Application of artificial neural network for defining the water quality in the river

Riyanto Haribowo[1]**, Very Dermawan**[1]**, Nevandria Satrya Yudha**[1]

[1]Water Resources Engineering Department, Universitas Brawijaya, Malang, 65145, Indonesia

riyanto_haribowo@ub.ac.id

Abstract. Predicting point and nonpoint source runoff of dissolved and suspended materials into their receiving streams is important to protecting water quality. Therefore, it is important to monitoring the condition of river water quality. The purpose of this study is to predict water quality in small streams using an Artificial Neural Network (ANN). The study focuses on small stream in tributary of Brantas River. The variables of interest are dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), pH and temperature (T). To validate the performance of the trained ANN, it was applied to an unseen data set from a station in the region. The result show that the prediction of DO is 6.03 mg/litre, pH is 6,47 mg/litre and temperature is 25.18°. With the relatively error was 15.63%, 12.64% and 14.12% respectively. It was finally concluded that ANN models are capable of simulating the water quality parameters.

Keywords: artificial neural network, river, water quality, modelling

1. Introduction

Rivers have received increasing amounts of attention due to their vicinity to large centres of population [6, 8]. Therefore, it is necessary to have reliable information on characteristic of water quality for effective pollution control and water resource management, especially in arid regions [5, 4].

In an effort to handle this problem, it is necessary to conduct accurate and efficient water quality monitoring activities as a reference to make efforts to manage the quality of river water [1, 7]. Therefore, the quality of river water can improve along with the increasing knowledge of human work more facilitated by the computer [3].

An Artificial Neural Network (ANN) is a computational method inspired by the studies of the brain and nervous system in biological organisms [9]. ANN models have been used increasingly in various aspects of science and engineering because of its ability to model both linear and nonlinear systems without the need to make any assumptions as are implicit in most traditional statistical approaches [2].

The main aim of the present work is to construct an artificial neural network (ANN) model of the Surabaya river water quality. The DO, BOD, COD, pH and temperature of the river water were taken as the dependent variables here and set of other parameters constituted the independent variables. In this study, ANN models have been identified for computing the DO, BOD, COD, pH and temperature of the river water.

2. Material and Methods

The data set used in this study was generated through continuous monitoring of the water quality of Surabaya river which is the tributary of Brantas river, the biggest river in East Java. This research was conducted on three water quality monitoring stations located in Surabaya river, namely Canggu Tambangan station, Perning Bridge and Jrebeng Bridge. Data of water quality parameter used is monthly data for 10 years (2006 -2015), and parameters used are DO, BOD, COD, pH and temperature. To calculate the discharge using monthly rainfall data for 10 years (2006-2015) from 3 nearest stations (Kemlaten, Krian, Bakalan). Field data for DO, pH and temperature were analyzed with the help of water quality checker (HORIBA U-50) device. Prediction analysis of water quality parameters by modeling the artificial neural networks using NeuroSolution 7.0 For Excel. The ANN network used for the present study is shown in Figure 1. In the modeling analysis used three scenarios that will be explained in the next section.
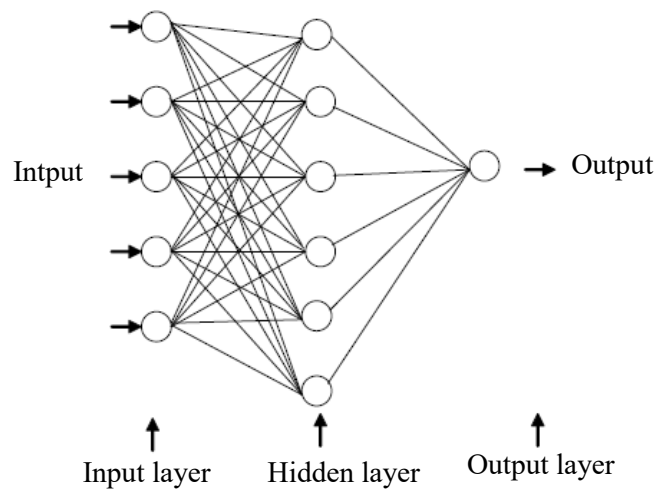
Figure 1. Structure of a multi-layer feed forward Artificial Neural Network model

