Artificial Neural Networks Combined with Sensitivity Analysis as a Prediction Model for Water Quality Index in Juru River, Malaysia

Mohd Fahmi Mohd Nasir [#1], Hafizan Juahir [#2], Nuraainaa Roslan [#3], Isahak Mohd [#4], Nur Aliaa Shafie [#5], Norlafifah Ramli [*6]

[#]*Department of Environmental Sciences, Faculty of Environmental Studies, 43400 UPM Serdang, Selangor, Malaysia*

[1]mfahmimnasir@gmail.com

[2]hafizan@env.upm.edu.my

[*]*Surface Water Monitoring Unit, Water and Marine Division, Department of Environment Malaysia, Federal Government Administrative Centre, 62574 Putrajaya, Malaysia*

***Abstract*- The manuscript describes the application of artificial neural networks (ANNs) for the series modeling of surface water quality prediction in Juru River, Malaysia. This is based on water quality data from twelve monitoring stations of Juru River (6 January 2003 until 5 November 2007) provided by Department of Environment (DOE), Malaysia. Thirty physicochemical parameters were involved in this analysis as input variables and water quality index as output variable. Three models were proposed to identify the most effective model in attempt to predict the WQI. Sensitivity analysis (SA) was carried out by using leave one out approach in order to indentify the most significant input-output relationship. The ANNs developed was successfully trained and tested using the available data sets and the performance of ANNs models was determined by coefficient of determination (R^2), coefficient of correlation (R) and root mean square error (RMSE). Results show that ANN-1 gives the higher value of R^2 (0.9942) and RMSE (2.8966), however this model was only trained and tested using all available parameters. The second model (ANN-2) gives R^2 value (0.9839) slightly higher compare to the third model, ANN-3 (0.9811). This is supported by the RMSE values which indicate that ANN-2 has a lower value compared to ANN-3 which is 2.1877 and 2.411. Hence, this study will trigger DOE to use ANNs in order to predict WQI other than using conventional method (WQI equation) that is currently being used by DOE. In addition, the ANNs managed to show remarkable prediction performance to predict the WQI in Juru River.**

***Keywords*- Artificial Neural Network; Sensitivity Analysis; Water Quality Index**

I. INTRODUCTION

Over the past few decades, increasing number of industrial areas blooms throughout the country resulting to enormous amount of anthropogenic activities being introduce to the environment which eventually effect the water bodies. Nowadays with the advanced science and technology, the increase in human population, industries, agriculture activities and urban development have causes the riverbanks to widen. Nonetheless, unstoppable human activities such as agricultural using pesticides, domestic sewage, factories effluents and even soil erosion due to improper development have lead to pollution in the water bodies [1].

However, water pollution mainly occurs due to the overloading of waste in the water system. The contamination of streams, lakes, underground water, bays or oceans by substances can be harmful to living things, not only human, but to the wildlife and plants. Pollutants that have been discharges into water body via point source and non-point source are difficult to identify due to indefinite origin of the pollutants. However, factors such as effluent from industrial area, sewage disposals and land clearings greatly influence the water quality. In fact, the water pollution not only affect human health, but also the entire environment especially aquatic lives that live in it [2].

Currently, ANNs has been widely used in solving environmental problem including water resources modeling and management problems [3]. Therefore, this study implemented ANNs to determine the most significant water quality variables that contribute to the water pollution. ANNs are universal estimators of non-linear mapping that are able to learn and to generalize relationship between input and output data from examples (training data) [4]. The dominant misunderstandings arising from the difference in terminology is that many scientists who employ ANNs to water resources issues declare that ANNs can 'learn from examples' and that is one of the dominant advantages ANNs have over other methods [5].

The strength in ANNs is pattern recognition and pattern classification but this program also can be use in predictive purposes [6]. ANN have been successfully used in groundwater parameter estimation [4], groundwater forecasting level [7] and also for groundwater numerical modeling and environmental design [3]. Artificial neurons were first developed by Warren McCulloch in 1943, a neurophysiologist, and Walter Pitts. At the beginning of 1980s, expert systems were represented the future of artificial intelligence and of computers in general [8].

