The Mapping of Water Quality Prediction using Adaptive Neuro-Fuzzy Inference System

M M Santoni¹, H N Irmanda² and R Astriratma²

¹Department of Informatics, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran, Jl. Rs. Fatmawati, Pondok Labu, South Jakarta, 12450, DKI Jakarta, Indonesia
²Department of Information System, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran, Jl. Rs. Fatmawati, Pondok Labu, South Jakarta, 12450, DKI Jakarta, Indonesia

Email: megasantoni@upnvj.ac.id

Abstract. This research proposed a mapping of water quality level prediction in China’s main watershed by using Adaptive Neuro-Fuzzy Inference System (ANFIS). ANFIS is a method that combine the advantages of Fuzzy Inference System (FIS) and Artificial Neural Networks (ANN). This research used data from The National Environmental Monitoring Center of China (CNEMC) website. Based on the result, the water quality level in 113 of 127 river nodes can be predicted correctly. The prediction results were mapped by using QGIS tools to help any stakeholders monitoring the water condition in China’s main watershed.

1. Introduction
Water is an important thing in human life. Water is used for activities or work for example regarding to agriculture, industrial, livestock, etc. Understanding water quality is necessary, so that it can be used optimally. Furthermore, water quality mapping is crucially needed to monitor water condition in an area.

Water quality measurement has an important role in environmental management. Water quality prediction can be used to help the government or any stakeholders for making a decision. The parameters that can be used to measure water quality are complicated. Moreover, the relationship between water quality parameters and the status of the water quality are hard to identified.

Some researchers have proposed a model of water quality prediction. Mehdi and Ashgar [1], used fuzzy logic approach in spatial analysis to model drinkable groundwater quality. The fuzzy approach produces a model that is better than the statistical approach. Spatial distribution mapping is able to help the stakeholder identifying drinkable groundwater supply. Caniani, et al [2] conducted a research about groundwater pollution risk measurement by using fuzzy logic approach. The result of fuzzy model can overcome the subjectivity and uncertainty of input data. The fuzzy logic approach is acceptable to model something meaningless and impressive. On the other hand, determining linguistic variables and membership function in fuzzy logic often take a long time and difficult to overcome a new situation. Based on research from Mayvilvaganan and Naidu [3] that compared fuzzy logic approach with Artificial Neural Networks (ANN) on groundwater level, ANN shows better prediction than fuzzy logic method.
Adaptive Neural Network Fuzzy Inference System (ANFIS) is proposed in this paper to predict water quality in China’s main watershed. ANFIS is a method that combine the advantages of a Fuzzy Inference System (FIS) and Artificial Neural Networks (ANN). The FIS is able to translate knowledge of experts in the form of rules but it commonly takes a long time to determine its membership function. Thus, learning method of artificial neural networks is necessary to automate the process and reduce the searching time. In other words, ANFIS method can be applied in the various field [4].

2. Data and methodology
This section explains the data and research methodology. First subsection contains data that used in this research. Second subsection explains Adaptive Neuro-Fuzzy Inference System (ANFIS).

2.1. Data
The data used in this research is obtained from The National Environmental Monitoring Center of China (CNEMC). CNEMC monitors river water quality condition in China’s main watershed such as Songhua River, Liaohe River, Haihe River, Huaihe River, Yellow River, Yangtze River, Pearl River, Taihu Lake, Chaohu Lake, Dianchi Lake, Qiantang River, and Minjian River.

The parameters that is used to measure river water quality are dissolved oxygen (DO), chemical oxygen demand (COD), and ammonia-nitrogen (NH3-N). The data is taken from monitoring data in 2013 from 51st and 52nd week with 127 river nodes. Data in the 51st week used as training data to predict river water quality in 52nd week. Overall China’s map can be seen in figure 1.