The mathematical model of mass balance method can be used to determine the average concentration of downstream flow from point source and non-point sources pollutants (Figure 2).
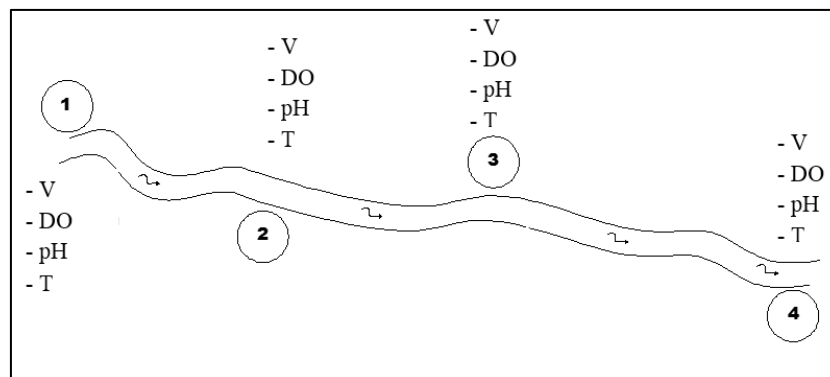
Figure 2. River Flow scheme for mass balance analysis

Information:
1. River flow before mixing with pollutant sources
2. Flow source stream A
3. Flow source stream B
4. River flow after mixing with pollutant sources.

3. Result and Discussion

3.1. Analysis of water quality using artificial neural networks (ANN)

In this study, the ANN method will be applied to predict water quality parameters (DO, BOD, COD, pH and temperature) at Jrebeng Bridge point. At the prediction stage with ANN method in this study will be made 3 (three) configuration. So it will be seen which configuration model has the best accuracy. The differences between the three scenarios lies in the input and desire variables.

3.1.1. Discussion of configuration results 1

Configuration I is conducted to predict parameters that can be measured directly in the field such as DO, pH and temperature. This is carry out to determine the current condition of water quality at the downstream point with no need to measure the downstream, only with measurement data in upstream and centre of the river. After that, also tried to predict BOD and COD only with input parameter of DO, pH and Temperature. From the results of ANN analysis in configuration 1 shows good results for DO, pH and Temperature parameters. But at the output of BOD and COD shows the RE is still high namely above 10% (Table 1). The lowest result in configuration 1 for DO, pH and temperature set in the configuration with the output of each parameter, then set in the composition of the same dataset and epoch namely the dataset 60:20:20 and epoch 5000 with RE DO = 4.50%, pH = 0.99% and Temperature =1.03%.

Table 1. Relative error for configuration I with output for each parameter

Parameter	Epoch	Relative Error (%)		
		Dataset Composition 50%:30%:20%	Dataset Composition 60%:20%:20%	Dataset Composition 60%:30%:10%
DO	1,000	10.55	8.73	9.21
	5,000	4.94	**4.50**	7.91
	10,000	10.00	6.13	9.78
BOD	1,000	**30.10**	32.95	41.72
	5,000	33.42	30.90	30.92
	10,000	30.30	40.34	36.18
COD	1,000	33.31	34.07	52.16
	5,000	40.10	36.01	42.64
	10,000	34.66	**28.53**	35.18
pH	1,000	1.15	1.17	1.29
	5,000	1.15	**0.99**	1.35
	10,000	1.44	1.12	1.20
T	1,000	1.32	1.43	1.38
	5,000	1.49	**1.03**	1.26
	10,000	1.50	1.30	1.48

3.1.2. Discussion of configuration results II

Configuration II is conducted to predict water quality in the downstream with water quality parameters, in which the measurement can be done in the laboratory, with the requirement having the data of river water quality that will be measured on the upstream and downstream of the river. For instance, we will measure BOD and COD, the requirements we must have BOD and COD data on upstream and downstream and direct measurable supporting data such as DO, pH, and temperature at 3 river sections ie upstream, middle and downstream. As well as required rainfall data from some nearby rain stations.