Additionally, this study uses ANNs to analyze the data combined with Sensitivity Analysis (SA) to identify the most effective approach to solve the water quality problem in Juru River, Penang. Juru River has been listed by the Department of Environment (DOE) as one of the seven polluted river in Malaysia [9]. DOE utilizes Water Quality Index (WQI) to evaluate the water quality status of the river. The major pollution indicators in polluted river are Biological Oxygen Demand (BOD), Ammoniacal Nitrogen (AN) and Suspended Solid (SS). In 2006, twenty two river basins were categorized as being polluted by BOD, forty one river basins by AN and forty two river basins by SS. The high BOD levels are generally contributed by untreated or partially treated sewage and discharges from agro-based and manufacturing industries directly into the river. The main sources of AN originated from domestic sewage and livestock farming, whilst the sources for SS were mostly earthworks and land clearing activities [9].

According to DOE (2005), the water is so toxic that it is unsafe for drinking even after being boiled. The Water Quality Index (WQI) remains at Class V indicating that the river is excessively polluted and no fish would survive in such rivers. As more lands in the Juru River Basin are turning into urbanized areas, the river is at risk of becoming an open wastewater sewers [10]. The information gathered and on-site investigations revealed that Juru River pollution is caused by industrial wastes from the nearby industrial areas, wet market waste, household waste and sewage wastewater from the residential or settlement areas [11, 12]. Human settlements including squatters along the riverbanks of Juru River are not equipped with proper sanitary system. Domestic wastes are being dumped directly to the river. The drainage system built in the industrial sites is directly channeled to the river.

Therefore these model aims to establish the inputs parameter for prediction of water quality and to provide the best set of input-output parameters for the development of learning and predicting procedures using ANN.

II. METHODOLOGY

A. Study Area

The Juru River is located in Seberang Perai Tengah in the state of Pulau Pinang. Additionally, the main course of Juru River is about 15.62 km with a total catchment area of 60.5 km^2. The instance, the upper parts are known as Permatang Rawa River; the middle area is Rambai River while the downstream channel is called Juru River. The river flows from Bukit Minyak area towards the west and discharges into the Straits of Melaka. There are three major tributaries namely as Ara River, Kilang Ubi River and Pasir River, including a drain called Parit 4. A large portion of the catchments areas are made up mainly from industrial and residential area.

The main types of land use in this basin are forest, agriculture, industrial and residential. Mostly, forested areas are located at the upstream east of the basin. In term of land use utilisation, industrial is considered to be one of the main economic activities in this river basin where it covers about 592.3 hectares of the area. According to land use map of Juru River Basin as shown in Fig. 2, main industrial area is located in the downstream and middle stream of the river. The orchard is the main agricultural crops (255.8 hectares), followed by paddy field (200.0 hectares). Vegetable also one of the agricultural crops which cover a total area of 6.6 hectares.

According to the DOE Report (2006), Juru River is listed as one of the polluted rivers in Malaysia [9]. This might be due to most industrial activities are located in the downstream and middle stream of the basin thus contributed to impurity of Juru River water quality. Small and medium industries are main polluters compared to multinational companies. This could be due to the fact multinational companies are more readily to abide by environmental standards than small and medium industries. The main source of pollution in Juru River Basin are recognize to come from point sources such as industrial effluents from sewage treatment plant and non-point sources such as agricultural runoff, residential, commercial and business, industry and others [11, 12]. The heavy industries will henceforth affect the cockle farming activities available in the estuary of Juru River thus consuming the cockles will eventually causes severe diseases due to enormous amount of pollutant being irresponsibly discharges into the river.

B. The Data Portion

Initially, the ANNs models were developed using available data of thirty water quality parameters and 308 samples data sets for Juru River. These data were collected starting from 6 January 2003 until 5 November 2007 from twelve water quality monitoring stations in Juru River Basin managed by DOE under the Ministry Natural Resources and Environment.

