Forecasting the Water Quality Class in a River Basin using an Artificial Neural Network with the Softmax Activation Function

[1,2]Shah Christirani Azhar, [2]Ahmad Zaharin Aris, [2]Mohd Kamil Yusoff and
[2]Mohammad Firuz Ramli
[1]Centre of Foundation Studies for Agricultural Science,
[2]Department of Environmental Sciences, Faculty of Environmental Studies,
Universiti Putra Malaysia, Serdang, 43400 Selangor, Malaysia
christirani@upm.edu.my, +60389468428; +60389466997

Abstract: Classification of river water quality needs an efficient method to reduce energy, save time and decrease the risk of errors. This study describes the application of an Artificial Neural Network (ANN) with the softmax activation function to forecast the Water Quality Class (WQC) under the National Water Quality Standard (NWQS) of the Muda River Basin (MRB) (Malaysia). The water quality was classified automatically without Water Quality Index (WQI) calculation. Two different sets of Water Quality Variables (WQVs) were applied as input variables. The modelling discover that the optimal network architecture was the 1:6-1:6-1:1 and used a 60-20-20% splitting plan. ANN1 with the six WQVs was selected to predict the WQC in the MRB. Predictions of the WQC rendered by this model for the training set were very accurate (96.8% correct, Percent Incorrect Prediction (PIP) = 3.2, CEE = 3.44). The approach presented is a very useful and offers a compelling alternative to forecasting of river class, mainly because WQI calculation involves a complex and lengthy calculations. Subsequently, this approach could be applied to water quality classification in other river basins for better water quality management.

Key words: Artificial neural network, softmax activation function, water quality modelling, Muda River basin, quality management, quality classification

INTRODUCTION

Water quality has deteriorated due to important factors like the increasing population, industrialisation, agriculturalisation and urbanisation (Semiromi *et al.*, 2011; Tyagi *et al.*, 2013). Therefore, an effective assessment of water quality must be conducted to recognise the level of pollution. The wide range of variables used to measure the water quality of such as pH, temperature, conductivity, turbidity, nutrients, metals and pathogens. Consequently, the monitoring system produces a massive of data and complicated. However, a tough task to transform complex data into meaningful information to better defines the status of the water quality. For this reason, the water quality parameters are classified to represents the contamination degree of surface water and reflects the status of water quality in a holistic manner (Zampella *et al.*, 2006; Santos Simoes *et al.*, 2008; Suratman *et al.*, 2009). The results of water quality classification should be presented meaningfully to those responsible for governing water quality and consumers. Besides, the classification of water quality can be used to

find out the effectiveness of water quality restoration measures when an action was taken on the identified pollution. In Malaysia, there are two primary methods adopted to classify river water quality, the National Water Quality Standards (NWQS) which obtained by considering the main uses of water and the Water Quality Index (WQI) which is rooted from the NWQS (Zainudin, 2010).

The WQI serves as the basis for the environmental assessment of a water course based on pollution load categorisation. The index is used to change a massive water quality data from multiple variables into a single number which represents the status of water quality. Further, the WQI classifies the water quality based on suitability for some target uses including domestic water supply, fisheries and aquatic propagation, livestock drinking, recreation and agriculture (DoE., 2014). The NWQS categorises rivers into five classes (I-V). Nevertheless, WQI calculation practiced in Malaysia involving long and complex calculation, similar to methods used by other countries (Cude, 2001; Liou *et al.*, 2004; Bordalo *et al.*, 2006; Abrahao *et al.*, 2007;

Alobaidy *et al.*, 2010; Thi Minh Hanh *et al.*, 2010). Therefore, the calculation of WQIs took longer, more effort and had to operate with careful to avoid inadvertent error during calculation. Therefore, this study suggests a direct means an efficient and fast method for the calculation and prediction of water quality accordance artificial neural networks. ANN modeling capable save computation time, decreases the risk of errors in calculation and effort (Gazzaz *et al.*, 2012).

