Prediction of Chemical Oxygen Demand In Dondang River Using Artificial Neural Network

Anita Talib and Mawar Idati Amat

Abstract—**The increase of human population every year results in increasing of water usage and demand. This phenomenon affects the overall water quality and change the flow of river ecosystems. Various residential and commercial building were built along Dondang river basin in Pulau Pinang. Irresponsible residents and factory manager treat the river as sites for illegal garbage dumping and waste disposal. As an effort to monitor river water quality in Malaysia, Alam Sekitar Malaysia (ASMA) collaborated with Department of Environment (DOE) Malaysia to measure water quality parameters every month. The data used for Dondang River was dated from January 1998 until December 2007 at three stations at different sites. Different variables in this model include Dissolve Oxygen (DO), Biological Oxygen Demand (BOD), Suspended Solid (SS), pH, ammonia (NH_3), temperature, nitrate (NO_3), total solid (TS) and phosphate (PO4).were used as inputs while the prediction is made for Chemical Oxygen Demand (COD). The number of hidden layer, processing elements, the value of learning rule and transfer were adjusted to achieve the minimum error and highest prediction accuracy. The prediction of COD were divided into training, cross validation and testing data. The best possible model estimation was choose from the model with higher generalization, less minimum square error and high r squared value. The model estimation then tested with Mean Sensitivity test to check the validity of the model and to determine the factors that affect the value of COD the most.. Sensitivity analysis showed that BOD is the most important variable that determine COD, followed by phosphate concentration, DO, susupended solids and temperature. The results are consistent with the effects of land use on the river water quality as Dondang River flows through mainly housing area and factories or industrial area.**

Index Terms—**Chemical oxygen demand, artificial neural networks, mean sensitivity test**

I. Introduction

Human activities give impacts on river water quality, whether by direct disposal or indirectly. The action of direct disposal polluted the river by illegal garbage dumping or chemical disposal from nearby factory. Meanwhile, we are unable to observe the people who polluted the river indirectly, or even to measure the value of pollutants dumped into the river. Direct discharge and disposals also contribute due to the process leaching after the heavy rain, usually when the farmers put excess fertilizers or pesticides on the crops, and also the wastes from pasture sites. In natural environment, oxygen in the atmosphere will absorb into the surface area of the water but human activities such as open fire and industrial combustion, also contributes to the nutrient enrichment in water in the form of acid rain production [1].

The Artificial Neural Networks (ANNs) technique was used in this study to predict the output value of the desired parameters. ANNs are developed through the use of computer software to recognize the patterns of the data by training data through supervised learning. The idea of developing ANNs was inspired from biological nervous system in human brain, with the ability to organize its neurons and learn through 'experience'. ANNs flexibility learning system and adaptive ability allows them to learn from linear and non-linear function. The benefit is that, ANNs are more superior than the conventional or traditional statistical technique as ANNs are able to process complex, non-linear and parallel information processing system but produce higher accuracy and greater complexity of prediction [2]. ANNs have also been used in studies of water research such as identification of land use using water quality data [3], eutrophication prediction and modelling of BOD in Melen River [4] .

ANNs have been vastly used for the past few years in many area of research, including bioinformatics, image analysis, speech recognition and financial forecasting. The structures of ANNs are compose of nodes interconnected with each other which represent relationship between each nodes. The number of nodes is determined and variables can be adjusted chosen to build neural model. The outcomes of the model are based on the nodes, hidden layer and variables chosen for the model.

Biological Oxygen Demand (BOD) is the amount of oxygen used by aerobic microorganism to break down the organic matters into more stable form [5]. Chemical Oxygen Demand (COD) is the amount of oxygen used to oxidize chemical substances through chemical processes.

II. Methodology and Site of Studies

Water quality data was obtained from Department of Agriculture (DoE) Malaysia with the cooperation of Alam Sekitar Malaysia (ASMA). As ANNs requires large data sets for time prediction, the data consist 10 years of sampling data, from January 1998 to December 2007. Raw data was checked thoroughly for the value and sample size, the distribution of the data and cleaned before they can be use for modeling purposes. Missing data or duplicate data are deleted to ensure that training will run smoothly and to reduce Minimum Square Error (MSE).

The ANNs structure used in this study is Multi-Layer

Manuscript received March 3, 2012; revised April 7, 2012.

A. Talib is from School of Distance Education, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia (e-mail: anita@usm.my)

M. I. Amat is from Centre For Marine And Coastal Studies, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

Perceptron (MLP), a network that feed the input data to the neural layer to produce desire output. MLP works through back-propagation method, by re-feed the error data from output and propagate back to the input. Hidden layer and number of nodes or Processing Elements (PEs) are chosen for model building, based on the complexity of the data and studies. The data was divided into 3 parts, 55% of the data were used as training, 17% for cross validation and 28% for the testing. Cross validation data is needed to observe the performance of the training. This is one of the method to make sure model estimation produce high quality prediction, by preventing the generalization of the data. However, if the model is trained for too long or excessive epoch set for the training, the model can be overtrained. Overtraining means the model has generalized on the data itself, but is not able to learn the patterns of the unseen data. Thus, overtrained model are unable to adapt to new situation if new and different types of data were introduced to the model.

To choose the best data from the model prediction, model validation was made by testing its sensitivity. Mean sensitivity was tested on the data to validate the model whether it is good for prediction. Parameter that has bigger influence with the output value will have higher correlation value with the output.