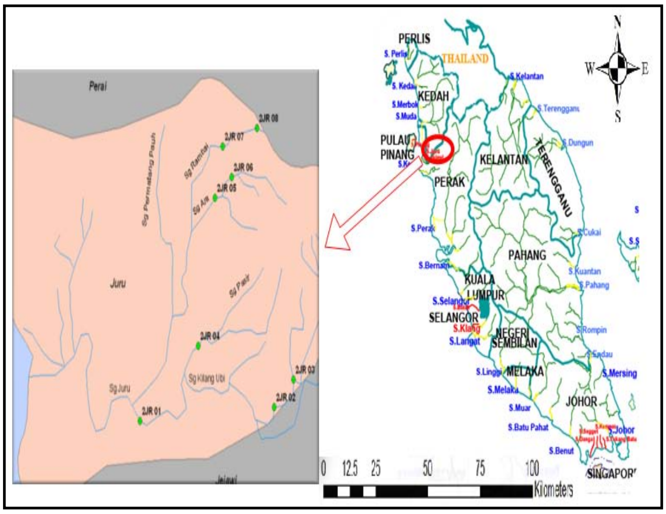


Fig 1 Water quality monitoring station supervised by DOE in Juru River Basin

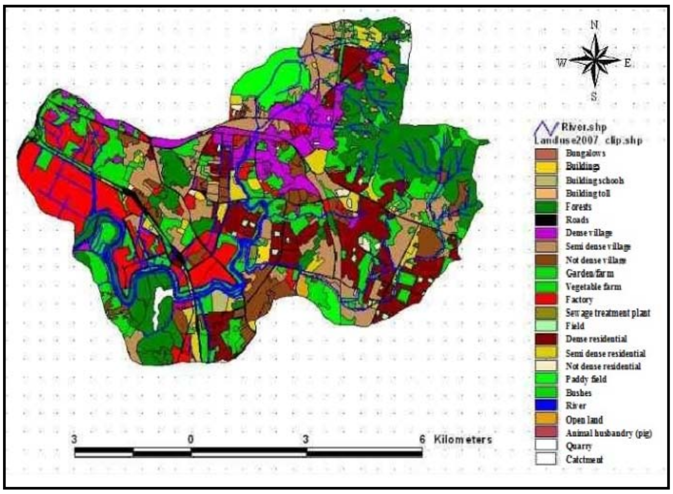


Fig 2 The varieties land use in the Juru River Basin

The thirty parameters analyzed were BOD, AN, SS, Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), , Dissolved Solid (DS), pH, Temperature (TEMP), Conductivity (COND), Salinity (SAL), Turbidity (TUR), Total solid (TS), Nitrate (NO_3), Chloride (Cl), Phosphate (PO_4), Arsenic (As), Mercury (Hg), Cadmium (Cd), Chromium (Cr), Lead (Pb), Zinc (Zn), Calcium (Ca), Iron (Fe), Potassium (K), Magnesium (Mg), Sodium (Na), Oil and Grease (OG), MBAS (Methylene Blue Active Substance), Escherichia coli *(E-coli)* and Coliform. These parameters were used as input to develop three ANNs models whereas WQI were used as output in these models.

According DOE (2004), the WQI serves as the basis for environmental assessment of a watercourse in relation to pollution load categorization and designation of classes for beneficial uses as provided under the National Water Quality Standards for Malaysia (NWQS) [13]. The WQI was derived using DO, BOD, COD, AN, SS and pH as the index.

C. Neural Network Prediction

In this study, three different ANNs models were developed. These models aim to determine the significant parameter that was affected by WQI. Initially, input for the first model used thirty parameters with WQI as desired output. While the second model only use six significant parameters. For this matter, the first step was conducted to reduce the insignificant parameter by using ANNs which includes the leave one out method based on the correlation between each parameter with WQI in order to recognize which parameter contribute most into the WQI of Juru River [14]. In leave one out method, variables were removed step-by-step to indicate the difference of R^2 (all parameters) and R^2 (leave one out parameter) value. However, the third model only uses six parameters based on the WQI. These three models were compared with the R^2 and RMSE value to define the best model.

Independent test set is used at different phase of learning and can be an indicator of the model performances. The available data is required to be divided into three sub phases; training phase, testing phase, and a validation phase. Since cross validation requires that validation data must be separated from the testing set before further analysis. This is a rule of thumb that can be applied where network geometry or input parameters are optimized by trial and error [5]. These neural networks consist of three layers, which comprise an input layer, a hidden layer, and an output layer (Fig. 3).The size of the input and hidden of network has been varies depending on prediction horizon, whereas the output layer has single node [7]. Each neuron in one layer is connected to the neurons in the next layer, whereas there are no connections between the units of the same layer even though the number of neurons in each layer may vary depending on the problem [4, 14].