Table 2. Relative error for configuration II with BOD and COD output

Parameter	Epoch	Relative Error (%)		
		Dataset Composition 50%:30%:20%	Dataset Composition 60%:20%:20%	Dataset Composition 60%:30%:10%
BOD	1,000	20.44	17.58	20.48
	5,000	18.49	**15.59**	16.15
	10,000	17.43	24.66	22.38
COD	1,000	23.62	**16.91**	18.54
	5,000	31.45	27.29	21.10
	10,000	24.95	18.51	19.25

From the results of ANN analysis in configuration II shows the RE is still above 10% (Table 2). The lowest results in configuration 2 for BOD and COD are in different epoch but still in the same data set composition ie 60:20:20. For BOD Located on epoch 5,000 with a RE of 15.59% whereas COD located on epoch 1,000 with RE 16.91%

3.1.3. Discussion of configuration results III

Configuration III is performed to predict BOD or COD parameters that cannot be measured directly in the field. Analyzes were performed using input data that can be measured directly in the field such as DO, pH and temperature. This was conducted to determine the current condition of BOD and COD condition without need to bring water sample to the laboratory but only with input parameter DO, pH and temperature, so it could save time and cost. Here is an example of the best RE configuration III as shown in Table 3.

Table 3. Relative Error configuration III with BOD and COD output

| Parameter	Epoch	Relative Error (%)		
		Dataset Composition 50%:30%:20%	Dataset Composition 60%:20%:20%	Dataset Composition 60%:30%:10%
BOD	1,000	29.89	**20.67**	31.55
	5,000	32.25	33.20	50.87
	10,000	40.56	41.65	33.46
COD	1,000	30.58	32.68	**26.85**
	5,000	38.06	28.62	28.25
	10,000	27.57	31.55	36.43

From the analysis results seen in the 3rd configuration apparently, the smallest RE is in the same epoch that is 1,000, but the results obtained is still better in the 2nd configuration, this is because in the 3rd configuration RE is still above 20%. So, it can be concluded that the form of network architecture in the 2nd configuration can be used in BOD and COD forecasting. After conducted 90 times process of ANN analysis using Neurosolution for excel then the analysis result of the model with the smallest RE can be seen in Table 4.

**Tabel 4.** Selection of the Smallest Relative Error

Parameter	Smallest RE			
	Epoch	Configuration	Dataset composition Train : cros val : test	RE (%)
DO	5,000	1	60 % : 20% : 20%	4.50
BOD	5,000	2	60 % : 20% : 20%	15.59
COD	1,000	2	60 % : 20% : 20%	16.91
pH	5,000	1	60 % : 20% : 20%	0.99
Temperature	5,000	1	60 % : 20% : 20%	1.03

### 3.2. Comparison between model results and actual data

Result of model and actual data analysis for DO, pH and temperature parameter at smallest RE located in the epoch and configuration as well as the same data set composition, namely epoch 5,000 configurations 1 and data set composition 60:20:20. Comparison between model of analysis result and actual data for DO, pH and temperature parameter can be seen in Figure 3-8. The analysis shows that the model is already approaching the actual data, so it can be concluded that the model can be used to predict water quality condition.
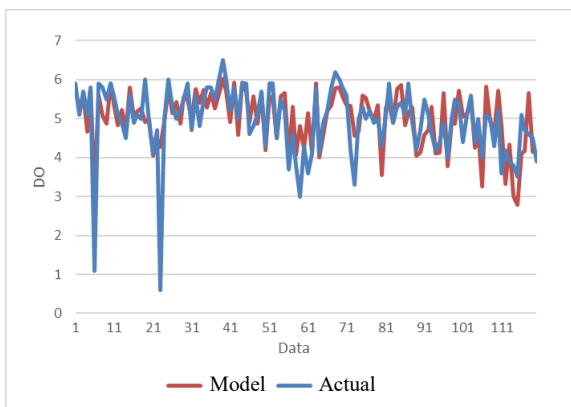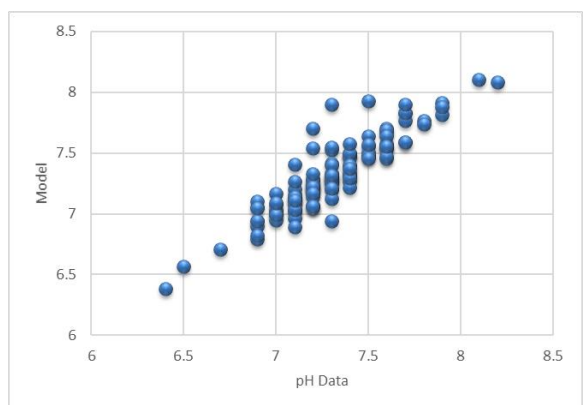

