WATER QUALITY CLASSIFICATION BY ARTIFICIAL NEURAL NETWORK - A CASE STUDY OF DONG NAI RIVER, VIETNAM

Nguyen Hien Than\(^1\),* Che Dinh Ly\(^1\), Pham Van Tat\(^2\)

\(^1\)Faculty of Environment and Natural Resources, Thu Dau Mot University, 06 Tran Van On, Phu Hoa, Thu Dau Mot City, Binh Duong

\(^2\)Hoa Sen University, 08 Nguyen Van Trang, District 1, Ho Chi Minh City

*Email: thanh@tdmu.edu.vn

Received: 30 June 2017; Accepted for publication: 18 October 2017

ABSTRACT

The Dong Nai River is the main source of supplied water for Ho Chi Minh City, Dong Nai, Binh Duong province and other areas. However, the water quality state of the Dong Nai River has been heavily pressured by discharged sources from urban areas, industrial zones, agricultural, domestic activities, etc. In this paper, the authors employed the artificial neural network model (ANNs) to classify water quality of Dong Nai River that apply a new tool to assess water quality in Vietnam. The monitoring data were used for eight years from 2007 to 2014 with 23 monitoring stations. Two neural network models including a multi-layer perceptron (MLPNN) and a generalized regression network (GRNN) were employed to classify water quality of the Dong Nai River. The results of the study showed that GRNN and MLPNN classified excellently water quality. Optimal structure of the MLPNN was \(H_1L_0I_1\) with model error about 0.1268 while the GRNN was error about 0.00001615. Comparing the result of water quality classification between the ANNs and the fuzzy comprehensive evaluation indicated that they were in close agreement with the respective values (the accurate rate of GRNN 100% and 98.5% of MLPNN).

Keywords: artificial neuralnetwork, water quality.classification, Dong Nai River.

1. INTRODUCTION

In lately years, water pollution has become the most concerned problem in many countries around the world. The assessment of long-term water quality changes is also a challenged problem [1]. Water resource pollution will cause serious effects such as water shortages, disease, ecological unbalance, etc. They affect economic development activities and human health [2]. Thus, the management of river water quality is a major environmental challenge [3]. The water quality monitoring and assessment are some of the important tasks in environmental management. Managers and authorities are not only easy to make plans and decisions in environmental protection, but also provide more information for the community [4]. The popular methods used to evaluate water quality as single-factor evaluation, the pollution index, the water quality index, fuzzy evaluation, multivariate analysis, and the artificial neural network method.
In relation to the single-factormethod, the concentration of all pollutants and parameters are evaluated in parallel with their water quality standard and the pollutionlevel depends on the score of the worst index [5]. This method is relatively simple and easy to operate but it does not reflect the comprehensive water quality [5, 6]. The pollution index method can give a quantitative description of water quality which is low, heavy or extremely heavy but it can’t distinguish functional categories of water [6]. To improve these disadvantage, the water quality index was initially proposed by Horton in 1965 and since then, a great deal of consideration has been given to the development of methods [7]. This index has been considered as the criteria for water classification. It is also a numeric expression employed to transmute immense quantities of body water description data into an index, which characterizes the presenting water quality degree [8]. Available water quality indexes, however, have some limitations such as incorporating a limited number of water quality variables and providing deterministic outputs [9].

Besides, the fuzzy comprehensive evaluation has been developed in a few decades ago. Fuzzy set theory concepts can be useful in water quality modeling, as they can provide an alternative approach to deal with problems in which the objectives and restrictions are not well defined or information about them is not precise [10]. The fuzzy comprehensive evaluation has verified to resolve difficulties of fuzzy boundaries and control the influence of monitoring mistakes on valuation results effectively. This method is accurate and objective, therefore the evaluation results can be guaranteed [11].

The artificial neural network (ANN) model is considered as a potentially beneficial technique for sophisticated non-linear phenomena [12, 13]. The theory of artificial neurons was first proposed by Warren M. and Walter P. in 1943 [14]. The general structure of the artificial neural networks is biologically stimulated by the human brain [12]. The most common network architectures, namely multi-layer perceptron, radial basis function and Kohonen’s self-organizations maps, were successfully applied to many fields [15]. Artificial neural network techniques with the ability to evaluates multi-variate data of water quality by virtue of what a complex visualization capability can provide an substitution to present traditional models.