Capability of conservative feedforward networks is limited in that the nodes in a single layer can only be connected to the nodes in the next layer. However, this flaw have been upgraded and now recurrent networks allows nodes in a layer are able to interconnect with nodes in the previous layer, next layer, the same layer or even with the nodes itself [15]. Other than that, feed forward networks require that dynamic systems to be treated thoroughly [16, 17]. Treatment for this system is achieved by including lagged inputs whereby recurrent networks are able to model dynamic properties without any added treatment [16, 17].

Determination of the number of connection weights and arrangement of network is generally done by fixing the number of hidden layers and choosing the optimum number of nodes in the layer. However (Kumar, 1993) stated that fixation of number of nodes is the best method compared to compromising the number of hidden layers. This method is implemented during training phase of the raw data where connections of nodes and connection weights are optimized [18].

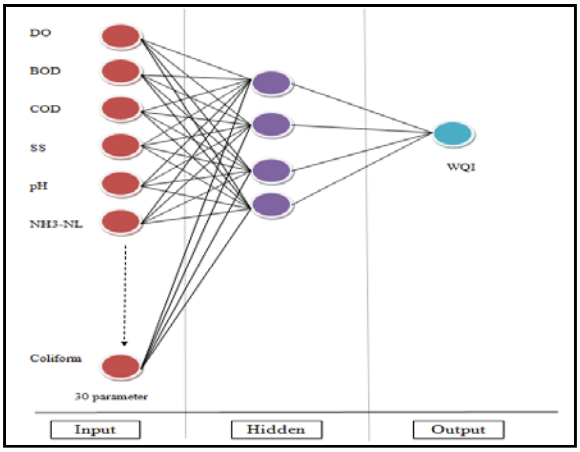


Fig 3 Example of network structure

Properties of small and large networks are different in a way that smaller networks require fewer physical resources, higher processing rate and can be implemented easily and economically. Smaller networks also implicate smaller surface error but more complicated and contains more local minima [19-21]. The use of more than one hidden layer provides greater flexibility and enables approximation of complex functions with fewer connection weights in many situations [22-25].

D. Sensitivity Analysis

Sensitivity analyses employed leave one out approach in order to indentify the most significant input-output relationship that have been carried out manually. It is used to provide information on the relative significance of the thirty input variables in each parameter along Juru River. However, the sensitivity were defined as the RMSE value indicates the performance of the network if the variable under consideration is removed from the analysis [26]. Thus, disappearance of more important variables results in higher RMSE values indicating that the network is affected to greater extent when these variables are not included [26]. Based on thirty parameters, there are only six significant parameters chosen will be determined to evaluate WQI value and then compared with six parameters proposed by DOE which is using conventional method (WQI equation).

E. Criteria of Model Performance Evaluation

Three different criteria are used in order to evaluate the effectiveness of each network and its ability to make precise prediction. The R^2 efficiency criterion, given by

$$R^2 = 1 - \frac{\sum(x_i - y_i)^2}{\sum y_i^2 - \frac{\sum y_i^2}{n}} \quad (1)$$

Also, the Root Mean Square Error (RMSE) calculated by

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \quad (2)$$

that indicates the discrepancy between the observed and calculated values. The lowest the RMSE, the more accurate the prediction is. Percent residual error (%RE) is given as

$$\%RE = \frac{x_i - y_i}{x_i} \times 100 \quad (3)$$

where x_i is the observed data, y_i the predicted data and n is the number of observation and representing the percentage of the initial uncertainty explained by the model. The best fit between observed and predicted values, which is unlikely to occur, would have RMSE = 0 and R^2 = 1 [7].

