] Barzegar, R., Adamowski, J., and Moghaddam, A. A. 2016. "Application of Wavelet-Artificial Intelligence Hybrid Models for Water Quality Prediction: a Case Study in Aji-Chay River, Iran." *Stochastic Environmental Research and Risk Assessment*: 1-23. doi: 10.1007/s00477-016-1213-y.
- [5] Talib, H., Abu, Y. and Rahman, N. N. Abdul. 2009, "Predicting Biochemical Oxygen Demand as Indicator of River Pollution Using Artificial Neural Networks." In *Proceedings of the 18th World IMACS/MODSIM Congress*.
- [6] Singh, K. P., Basant, Ankita, M. A., and Jain, G. 2009. "Artificial Neural Network Modeling of the River Water Quality—A Case Study." *Ecological Modelling* 220 (6): 888-895.
- [7] Adamowski, J. and Chan, H. F. 2011. "A wavelet Neural Network Conjunction Model for Groundwater Level Forecasting." *Journal of Hydrology* 407: 28-40.
- [8] Panda, S. S., Garg, V., and Chaubey, I. 2004. "Artificial Neural Networks Application in Lake Water Quality Estimation Using Satellite Imagery." *Journal of Environmental Informatics* 4: 65-74.
- [9] Palani, S., Lioung, S. Y., and Tkachich, P. 2008. "An ANN Application for Water Quality Forecasting." *Marine Pollution Bulletin* 56: 1586-1597. doi: 10.1016/j.marpolbul.2008.05.021.
- [10] Thair, S. K., Abdul, J., Hameed, M., and M, Ayad S. 2014. "Prediction of Water Quality of Euphrates River by Using Artificial Neural Network Model (Spatial and Temporal Study)." *International Research Journal of*

Application of a Neural Network Technique for Prediction of the Water Quality Index in the Dong Nai River, Vietnam

- Natural Sciences* 2 (3): 25-38.
- [11] Farmaki, E. G., Thomaidis, N. S., and Efstatouliou, C. E. 2009. "Comparative Use of artificial Neural Networks for the Discrimination of the Water Eservoirs of Athens according to Their Elemental Content." In *Proceedings of the 11th international Conference on Environmental Science and Technology*, 254-263.
- [12] Gorashi, F. and Abdullah, A. 2012. "Prediction of Water Quality Index Using Back Propagation Network Algorithm Case Study: Gombak River." *Journal of Engineering Science and Technology* 7: 447-461.
- [13] Zhao, Y., Guo, L., Liang, J. B., and Zhang, M. 2016. "Seasonal Artificial Neural Network Model for Water Quality Prediction via a Clustering Analysis Method in a Wastewater Treatment Plant of China." *Desalination and Water Treatment* 57 (8): 3452-3465.
- [14] Kanani, S., Asadollahfardi, G., and Ghanbari, A. 2008. "Application of Artificial Neural Network to Predict Total Dissolved Solid in Achechay River Basin." *World Applied Sciences Journal* 4: 646-654.
- [15] Abyaneh, H. Z. 2014. "Evaluation of Multivariate Linear Regression and Artificial Neural Networks in Prediction of Water Quality Parameters." *Journal of Environmental Health Science & Engineering* 12 (40): 1-8.
- [16] Ringler, C., Huy, N. V., and Msangi, S. 2006. "Water Allocation Policies for the Dong Nai River Basin in Vietnam: an Integrated Perspective." *Journal of the American Water Resources Association* 42 (6): 1465-1482.
- [17] Sở Tài nguyên và Môi trường tỉnh Đồng Nai. 2012. *Quan trắc Nước mặt trên địa bàn Tỉnh Đồng Nai năm 2012*, Đồng Nai.
- [18] Mekanik, F., Lee, T. S., and Imteaz, M. A. 2011, "Rainfall Modeling Using Artificial Neural Network for a Mountainous Region in West Iran." In *Proceedings of the 19th International Congress on Modelling and Simulation*, 3518-3524.
- [19] Tổng Cục Môi Trường. 2011. *Quyết định số 879/QĐ-TCMT về việc Ban hành Sổ tay Hướng dẫn Tính toán Chỉ số Chất lượng nước*, Hà Nội.