![China's map](image)

Figure 1. China’s map

2.2. Adaptive Neuro Fuzzy Inference System (ANFIS)
Adaptive Neuro-Fuzzy Inference System is an algorithm that can combine Fuzzy Inference System (FIS) and artificial neural networks (ANN). ANFIS architecture can be seen in figure 2. Based on figure 2, there are two inputs fuzzy inference system, which are $x_1$ and $x_2$, and the output is $Y$. In the Sugeno model order one, set of rules used the linear combination from the existing input, that can be expressed as follows:

Rule 1 :IF $x_1$ is $A_1$ AND $x_2$ is $B_1$ THEN $Y_1 = p_1x_1 + q_1x_2 + r_1$

Rule 2 :IF $x_1$ is $A_2$ AND $x_2$ is $B_2$ THEN $Y_2 = p_2x_1 + q_2x_2 + r_2$

In order to get average weight value from the output, it can be calculated by using equation 1.
2.3. Performance evaluation

The result of ANFIS model is evaluated by using statistical evaluations such as Correlation Coefficient (CORR), Nash-Sutcliffe Coefficient of Efficiency (NSCE), Root Mean Square Error (RMSE), and accuracy value [5]. CORR is a value that measures the correlation between two variables [6]. CORR can be calculated by using equation 2.

\[
\text{CORR} = \frac{\sum_{n=1}^{N}(A_n - \bar{A}) \times (P_n - \bar{P})}{\sqrt{\sum_{n=1}^{N}(A_n - \bar{A})^2 \times \sum_{n=1}^{N}(P_n - \bar{P})^2}}
\]

Based on equation 2, \(A_n\) is the actual risk value for \(n\)th data, \(\bar{A}\) is average of the actual risk value, \(P_n\) is the predicted risk value using ANFIS model for \(n\)th data, \(\bar{P}\) is average of the predicted risk value, and \(N\) is number of testing data.

NSCE gives information about the precision level of correlation between the predicted and the actual values [7]. NSCE can be calculated using equation 3.

\[
\text{NSCE} = 1 - \frac{\sum_{n=1}^{N}(A_n - P_n)^2}{\sum_{n=1}^{N}(A_n - \bar{A})^2}
\]

RMSE measures error that obtained from sum square of deviation between the predicted and the actual values [8]. RMSE can be calculated using equation 4.

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N}(A_n - P_n)^2}
\]

Accuracy indicates the ratio between the sum of data identified correctly with the total number of testing data. Accuracy can be calculated using equation 5.
\[ \text{Accuracy}= \frac{\text{the number of data identified correctly}}{\text{the total number of testing data}} \]  

3. Result and Discussion
The ANFIS model is used to predict the level of water quality in China’s watershed. Data in the 51st week used as training data by using ANFIS method. The model that produced from ANFIS training, is used to predict water quality level data in the 52nd week. Afterward, the prediction result is compared with the actual data by using statistical evaluation.

The prediction of water quality level is categorized into six categories, denoted as quality 1, 2, 3, 4, 5 and 6. Each value that is obtained from the ANFIS model represents the value of water quality level. The range of ANFIS values are converted into water quality level that defined in Table 1. The number of rules produced by ANFIS are eight rules.

| Table 1. Water quality risk level |
|----------------------------------|
| Water quality level | The range of ANFIS value |
|---------------------|-------------------------|
| Quality 1 | 0.00 – 1.50 |
| Quality 2 | 1.50 – 2.50 |
| Quality 3 | 2.50 – 3.50 |
| Quality 4 | 3.50 – 4.50 |
| Quality 5 | 4.50 – 5.50 |
| Quality 6 | > 5.50 |

This study used several types of the membership function, each of them has two types of linguistic variables. The experiments are conducted with various epoch values from 1 to 500. The evaluation result can be seen in Table 2.

| Table 2. The evaluation result of ANFIS using several types of membership function |
|-----------------------------------------------|
| Membership function types | Accuracy | CORR | NSCE | RMSE |
|------------------------------|----------|------|------|------|
| gbellmf                     | 85.8268  | 0.9534 | 0.9057 | 0.4066 |
| trimf                       | 85.8268  | 0.9527 | 0.9057 | 0.4066 |
| trapmf                      | 88.9764  | 0.9616 | 0.9237 | 0.3659 |
| gaussmf                     | 85.0394  | 0.8586 | 0.7216 | 0.6987 |
| gauss2mf                    | 87.4016  | 0.8726 | 0.7351 | 0.6816 |
| pimf                        | 88.1890  | 0.9571 | 0.9147 | 0.3868 |
| dsigmf                      | 86.6142  | 0.9168 | 0.8384 | 0.5324 |
| psigmf                      | 86.6142  | 0.9168 | 0.8384 | 0.5324 |

Table 2 shows that the trapezoidal membership function produces the best model evaluation result. The trapezoidal membership function is a membership function type that is often used in fuzzy logic. Based on 127 river nodes that have been tested, the best accuracy is 88.9764. It means 113 river nodes are predicted correctly and 14 river nodes are predicted incorrectly. The accuracy is sufficient which is supported by CORR, BSCE and RMSE values.