COD has been set as the output of the training. Nine parameters chosen for the input of this model are Dissolve Oxygen (DO), Biological Oxygen Demand (BOD), Suspended Solid (SS), pH, ammonia (NH3), temperature, nitrate (NO3), total solid (TS) and phosphate (PO4). The number of processing elements (PEs), transfer option and learning rule are adjust until the best model prediction is achieve. Variables for transfer option tested for this studies include Tanh Axon, Sigmoid Axon, Linear Tanh Axon and Linear Sigmoid Axon. Meanwhile, the learning rule option can be choose from Momentum, Conjugate Gradient and Levenberg Marquadtt.

Dondang River runs from the west of Pulau Pinang, connected with other rivers and run through the island to the northeast of Pulau Pinang. Water quality data were taken from different sampling sites, from 2PG04, 2PG05 and 2PG06 station. The land use along the river is mainly housing area and factories or industrial area (see Fig 1).

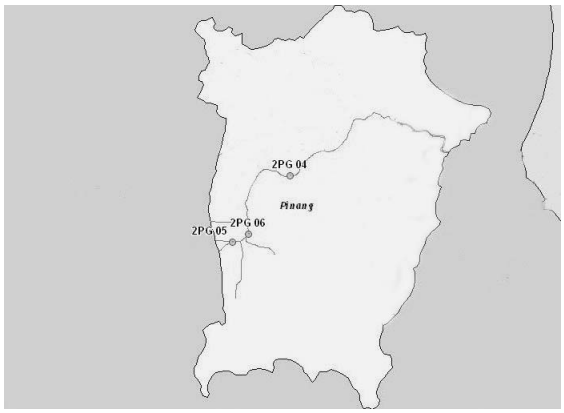

Fig. 1. The location of dondang river, pulau pinang.

III. Results and Discussion

By using Levenberg Marquadtt method, the training stops when MSE reached 0.002 and cross validation MSE reached 0.004. The average r value for training, cross validation and testing performance is 0.94. Meanwhile, r^2 value for the prediction model is 0.83. The prediction of COD data is as shown in Fig.2 below.

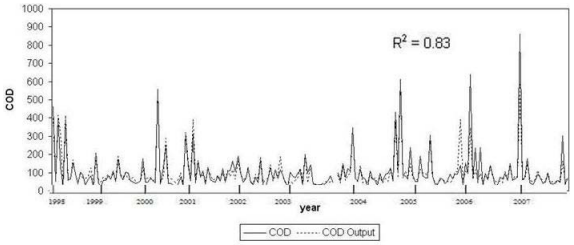

Fig. 2. Chemical oxygen demand estimation for dondang river.

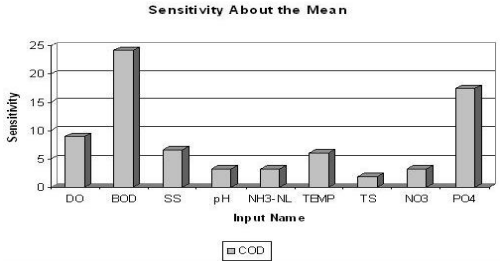

Fig. 3. Mean sensitivity test of dondang river.

When the mean sensitivity was tested (see Figure 3), BOD turns out to be the most significant value that influence the value of COD. This result was expected when the neural model was developed. COD and BOD process is related to each other as chemical reaction and oxidization process occur when microorganisms break down organic matter into more stable form. The second factor to contribute to the COD value detected is phosphate concentration. The COD value tend to be higher when the phosphate concentration is high. As phosphate takes important part in oxidization and energy-release process, there is a direct link between increase in phosphate concentrations and microorganisms. Thus, higher growth of microorganisms, consume more oxygen leading to low DO together with and an increase in BOD.

DO shows as the third factor in the mean sensitivity test, follow by suspended solid and temperature. This is more likely because the concentration of COD is not directly affected by DO concentration. This is possibly due to DO concentration is usually affected by other criteria too, such as pH and temperature of the water. According to [6] the rates of chemical and biological reaction are affected by temperature value, which is also one of the important factors that affect other parameters such as DO, BOD and COD. Apart from pH and temperature, other parameters include, ammonia, and nitrate. Total solid has a low effect on the value of COD but still show some correlation with the COD value. The land use along the river are mainly housing area and factories or industrial area, and these are the major contributors to the nutrient inputs affecting the water quality of Dondang river, as evident from the results of this study.

IV. Conclusions

Neural networks can be applied to study the factors that affect water quality of Dondang River. This study has shown

that COD can be used as a predictor variable to assess river water quality. Sensitivity analysis showed that BOD is the most important variable that determine COD, followed by phosphate concentration. The other important factors are DO, suspended solids, nitrate, ammonia and temperature. The results are consistent with the effects of land use on the river water quality as Dondang River flows through mainly housing area and factories or industrial area.

ACKNOWLEDGMENT

The authors would like to thank Universiti Sains Malaysia for the support of this project from the USM short-term Grant.

REFERENCES

[1] L. Håkanson, Water pollution : Methods and criteria to rank, model and remediate chemical threats to aquatic ecosystem. Backhuys Publishers, Leiden, 1999.

[2] S. Haykin, "Neural Networks: A comprehensive foundation," *Pearson Prentice Hall Inc*, 2005.

[3] H. Ha, M. K. Sternstrom, Identification of Land Use with Water Quality Data in Stormwater Using Neural Network. Journal of Water Research, pp. 4222-4230,2003.

[4] E. Dogan, B. Sengorur, and R. Koklu, "Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique," *Journal of Environmental Management*, pp. 1229-1235, 2009.

[5] B. A Cox, "A review of dissolve oxygen modeling techniques for lowland rivers," The Science of the Total Environment 314-316 : pp. 303-334, 2003.

[6] A. Drolc and J. Z. Koncan, "Water quality modeling of the River Sava," *Slovenia. Water Research* , pp. 2587 – 2592, 2006.