ANNs are widely accepted as a technology that can deal with non-linear problems. ANNs have been successfully used in hydrology-related areas such as flow forecasting (Kentel, 2009; Adamowski and Sun, 2010; Huo *et al.*, 2012), groundwater modeling (Huang *et al.*, 2011; Yoon *et al.*, 2011); water quality (Dogan *et al.*, 2009; Singh *et al.*, 2009; Khalil *et al.*, 2011), precipitation forecasting (Kim and Pachepsky, 2010; Melesse *et al.*, 2011; Rajaee, 2011), water management policy (Iliadis and Maris, 2007; Glezakos *et al.*, 2009), hydrologic time series (Wu and Chau, 2010; Khashei and Bijari, 2011) and reservoir operations (Chaves and Kojiri, 2007; Chaves and Chang, 2008; Wang *et al.*, 2011). Nevertheless, the development of predictive models using ANN for classification based on NWQS has not been conducted. The main objectives of this study was to: determine the best model to predict river's Water Quality Class (WQC) and demonstrate the ability of the ANN to provide a useful model to predict WQC in the river basin.

MATERIALS AND METHODS

Study area: The Muda River Basin (MRB) is located in the Northwest corner of Peninsular Malaysia (Fig. 1) and covers an area of about 4.210 km^2. The main river in the basin is the Muda River with a length of 180 km. The average annual temperature is 27°C while the mean annual precipitation in the whole basin is 2.300 mm. The Muda River belongs to the state of Kedah and Penang (Julien *et al.*, 2009). Therefore, this river is the major essential source of water for domestic and agricultural uses in Kedah and Penang. Up to 96.0% of Kedah's water supply and 80.0% of Penang's water supply are drawn from the Muda River. The major land use in the basin is for agriculture and forestry (Fig. 1). Agriculture-related operations, like animal and crop farming, cover approximately 55.0% of the total land area of the MRB. Meanwhile, about 35.0% of the catchment area is still covered by forest. Therefore, agro-based industries such as rubber processing, palm oil factories and logging are the main economic activities in the area. In general, the water quality in the MRB is classified as class II (DoE., 2014).

The data set: The water quality dataset was obtained from the Department of Environment (Malaysia). The data set covers the period from 1998-2013. The data contains the values of selected monitored WQVs for nine monitoring stations of the main river and the tributaries. Locations of the river monitoring stations are presented in Fig. 1. The data sets taken for this study comprise of six WQVs: Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), pH and ammonia nitrogen (NH_3-N). Descriptive statistics for these variables are shown in Table 1 and 2.

The water quality index: Status of river's water quality in Malaysia is assessed using the formula DOE-WQI issued by the Department of Environment (Malaysia). The WQI is computed based on the concentration of DO, BOD, COD, SS, pH and NH_3-N (DoE., 2014). Calculations are performed on the parameter sub-indices. The DOE-WQI was calculated as follows:

$$WQI = 0.22*SI_{DO} + 0.19*SI_{BOD} + 0.16*SI_{COD} + 0.15*SI_{NH3-N} + 0.16*SI_{SS} + 0.12*SI_{pH}$$

Table 1: Water quality properties measured between 1998 and 2013 in MRB

WQVs/units	Minimum	Maximum	Mean	SD
DO (mg/L)	2.09	8.15	5.91	1.33
BOD (mg/L)	0.85	32.78	3.14	5.12
COD (mg/L)	5.88	110.75	23.78	15.93
SS (mg/L)	2.67	124.00	49.08	31.44
pH (unit)	5.64	7.64	6.72	0.39
NH_3-N (mg/L)	0.01	39.78	2.02	6.27

Table 2: The sub-index calculation formulae for the WQI (DoE., 2014)