The aim of this investigation seek an automated methodology and a superior ANN model that can quickly and efficiently classify the water quality of Dong Nai River. Besides, this study also compared the capable of water quality classifications between a multi-layer perceptron neutral network (MLPNN) and a generalized regression neutral network (GRNN). This study contributed to improving plan and management source of water quality in Vietnam.

2. MATERIAL AND METHODS

2.1 The data of the study

The use of data in this study was collected from 23 monitoring sites during the recorded time period from 2007 to 2014 in Departments of Natural Resources and Environment of Dong Nai and Binh Duong provinces. Four stations were located in Binh Duong provinces that collected one every two months and monthly collected surface water samples in Dong Nai Environmental Monitoring Center. Monitoring stations have been divided into five sections: Section 1 from SW-DN-01 to SW-DN-02 observes water quality of Dong Nai River at upper stream (agriculture area), section 2 from SW-DN-03 to SW-DN-06 monitors water quality before running Bien Hoa City, section 3 from SW-DN-07 to SW-DN-15 assesses impacts on urban and industrial activities, section 4 from SW-DN-16 to SW-DN-19 keeps watch on water
quality in industrial zone (Long Thanh, Nhơn Trach), and section 5 from DN1 to DN4 evaluates effects of industrial activities on Bình Dương province.

2.2 Research methodologies

2.2.1 The water quality index (Decision no. 879/2011/QD-TCMT)

In 2011, the Vietnam Environment Administration proclaimed the Decision no. 879/2011/QD-TCMT to calculate the water quality index. In this study, the water quality index was calculated from 9 parameters: BOD₅, COD, N-NH₄, turbidity, TSS, Coliform, saturation DO%, pH and water temperature. Using the scheme benchmark of Decision no. 879 to determine the water quality index value to compare and evaluate the water quality level: [0-25; 26-25; 51-75; 76-90; 91-100] = [I; II; III; IV; V]

2.2.2 Grade of water quality by artificial neural network

The authors have investigated the ability to classify the water quality of Dong Nai river using two kinds of GRNN and MLPNN based on surface water quality standards of Vietnam. The process of the water quality classification is showed in Fig.1:

Step 1: Define the parameters involved in the classification models. In this study, the parameter in the ANN models was used to classify water quality being the same parameters in the WQI.

Step 2: Creating a numbers of standard sample data based on the Water Quality Standard. The sample data was generated randomly using Excel 2016 software (Data> Data analysis> Random Number Generation).

Step 3: Standardizing the sample data created by the following formula: \( X_n = \frac{(X - X_{\text{Min}})}{(X_{\text{Max}} - X_{\text{Min}})} \)

Step 4: Training: input and output data for the model were displayed in the Figure 2.

Step 5: Evaluate the training model.

Step 6: Implement the water quality hierarchy based on the established network model.

To increase objectivity for comparing the degree of pollution of water by ANNs. Training and test data are set up as individual files and are used in the same way for all water quality grading methods by GRNN, MLPNN and WQI.

![Figure 1. Process of water quality classification using ANN.](image)

![Figure 2. The structure of the data.](image)
3. RESULTS AND DISCUSSION

3.1 Building standard data

800 samples of training and testing data were randomly generated from Excel 2016 software. The number of samples produced corresponding to each pollution level in the water quality standard was 160 samples for each parameter.

3.2. Model of water quality classification by artificial neural network

The generated data is divided into two data sets following training and testing rate were 70:30, which means that 560 samples were used for network training and 240 samples using for testing. The data sets were standardized before implemented the ANN training. The results of building models showed that the MLPNN obtained the optimal network structure were 9 input variables, 4 hidden nodes and 1 output (H4i2O1) because this model gives the minimum root-mean-square error (RMSE). The GRNN model gave very good performance and very low error of the model. Besides, the training time of this model was very fast. The results of the network training for classifying water quality in Dong Nai River showed that the both GRNN and MLPNN were good performance. The RMSE error of the GRNN is near zero with 0 % bad predictions. The model testing error is also close to zero, and the percent of bad predictions rate was 0 %. The MLPNN model (H4i2O1) was a training error of 0.01268 and the share of the false predictions is 0 %. The testing model obtained RMSE of 0.02243 with 0 % bad prediction rate. Thus, the GRNN model was better than the MLPNN model.