III. RESULTS AND DISCUSSIONS

The results depicted in Table 4 show the models performance evaluations of the significant parameters for forecasting WQI values. There are two models that are implemented in this study recognized as Model A and Model B. The model A consist of two models, the first model involves thirty parameter as inputs with three hidden layers and one output (WQI) while the second model explains six significant parameters after conducting the sensitivity analysis. Meanwhile, Model B was developed by considering the six parameters proposed by DOE. The model ANN-1 with all the input parameters were selected as the most appropriate model for WQI forecasting with high R^2 is 0.9839 and low RMSE is 2.1877 as compared to other model. The depicted model ANN-2 and ANN-3 indicates that out of all the six significant parameters there is one parameter differ from both model which is for ANN-2 as it consider mercury whereas ANN-3 consider pH in the model. Moreover, R^2 value for ANN-2 is greater than ANN-3 which is 0.9839 and 0.9811. Besides that, RMSE value for ANN-2 (2.1877) is lowers than ANN-3 (2.4111). The best network model was chosen based on the highest coefficient of determination (R^2) and lowest RMSE for each parameter [7].Therefore, the ANN model are undoubtedly capable to be used as an alternative procedure in order to predict WQI rather than using conventional method (WQI equation) which is currently being used by DOE. Moreover, the ANN also consider as supervise pattern recognition techniques therefore, the input-output relationship determined are given by specific pattern (condition) due to particular study location without loss of much information [14].

Table 4 The performance of ANN models

Input data	Network	No. of Observation	R^2	R	RMSE
Model A	[30,3,1] (ANN-1)	308	0.9942	0.9971	2.8965
	[6,3,1] (ANN-2)	308	0.9839	0.9919	2.1877
Model B	[6,3,1] (ANN-3)	308	0.9811	0.9905	2.4111

Table 5 shows the results of SA which indicates the determination of coefficient for each parameter affecting the WQI. The leave one out approach excluded one variable at a time in order to determine the percentage of contribution poses by the variable that would affect the R^2 values [26]. Before applying leave one out approach, thirty parameters were used to predict WQI. Henceforth, six parameters indicate the highest percentage of contribution from the overall water quality parameter. The percentage of contribution of each parameter shows the highest concentration when eliminating DO from all the parameters with 43.35% that explains 96% of all the data followed by BOD with 23.43% contribution (R^2=0.9785). On contrary, the lowest percentage of contribution is SS with 3.40% (R^2=0.9944**)**.

A. ANNs Models

Fig. B (a) to D (a) shows the graph plotting for observed WQI and predicted WQI. The figure signifies that some of overall observations were out from the range of upper and lower boundary (95% mean of the confidence interval). Therefore, this proved that these models are able to predict WQI values from the all available inputs with negligible precision.

Fig. B (c) to D (c) demonstrated the performance of the ANN models of Juru River representing the training and testing based on actual WQI and predicted WQI. The model developed produced good accuracy for both training (66.76%) and testing sets (33.33%) from overall data set. Which further explain the network output almost equal to output [4] and high accuracy for the cross validation with minimum value of RMSE [27]. As shown in Fig. B (c) the predicted WQI values from the training set able to follow the pattern recognize by the training set and produce high reliability and goodness-of-fit (R^2=0.9942).

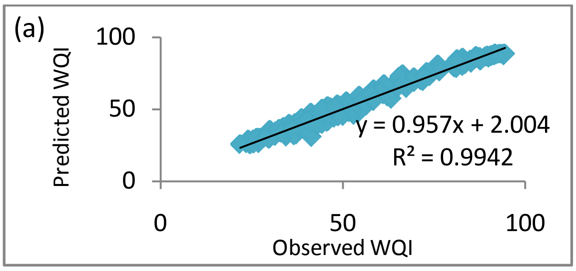
Table 5 indicate the contribution for each parameter