Parameters	Values[a]	Sub-index equation
DO	$X \leq 8$	$SI_{DO} = 0$
(Saturation %)[b]	$8 < X < 92$	$SI_{DO} = -0.395 + 0.03*X^2 - 0.0002*X^3$
	$X \geq 92$	$SI_{DO} = 100$
BOD	$X \leq 5$	$SI_{BOD} = 100.4 - 4.23*X^3$
	$X > 5$	$SI_{BOD} = (108*e^{-0.55*X}) - 0.1*X$
COD	$X \leq 20$	$SI_{COD} = 99.1 - 1.33*X$
	$X > 20$	$SI_{COD} = (103*e^{-0.0157*X}) - 0.04*X$
NH_3-N	$X \leq 0.3$	$SI_{NH3-N} = 100.5 - 105*X$
	$0.3 < X < 4$	$SI_{NH3-N} = (94*e^{-0.573/x}) - 5|X-2|$
	$X \geq 4$	$SI^{NH3-N} = 0$
SS	$X \leq 100$	$SI_{SS} = (97.5*e^{-0.00676*X}) + 0.05*X$
	$100 < X < 1000$	$SI_{SS} = (71*e^{-0.0016*X}) - 0.015$
	$X \geq 1000$	$SI_{SS} = 0$
pH	$X < 505$	$Si_{pH} = 17.2 - 17.2*X + 5.02*X^2$
	$5.5 \leq X < 7$	$Si_{pH} = -242 + 95.5*X - 6.67*X^2$
	$7 \leq X < 8.75$	$Si_{pH} = -181 + 82.4*X - 6.05*X^2$
	$X \geq 8.75$	$Si_{pH} = 536 - 77*X + 2.76*X^2$

[a]X is the concentration of the indicated parameter in mg/L, except for pH and DO [b]DO (Saturation%) = [DO (mg/L)*12.795]-0.05
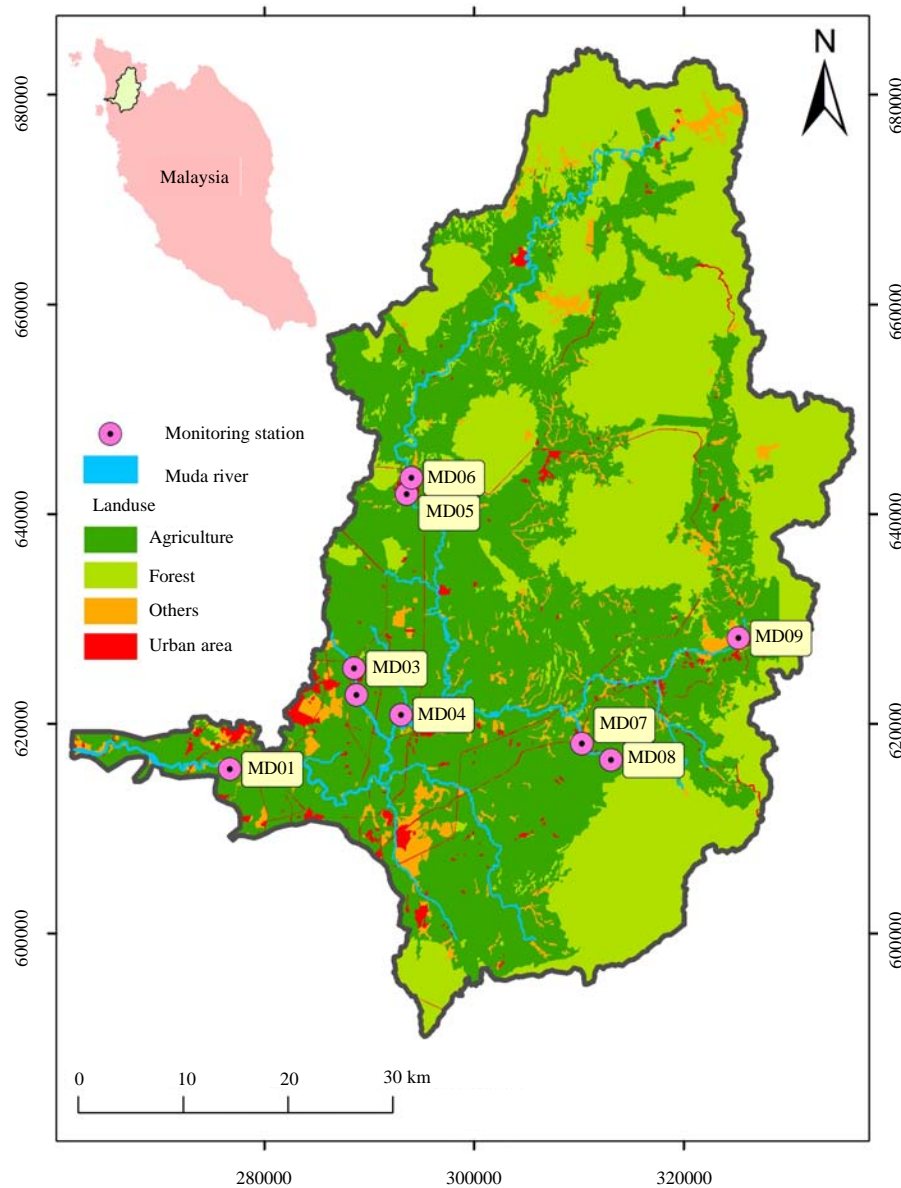
Fig. 1: Muda River basin and its land use. The figure shows the location of Muda River basin and the distribution of monitoring stations