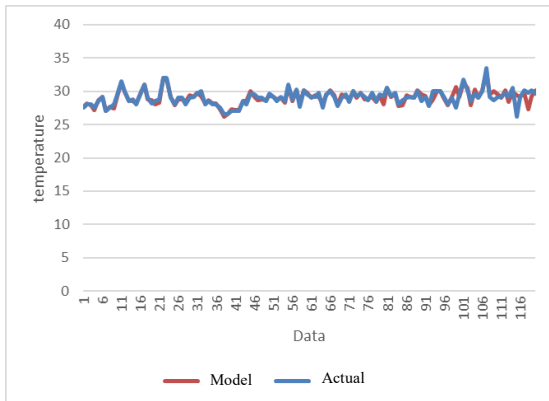**Figure 3.** Comparison between DO actual and model
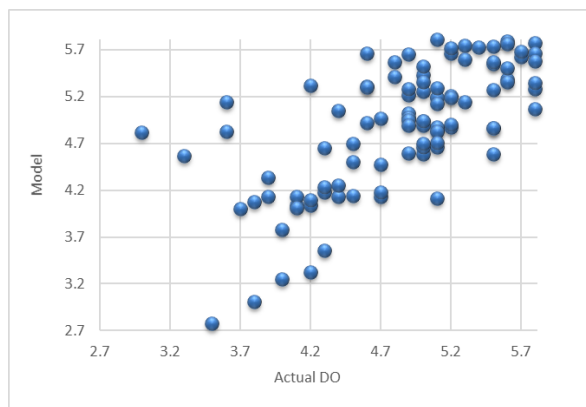
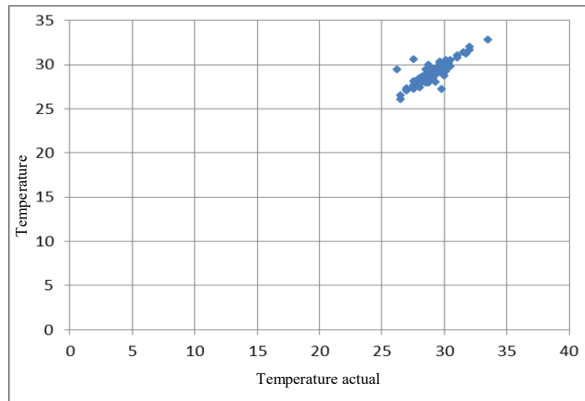**Figure 4.** Distribution of DO actual and model
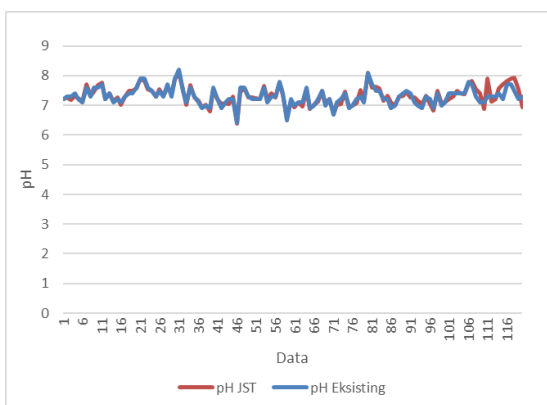
Figure 5. Comparison of pH actual and model