3.3 Compare the classification results between WQI and ANN

Spearman correlation coefficient used to evaluate the relationship between the models. In this study, therefore, it is used to compare the correlation between water quality classification models. The GRNN and MLPNN models were high correlated with all models (Table 1). Especially, GRNN, MLPNN values and actual values were the highest correlation. In contrast, WQI and actual value were lower correlation.

| Table 1. Pearson correlation between ANN, WQI and actual values ($\alpha = 0.01$). |
|---|---|---|---|
| GRNN | MLPNN | WQI | Actual |
| Pearson Correlation | 1 | | |
| Sig. (2-tailed) | | | |
| MLPNN | Pearson Correlation | 1.000 | 1 |
| Sig. (2-tailed) | 0.000 | | |
| WQI | Pearson Correlation | 0.927 | 0.927 | 1 |
| Sig. (2-tailed) | 0.000 | 0.000 | |
| Actual | Pearson Correlation | 1.000 | 1.000 | 0.927 | 1 |
| Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | |

Evaluating the accuracy of each model based on the confusion matrix was shown in Table 2 to Table 5. The results indicated that the GRNN and MLPNN classified accuracy 100% of water quality level. The water quality evaluation using the WQI only obtained 63% correct water quality level. The Kappa coefficient also demonstrated that the GRNN and MLPNN were good water quality classification.
The study also used SPSS 18.0 to verify the differences of two GRNN and MLPNN. The results of the paired samples test between GRNN and MLPNN illustrated that both models did not differ from the actual values. On the other hand, the pairwise comparison between the WQI model and the real value obtained standard deviation = 0.64 and \( p = 0 \). This showed that the classifying results of this method were not in close agreement with the actual value.

**Table 2.** The confusion matrix of GRNN.

| Classification | Actual | Total |
|---------------|--------|-------|
| I             | 42     | 42    |
| II            | 0      | 0     |
| III           | 0      | 55    |
| IV            | 0      | 47    |
| V             | 0      | 50    |

Spearson Chi-Square = 960, \( p = 0 \), Kappa = 1
Correct rate was 100%

**Table 3.** The confusion matrix of WQI.

| Classification | Actual | Total |
|---------------|--------|-------|
| I             | 42     | 51    |
| II            | 0      | 37    |
| III           | 0      | 22    |
| IV            | 0      | 20    |
| V             | 0      | 110   |

Spearson Chi-Square = 560.6, \( p = 0.537 \), \( p = 0 \)
Correct rate was \( \frac{42 + 37 + 22 + 0 + 50}{240} = 63\% \)

**Table 4.** The confusion matrix of MLPNN

| Classification | Actual | Total |
|---------------|--------|-------|
| I             | 42     | 42    |
| II            | 0      | 0     |
| III           | 0      | 55    |
| IV            | 0      | 47    |
| V             | 0      | 50    |

Spearson Chi-Square = 960, \( p = 0 \), Kappa = 1;
\( p = 0 \), Correct rate was 100%

**Table 5.** Statistics of Paired comparison Samples

| Pair | Mean  | N | Std. Deviation | SE Mean |
|------|-------|---|----------------|---------|
| Pair 1 | GRNN  | 3.0708 | 240 | 1.38702 | 0.08953 |
|       | Actual| 3.0708 | 240 | 1.38702 | 0.08953 |
| Pair 2 | MLPNN | 3.0708 | 240 | 1.38702 | 0.08953 |
|       | Actual| 3.0708 | 240 | 1.38702 | 0.08953 |
| Pair 3 | WQI   | 3.4208 | 240 | 1.65988 | 0.10715 |
|       | Actual| 3.0708 | 240 | 1.38702 | 0.08953 |

3.4 The results of classifying water quality of Dong Nai river

From the built ANN, the authors assessed the water quality of Dong Nai river in 2014. The results were shown in that the average pollution level of the Dong Nai River was grade III - medium pollution level. An average pollution level of the Section 1-4 were III and Section 5 were II. Thus, the points were the pollution level from low to moderate pollution. Areas affected by industrial and urban activities were higher pollution than others. Using t-test two sample assuming equal variances to check the difference between the GRNN and MNGNN models indicated that there were not different among them (\( p = 0.098 \)). This illustrated that the GRNN and MNGNN models achieved the similar results.