where, SI stands for Sub-index which is calculated as in Table 2. River classes calculated using WQI, provide information about the beneficial uses of the river. These classes based on the standard limits for selected pollutants. For this purpose, the Department of Environment (Malaysia) have determined the parameters used as indicators and WQI that characterize each of WQC, defined usage each class (Table 3). The WQCs determined by comparing the value of range of WQI for each class.

Discriminant analysis: Discriminant Analysis (DA) is a classification method or more naturally occurring classes/groups (Astel *et al.*, 2006; Zhou *et al.*, 2007) to identify the variables that distinguish between the groups. It is a technique that could apply to classify objects into groups based upon one set of independent variables (Adomako *et al.*, 2011) and construct the discriminant function (Alberto *et al.*, 2001; Singh *et al.*, 2004, 2005). Discriminant functions are calculated using the following Eq. 1:

Table 3: Status of water quality and water use (DoE., 2014)

WQC[a]	WQI[b]	Water status	Water use
I	>92.7	Very good	Conservation of natural environment Water supply 1 practically no treatment is necessary fishery 1-Very sensitive aquatic species II
II	76.5-92.7	Good	Good IIA: water supply II-Conventional treatment required Fishery II-Sensitive aquatic species IIB: recreational use with body contact
III	51.9-76.5	Average	Water supply III-Extensive treatment required Fishery III: common,of economic value and tolerant species Livestock drinking
IV	31.0-51.9	Polluted	Irrigation
V	<31	Very polluted	None of the above

[a]WQC; Water Quality Class [b]WQI; Water Quality Index

Table 4: The different combinations of variables

Data	Variable sets
DS1	DO, BOD, COD, SS, pH and NH$_3$-N
DS2	Selection based on DA of DS1 (DS1 as IVs[a] and WQC[b] as Dv[c])

[a]IVs; Independent Variables [b]WQC; Water Quality Class [c]DV; Dependent Variable

$$f\left(G\right)_i = k_i + \sum_{j=1}^{n} w_{ij} P_{ij}$$

Where:

i   = The number of Groups $(G)_i$

k_i = The constant to each group

n   = The number of parameters to classify one set of data into a specified group

w_j = The weight coefficient for specific Parameter (P_j)

The weight coefficient is used to maximize the distance between the means of the dependent variable DA has been carried out in stepwise mode on raw water quality data (Singh *et al.*, 2004; Santos-Roman *et al.*, 2007; Boyacioglu and Boyacioglu, 2008; Su *et al.*, 2011). In stepwise mode, the variables are inserted one by one, starting from the most significant parameter until no significant changes obtained. DA was applied to the raw data grouping of the water quality into five groups defined by the NWQS. Groups (I-V) act as the dependent variable while the WQVs as independent variables. Consequently, a DA was conducted in this study to select the best sets of input variables for the ANN. Therefore, the division of water quality data used in ANN modelling is shown in Table 4.