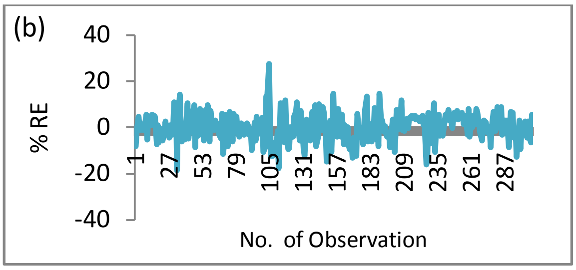
Model A	R^2	Difference R^2	Contribution (%)
All parameters (30)	0.9971		
L-As	0.9961	0.001	1.26
L-BOD	***0.9785***	***0.0186***	***23.43***
L-Ca	0.9965	0.0006	0.76
L-Cd	0.9968	0.0003	0.38
L-Cl	0.9975	-0.0004	-0.5
L-COD	***0.9922***	***0.0049***	***6.17***
L-COND	0.9972	-0.0001	-0.13
L-Cr	0.9973	-0.0002	-0.25
L-DO	***0.9629***	***0.0342***	***43.35***
L-DS	0.9975	-0.0004	-0.5
L-*E-coli*	0.997	0.0001	0.13
L-Fe	0.9971	0	0
L-Hg	***0.9932***	***0.0039***	***4.91***
L-K	0.9973	-0.0002	-0.25
L-MBAS	0.9959	0.0012	1.51
L-Mg	0.9969	0.0002	0.25
L-Na	0.9956	0.0015	1.89
L-AN	***0.9894***	***0.0077***	***9.7***
L-NO_3	0.9973	-0.0002	-0.25
L-OG	0.9972	-0.0001	-0.13
L-Pb	0.9963	0.0008	1.01
L-pH	0.9963	0.0008	1.01
L-PO_4	0.9958	0.0013	1.64
L-SAL	0.9976	-0.0005	-0.63
L-SS	***0.9944***	***0.0027***	***3.4***
L-TEMP	0.9972	-0.0001	-0.13
L-TS	0.9976	-0.0005	-0.05
L-TUR	0.995	0.0021	2.64
L-Zn	0.9966	0.0005	0.63
L-Coliform	0.9974	-0.0003	-0.38
Total		0.0794	100

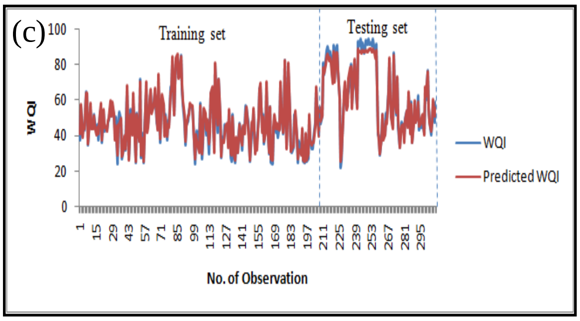
Note: L; leave one out

Residual error graph indicate the contrast of the observed and predicted WQI value (Fig. B (b) to D (b)). The residual values for each observation were in the ranged -19 to 28 for Model A (ANN-1), -15 to 10 for Model A (ANN-2) and -21 to 21 for Model B (ANN-3). In spite of the range is quite broad, the residual data were evenly distributed in the zero values. The verification and applicability of the model was influence by the existence of the outlier observations as shown also in Fig. B (a) to D (a).

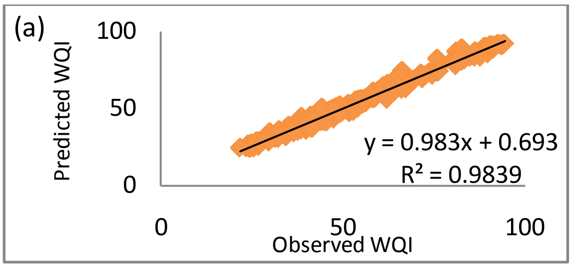
B. Model A (ANN-1)

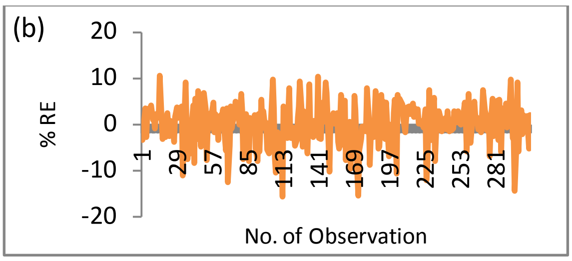


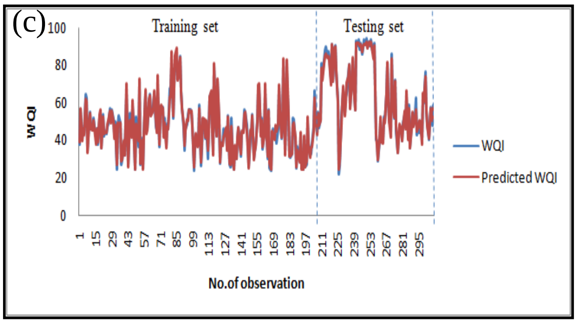




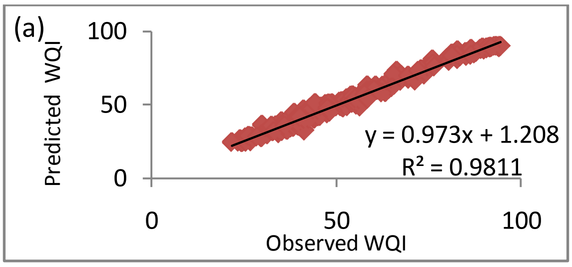
C. Model A (ANN-2)