CORR value that closer to 1, indicates that there is a linear correlation between the prediction and the actual variable. The more variables can be predicted correctly, the higher the CORR value will be obtained. It also happens on the NSCE value. A good model produces an NSCE value close to 1, which means the precision level in water quality prediction is getting closer to the actual water quality condition. Meanwhile, in RMSE value, a good model produces a RMSE value close to 0. It means the level of error is imperceptible. The best result of ANFIS model prediction is mapped by using QGIS Valniera 2.2.0 tools. The overall, mapping result for each water quality level depicted in Figure 3.
Figure 3. The result mapping of water quality level

The prediction result of testing data can be seen in the confusion matrix. Confusion matrix represents the correct and incorrect prediction by using the ANFIS model (can be seen in Table 3).

| Actual | Quality 1 | Quality 2 | Quality 3 | Quality 4 | Quality 5 | Quality 6 | ∑     |
|--------|-----------|-----------|-----------|-----------|-----------|-----------|--------|
| Quality 1 | 18        | 3         | 0         | 0         | 0         | 0         | 21     |
| Quality 2 | 3         | 48        | 3         | 0         | 0         | 0         | 54     |
| Quality 3 | 0         | 1         | 25        | 4         | 1         | 0         | 31     |
| Quality 4 | 0         | 0         | 0         | 5         | 1         | 0         | 6      |
| Quality 5 | 0         | 0         | 1         | 2         | 3         | 1         | 7      |
| Quality 6 | 0         | 1         | 0         | 0         | 2         | 5         | 8      |
| ∑       | 21        | 53        | 29        | 11        | 7         | 6         | 127    |

Based on Table 3, the water quality on quality 4, 5 and 6, have small values. It means the model cannot predict correctly in that category. This condition occurs because the number of data for each category have to be the same during training. However, the ANFIS model is sufficient to predict the level of water quality in China’s main watersheds.

4. Conclusion

In this research, a prediction system and water quality mapping in China’s main watersheds were conducted. The prediction model was obtained by using Adaptive Neuro-Fuzzy Inference System (ANFIS). The prediction of water quality level is categorized into six categories. Each value that is obtained from the ANFIS model represents the value of water quality level. The number of rules produced by ANFIS are eight rules. The trapezoidal membership function produced the highest evaluation value compared to other membership functions. Based on the data that has been used, the water quality level in 113 of 127 river nodes can be predicted correctly. The best accuracy is 88.9764. The prediction results were mapped by using QGIS tools to help any stakeholders monitoring the water condition in China’s main watershed.
Reference

[1] Kord M 2014 Spatial analysis of Ardabil plain aquifer potable groundwater using fuzzy logic Journal of King Saud University - Science vol. 26(2) pp. 129–140

[2] Caniani D, Lioi D S, Mancini I M, Masi S and Sdao F 2011 Fuzzy Logic Model Development for Groundwater Pollution Risk Assessment European Water Publication 35 pp. 13–22

[3] Mayilvaganan M K 2011 ANN and Fuzzy Logic Models for the Prediction of groundwater level of a watershed International Journal Computer Science and Engineering 3(6) pp. 2523–2530

[4] Altaher A, Almomani A and Ramadass S 2012 Application of Adaptive Neuro-Fuzzy Inference System for Information Security 8(6) pp. 983–986

[5] Yan H, Zou Z and Wang H 2010 Adaptive Neuro Fuzzy Inference System for Classification of Water Quality Status Journal of Environment Science 22(12) pp. 1891–1896

[6] Mukaka M M 2012 Statistics Corner: A Guide To Appropriate Use of Correlation Coefficient in Medical Research Malawi Medical Journal 24(3) pp. 69-71

[7] Harmel R D and Smith P K 2007 Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modelling Journal of Hydrology 337(3-4) pp. 326-336

[8] Moriasi D, Gitau M, Pai N and Daggupati P 2015 Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria Transaction of ASABE (American Society of Agricultural and Biological Engineers) 58(6) pp. 1763-1785