Artificial neural network: The ANN can associate the input and output variables without requiring any prior knowledge related to variables involved (Cevoli *et al.*, 2011). The ANN applications for the data processing based on the analogy with biological neuron system
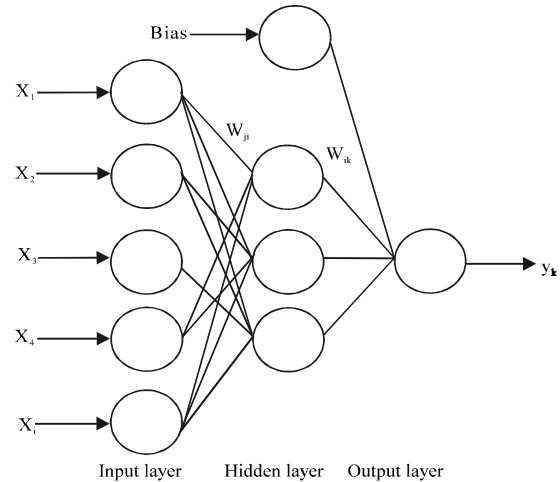

Fig. 2: A typical multilayer perceptron artificial neural network. In this figure, x_i is the ith input to the network, W_{ij} is interconnection weight between the ith input and jth hidden neuron, W_{ik} is interconnection weight between the hidden and output layers, y_k is "Net", network output

(Kruzlicova *et al.*, 2009). The neural network includes an input layer, a hidden layer and an output layer (Fig. 2). The basic unit of the neural network is the neuron. The formation of neural network generated by joining each neuron in the input layer to each neuron in the hidden layer and further connected to neurons in the output layer. The optimal number of neurons in the hidden layer determined by a trial-and-error procedure (Alizadeh and Kavianpour, 2015). A neural network should be trained before it can be put to use. The network is trained by repeatedly using a set of input-target data (Mandal and Roy, 2006) until there is no significant change in synaptic weights. The goal of the training to correct the position of the output vector closest to the target vector with improving the weight. Once, the training process has been completed, the network should be able to predict correctly any new data (Kruzlicova *et al.,* 2009). The hidden (x_i) and output (y_i) neuron is defined as follows:

$$x_i = f\left(v_i\right)$$

$$y_i = f\left(v_i\right)$$

where, $f(v_i)$ is the activation function applied in the hidden or output layers. The softmax function applied for classification is frequently carried out at the final layer of a network. From the statistical point, neural network model with softmax activation function could be seen as multiple logistic regression models (Bishop, 1995). The softmax functions are described by:

Table 5: Splitting schemes for ANN modelling

Splitting scheme	Training set (%)	Validation set (%)	Testing set (%)
I	80[a]	10	10
II	75	10	15
III	70	10	20
IV	70	15	15
V	65	15	20
VI	60	20	20
VII	60	15	25
VIII	50	25	25

[a]The percentage equivalent to the number of data

$$f(v_i) = \frac{e^{vi}}{\sum_{j=1}^{n} e^{vi}}$$

where, e is the Euler's number (e≈2.71828) and v_i is the net signal which corresponds to the sum of the weighted inputs from the previous layer given by:

$$v_i = \sum_{j=1}^{p} w_{ij} x_j + b_i$$

Where:

j = 1, p concerns neurons

x_j = The previous layer which precede the given neuron i

w_{ij} = The weight

b_i = The bias (offset)

An additional input (bias) added to the neuron that represents a threshold phenomenon. Internal activation level specified by the neurons that comprise the transfer function. The input converted to a suitable transfer function for determined the output (Bahramparvar *et al.*, 2014).

This research task aspired to produce an ANN Model predictive of the WQC using water quality parameters as predictors. Two different models were developed to predict river class based on two different sets of WQVs were applied as input variables. The models ANN1 and ANN2 were developed based on the datasets DS1 and DS2, respectively. The ANN design derived from water quality data randomly split into the training, testing and validation data sets with seven partitioning schemes (Table 5). The ANN's design is not restricted to the total number of samples and input variables (Anderson and Whitcomb, 2005). However, when the number of the training of data reduced, the quality of forecasts decreased significantly (Zhang and Friedrich, 2003).