Figure 6. Graph distribution of pH actual and model

Figure 7. Comparison between temperature actual and model

Figure 8. Graph of data distribution the output temperature of ANN and actual temperature

3.3. Water Quality Prediction

For water quality prediction, the next conducted is to add the data of direct measurement results in the field which then put as input data and desired data using one of the running results of the parameter with the best output that has the smallest RE %. From the analysis result, it is known that the prediction of DO, pH and temperature conditions is 6.03 mg/litre, 6.47 and 25.18°C. Meanwhile, the differences of the result between model and field data for DO, pH and temperature respectively were 1.12, 0.73 and 4.14, with an average RE of 14.14% (Table 5).

Tabel 5. Comparison of Model and Measurement Data in the Field

Point	Parameter	Value		Difference	KR %
		Model	Field		
Jembatan Jrebeng	DO (mg/litre)	6.03	7.15	1.12	15.63
	pH	6.47	5.74	0.73	12.64
	Temperature (°C)	25.18	29.32	4.14	14.12

4. Conclusions

The smallest RE value of DO amounted 4.50%, pH 0.98% and Temperature 1.0267% located in configuration 1 with target of each parameter using epoch 5,000 and dataset composition of training 60%, cross validation 20% and testing 20%. For the smallest RE of BOD 15.58% and COD 16.90% with configuration model 2, epoch 5,000, and dataset composition of Training 60% - cross validation 20% - testing 20%. Predicted result value for DO parameter is 6.03 mg/litre and the field measurement value are 7.15 mg/litre has a difference of 1.12 mg/litre with a RE of 15.63%. The predicted value of ANN for pH amounted 6.46 and the field measurement value 5.74 has a difference of 0.73 with a RE of 12.64%. The predicted value of ANN for temperature amounted 25.18°C and the field measurement value 29.32°C has a difference of 4.14°C with a RE of 14.12%. The analysis results show that with an average RE rate of less than 15%, the ANN model can be used to predict water quality conditions, which will make it easier to predict water quality conditions for better river management. Suggestions for future research to use more data from multiple stations to obtain more accurate results and represent the overall river condition.

References

[1] Boucher, M.A., Perreault, L., Anctil, F., 2009. Tools for the assessment of hydrological ensemble forecasts obtained by neural networks. J. Hydroinform. 11 (3e4), 297e307.

[2] Chang, Y.-T., Lin, J., Shieh, J.-S., Abbod, M.F., 2012. Optimization the initial weights of artificial neural networks via genetic algorithm applied to hip bone fracture prediction. Adv. Fuzzy Syst

9.

[3]   Khalil, B., Ouarda, T.B.M.J., St-Hilaire, A., 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. J. Hydrol. 405, 277e287.

[4]   Kisi O., Ay M., 2014. Comparison of Mann–Kendall and innovative trend method for water quality parameters of the Kizilirmak River, Turkey J. Hydrol., 513 (2014), pp. 362-375.

[5]   Motaghian H.R., Mohammadi J., 2011. Spatial estimation of saturated hydraulic conductivity from terrain attributes using regression, kriging, and artificial neural networks Pedosphere, 21 (2) (2011), pp. 170-177.

[6]   Nilsson C., Reidy C.A., Dynesius M., Revenga C., 2005. Fragmentation and flow regulation of the world's large river systems. Science, 308 (2005), pp. 405-408.

[7]   Riyanto H, Minami Y, Masahiko S, Tsuyoshi I, Koichi Y, Takaya H, Ariyo K,. 2017. Behavior of toxicity in river basins dominated by residential areas. Contemporary Engineering Sciences, 10 (7)(2-13), pp. 305-315.

[8]   Wang X.L., Lu Y.L., Han J.Y., He G.Z., Wang T.Y., 2007. Identification of anthropogenic influences on water quality of rivers in Taihu watershed. J. Environ. Sci. (China), 19 (2007), pp. 475-481.

[9]   Yegnanarayana, B. (1999). Artificial Neural Networks. Prentice-Hall of India Pvt. Ltd., New Delhi.