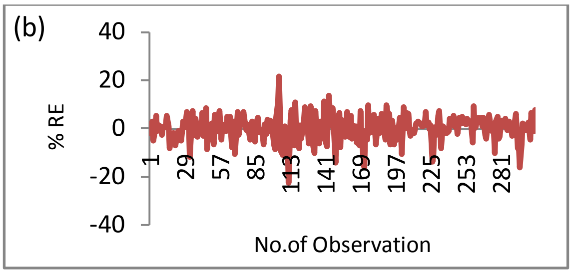


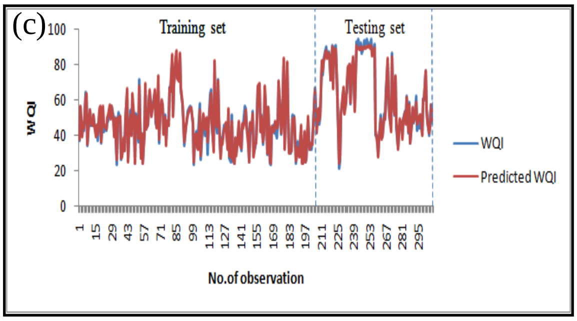




D. Model B (ANN-3)







IV. CONCLUSIONS

The models show the ability of ANNs to predict WQI by using only six significant parameters after conducting the sensitivity analysis and six parameters that was proposed by DOE which are compared to all the available water quality parameters (thirty inputs). The analyzed data show better prediction performance in using leave-one-out method and the most effective is Model A (ANN-2) since it gives the higher R^2 value (0.9839) compared to Model B (ANN-3) where R^2 value (0.9811). This study proved that ANNs are undoubtedly capable to be an alternative method in order to predict WQI rather than using conventional method (WQI equation) which is currently being used by DOE. The applications of ANNs for this study confirm that most of the parameter is significant to predict WQI even though it allows a reduction of the number of water quality parameter to identify the correct regional samples. The results showed that all the ANNs models are successful in predicting WQI according to the most significant parameter from Model A in Juru River Basin better than thirty parameters and six parameters proposed by DOE-WQI. Despite various limitations, ANNs is absolutely known as one of the best tool as it benefited in terms of their flexibility in terms of data requirements.

In general, the predictions from all the three models show that ANNs is an excellent predicting tool for WQI and a very useful for helping decision makers as part of the Juru River management measures.

ACKNOWLEDGEMENT

The authors are gratefully acknowledged to the Department of Environment (DOE), Malaysia for providing the field data. Not forgotten, the author would like to express greatest gratitude to the lab members for their wise ideas, encouragement and guidance throughout the making of this paper.

REFERENCES

[1] Niemi, G. J., Devore, P., Detenbeck, N., Taylor, D., Lima, A., Overview of Case Studies on Recovery of Aquatic Systems from Disturbance. *Environmental Management*, *14*, 571-587, 1990.

[2] Najah, A., Elshafie, A., Karim, O. A., Jaffar, O., Prediction of Johor River Water Quality Parameters Usiang Artificial Neural Networks. *European Journal of Scientific Research 28*, (3), 422-435, 2009.

[3] Nikolos, I. K., Stergiadi, M., Papadopoulou, M. P., Karatzas, G. P., Artificial neural networks as an alternative approach to groundwater numerical modelling and environmental design. *Hydrological Process, 22,* 3337-3348, 2008.

[4] Garcia, L. A., Shigidi, A., Using neural networks for parameter estimation in ground water. *Journal of Hydrology*, *318*, 215–231, 2006.

[5] Maier, H. R., Dandy, G. C., Neural network for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, *15*, 101-124, 2000.

[6] Dayoff, J. E., *Neural network architectures*. Van Nostrand Reinhold, New York, 1990.

[7] Daliakopoulos, I. N., Coulibaly, P., Tsanis, I. K., Groundwater level forecasting using artificial neural networks. *Journal of Hidrology*, *309*, 229-240, 2005.