The transfer function parameters measured during the training phase. Whereas the correct classification for determining neural network performance estimated during the testing phase. All phases should classify correctly to obtain an ideal network. In the training phase, the error is reduced to a minimum level by changing the weight. The output value O_i compared with a target value T_i in estimative the error functions. The training step was stopped when the Cross Entropy Error (CEE) converged and was <0.0001:

$$CEE = \frac{\sum_{i=1}^{N} T_i \ln O_i + (1 - T_i) \ln(1 - O_i)}{N}$$

Where:

O_i = The ith predicted output value

T_i = The ith observed target value

N = The number of data

RESULTS AND DISCUSSION

The ANN1 optimal architecture was 1:6-1:6-1:1 with 60-20-20% splitting plan. This model has 0% incorrect prediction in both the validation and testing phases (Fig. 3). Predictions of the WQC rendered by this model for the training set were very accurate (96.8% correct; PIP = 3.2). The CEE for training set was 3.44. During the training process, ANN1 could accurately predict class I, class II and IV. As regards class III, ANN1 predicted 77.8% correct during the training process. Nevertheless, during the validation phase, class III was predicted accurately for all cases (Table 6). On the other hand, class V was not recorded for any of the monitoring stations at Muda River during the study period.

Further, to improve WQC prediction for the Muda River, stepwise elimination DA was conducted to identify the most significant input variables to best describe the desired output (WQC) without losing much information. An analysis of the performance of ANN2 was extended by applying stepwise DA on the input and output data in order to identify the input variables discriminating the most between classes. Only three variables (DO, BOD and COD (DS2)) were found to be the best discriminators. Taking these analysis outputs into consideration, the researcher developed ANN2.

The ANN2's results showed that the optimum network architecture was 1:3-1:3-1:1 with the 65-15-20% partitioning scheme. The value of CEE for the training set was 7.2. The model's performance on the training set was very good. It had an accuracy of 98.4% and a PIP value of 1.6. The performance during the testing phase was very highly accurate (100%) and the PIP was equal to 0.0. However, the performance of this model on the validation set was the lowest, prediction accuracy was 91.7% (PIP = 8.3). Table 6 shows that this model has good performance in predicting class II and III during all phases (the training, testing and validation phases). In addition, ANN2 could predict class IV correctly during the training and testing phases. However, all the data for class I were predicted incorrectly during the training and validation

Table 6: Classification matrix for the ANN-WQC Model

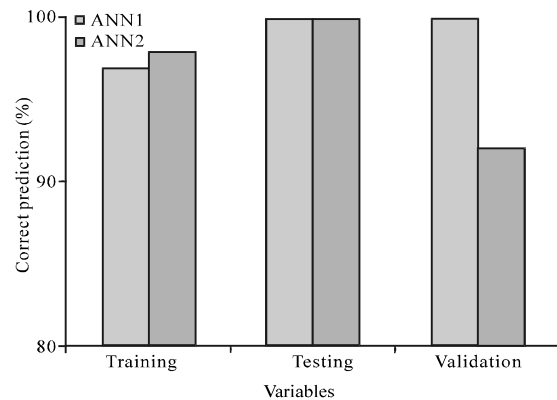| Sample observed/	Correct classification (%)	
Data set	River class	ANN2
Training		
I	100	0
II	100	100
III	78	100
IV	100	100
V	-	-
Overall (%)	97	98
Testing		
I	-	-
II	100	100
III	-	100
IV	-	100
V	-	-
Overall (%)	100	100
Validation		
I	-	0
II	100	100
III	100	100
IV	-	0
V	-	-
Overall (%)	100	92

Fig. 4: Comparison between the ANN models predictive of the WQC
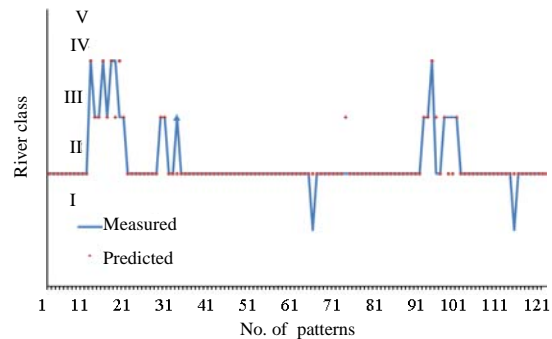

Fig. 5: Measured and predicted of river class

the fact that the PIP of ANN1 was 0% in prediction phase. Therefore, Model ANN1 used the 60-20-20% splitting plan with the six WQVs (DO, BOD, COD, pH, SS, NH_3-N) was selected to predict the WQC in the MRB. Figure 5 compares between ANN1-predicted and measured river class for each individual observation.