![Result of classifying water quality in Dong Nai river in 2014.](image)
4. CONCLUSION

The results of the development of the water quality classification using artificial neural networks showed that the ANNs were good performance for water quality classification. The results of water quality classification obtained high correct. Besides, the ANNs also provided the results of water quality classification in close agreement with Vietnam's surface water quality standards. The ANNs was the useful new tool to classify pollution levels of water with flexibility and rapid assessment results. The ANNs should be applied in the environmental field to support and improve traditional methods or problems that we cannot solve before such as assessment, classification and prediction.

REFERENCES

1. Diamantopoulou M. J., Antonopoulos V. Z., and Papamichail D. M. - The use of a neural network technique for the prediction of water quality parameters of Axios River in Northern Greece, European Water 11 (12) (2005) 55-62
2. Ahmed A. K., Jafri M. Z. M., Lim H. S., Ali A. A. Z., Kussay N. M., and Anwar H. A. S. - A feed-forward Hopfield neural network algorithm (FHNNA) with a colour satellite image for water quality mapping, IOP Conference Series: Earth and Environmental Science 37 (1) (2016) 012075.
3. Juahir H., Zain S. M., Toriman M. E., Mokhtar M., and Man H. C. - Application of artificial neural network models for predicting water quality index, Journal Kejuruteraan Awam 16 (2) (2004) 42-55.
4. Juan D., González H., Luis F., Carvajal S., and Francisco M. T. B. - Water quality index based on fuzzy logic applied to the aburra river basin in the jurisdiction of the metropolitan area, Medellin 171 (2012) 50-58.
5. Yang D., Zhengn L., Song W., Chen S., and Zhang Y. - Evaluation indexes and methods for water quality in ocean dumping areas, Procedia Environmental Sciences 16 (2012) 112–117.
6. Xing Z., Fu Q., and Liu D. - Water quality evaluation by the fuzzy comprehensive evaluation based on EW method, in Eighth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE: China, 2011.
7. Boyacioglu H. and Gundogdu V. - Efficiency of water quality index approach as an evaluation tool, Ecological Chemistry and Engineering 20 (2) (2013) 247-255.
8. Reza R. and Singh G. - Application of water quality index for assessment of pond water quality status in Orissa, India, Current World Environment 5 (2) (2010) 305-310.
9. Babaei S. F., Hassani A. H., Torabian A., Karbassi A. R., and Hosseinzadeh L. F. - Water quality index development using fuzzy logic: a case study of the Karoon River of Iran, African Journal of Biotechnology 10 (2011) 10125-10133.
10. Jun J. and Yanfang X. - Assessing water quality using fuzzy theory and information entropy in Baiyangdian, China, IEEE, 2 (Water Resource and Environmental Protection (ISWREP), 2011, pp. 1047-1050.
11. Haiyan W. - Assessment and prediction of overall environmental quality of Zhuzhou City, Hunan Province, China, Environmental Management 66 (2002) 329-340.
12. Patki V. K., Shrihari S., and Manu B. - Water quality prediction in distribution system using cascade feed forward neural network, International Journal of Advanced Technology in Civil Engineering 2 (1) (2013) 84-91.

13. Barzegar R., Adamowski J., and Moghaddam A. A. - 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, 2016, pp. 1-23.

14. Sarani N., Soltani J., Sarani S., and Moasher A. - 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 (2012) 2277 – 4394.

15. Farmaki E. G., Thomaidis N. S., and Efstathiou C. E. - Comparative use of artificial neural networks for the discrimination of the water reservoirs of athens according to their elemental content, in Proceedings of the 11th international Conference on Environmental Science and Technology, Chania, Crete, Greece, 2009.