[8] Juahir, H., Zain, S.M., Yusoff, M.K., Hanidza, T.I.T., Armi.A.S.M., Toriman, M.E., Mokhtar, M., Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environ Monit Assess*, *DOI 10.1007/s10661-010-1411-x*, 2008.

[9] DOE *Malaysia Environmental Quality Report 2005*; Department of Environment, Ministry of Natural Resources and Environment Malaysia: 2006.

[10] Toriman, M. E., Hashim, N., Hassan, A. J., Mokhtar, M., Juahir, H., Gasim, M. B., Abdullah, M. P., Study on the Impact of Tidal Effects on Water Quality Modelling of Juru River, Malaysia. *Asian Journal of Scientific Research*, *4*, (2), 129-138, 2011.

[11] Al-Shami, S. A., Rawi, C. S. M., Ahmad, A. H., Nor, S. A. M., Distribution of Chironomidae (Insecta: Diptera) in polluted river of the Juru River Basin, Penang, Malaysia. *Journal of Environmental Sciences*, *22*, (11), 1718-1727, 2010.

[12] Alkarkhi, A. F. M., Easa, A. M., Comparing Discriminant Analysis and Logistic Regression Model as a Statistical Assessment Tools of Arsenic and Heavy Metal Content in Cockles. *Journal of Sustainable Development*, *1*, (2), 102-106, 2008.

[13] DOE *Malaysia Environmental Quality Report 2003*; Department of Environment, Ministry of Natural Resources and Environment Malaysia: 2004.

[14] Juahir, H., Zain, S.M., Toriman, Mokhtar, M., Man, H. C., Application of Artificial Neural Network Models for Predicting Water Quality Index. *Jurnal Kejuruteraan Awam*, *16*, (2), 42-55, 2004.

[15] Warner, B., Misra, M., Understanding neural networks as statistical tools. *American Statistician*, *50*, (4), 284-293, 1996.

[16] Gencay, R., Liu, T., Nonlinear modelling and prediction with feedforward and recurrent networks. *Physica D*, *108*, (1-2), 119-134, 1997.

[17] Krishnapura, V. G., Jutan, A., ARMA neuron networks for modeling nonlinear dynamical systems. *Canadian Journal of Chemical Engineering*, *75*, (3), 574-582, 1997.

[18] Kumar, K. K., Optimization of the neural net connectivity pattern using a backpropagation algorithm. *Neurocomputing*, *5*, 273–286, 1993.

[19] Castellano, G., Fanelli, A. M., Pelillo, M., An iterative pruning algorithm for feedforward neural networks. *IEEE Transactionson Neural Networks*, *8*, (3), 519–531, 1997.

[20] Towell, G. G., Craven, M. K., Shavlik, J. W. In *Constructive induction in knowledge-based neural networks*, Proceedings of the 8th International Workshop on Machine Learning, San Mateo, 1991; Kaufman, M., Ed. San Mateo, pp 213–217.

[21] Bebis, G., Georgiopoulos, M., Feed-forward neural networks: Why network size is so important. *IEEE Potentials* 1994, 27–31.

[22] Flood, I., Kartam, N., Neural networks in civil engineering. I: Principles and understanding. *Journal of Computing in Civil Engineering*, *8*, (2), 131–148, 1994.

[23] Ripley, B. D., Neural networks and related methods of classfication. *Journal of the Royal Statistical Society B*, *56*, (3), 409–456, 1994.

[24] Tamura, S. i., Tateishi, M., Capabilities of a four-layered feedforward neural network: Four layers versus three. *IEEE Transactions on Neural Networks*, *8*, (2), 251–255, 1997.

[25] Sarle, W. S. In *Neural networks and statistical models*, In: Proceedings of the Nineteenth Annual SAS Users Group International Conference, SAS Institute, 1994; SAS Institute, 1994; pp 1538–1550.

[26] Lee, J. H. W., Huang, Y., Dickman, M., Jayawardena, A. W., Neural network modelling of coastal algal blooms. *Ecological Modelling*, *159*, 179-201, 2003.

[27] Rossel, R. A. V., Behrens, T, Using data mining to model and interprets oil diffuse reflectance spectra. *Geoderma*, *158*, 46–54, 2010.