The ANN approach presented in this study demonstrates the potential of producing models to foresee the WQC in the river basin for any monitoring station and period. Thus, the ANN is a useful alternative and is more powerful than traditional methods of water quality classification that are determined by the WQI. Computation value of the WQI involves lengthy calculations and the use of multiple sub-index formulae for each parameter. The study provides a protocol of water quality classification by ANN technique. The method can be applied successfully to any river basin in the world.

CONCLUSION

The study described the application of ANN with the softmax activation function to developing a model capable

Fig. 3: Model performance according to splitting plan in a) ANN1 and b) ANN2

phases. In this respect, class V was not recorded for any of the monitoring stations at the Muda River during the study period. Overall, the percentage of correct prediction of all phases for ANN1 and ANN2 represented in Fig. 4. The results show that performances of ANN1 and ANN2 were almost similar, both models had perfect prediction power (100%) in the testing phase, despite

of prediction the WQC under NWQS. The different models developed with two separate combinations of WQVs as input variables. The modelling efforts suggests that the optimal network architecture was 1:6-1:6-1:1 with the 60-20-20% splitting plan with the six WQVs (DO, BOD, COD, pH, SS, NH_3-N). Predictions of the WQC rendered by this model for the training set were very accurate (96.8% correct; PIP = 3.2; CEE = 3.44). The study results emphasized that the ANN can be applied to predict rapidly and efficiently the WQC of the river water. The findings should encourage the manager of the monitoring system classifies water quality using ANN method to replace traditional methods. Therefore, the ANNs are recommended for the classification of river water quality.

ACKNOWLEDGEMENT

REFERENCES

Abrahao, R., M. Carvalho, W.R. Da Silva, T. Machado and C. Gadelha *et al.*, 2007. Use of index analysis to evaluate the water quality of a stream receiving industrial effluents. Water SA, 33: 459-466.

Adamowski, J. and K. Sun, 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J. Hydrol., 390: 85-91.

Adomako, D., A. Gibrilla, T.T. Akiti, R. Fianko and P. Maloszewski, 2011. Hydrogeochemical evolution and groundwater flow in the Densu River Basin, Ghana. J. Water Resour. Prot., 3: 548-561.

Alberto, W.D., D.M. del Pilar, A.M. Valeria, P.S. Fabiana, H.A. Cecilia and B.M. de los Angeles, 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River Basin (Cordoba-Argentina). Water Res., 35: 2881-2894.

Alizadeh, M.J. and M.R. Kavianpour, 2015. Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. Mar. Pollut. Bull., 98: 171-178.

Alobaidy, A.H.M.J., H.S. Abid and B.K. Maulood, 2010. Application of water quality index for assessment of Dokan lake ecosystem, Kurdistan region, Iraq. J. Water Res. Prot., 2: 792-798.

Anderson, M.J. and P.J. Whitcomb, 2005. RSM Simplified-Optimizing Processes Using Response Surface Methods for Design of Experiments. Productivity Press, New York.

Astel, A., M. Biziuk, A. Przyjazny and J. Namiesnik, 2006. Chemometrics in monitoring spatial and temporal variations in drinking water quality. Water Res., 40: 1706-1716.

Bahramparvar, M., F. Salehi and S.M. Razavi, 2014. Predicting total acceptance of ice cream using artificial neural network. J. Food Process. Preserv., 38: 1080-1088.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Oxford, UK., ISBN-13: 9780198538646, Pages: 482.

Bordalo, A.A., R. Teixeira and W.J. Wiebe, 2006. A water quality index applied to an international shared river basin: The case of the Douro River. Environ. Manage., 38: 910-920.

Boyacioglu, H. and H. Boyacioglu, 2008. Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. Environ. Geol., 54: 275-282.

Cevoli, C., L. Cerretani, A. Gori, M.F. Caboni and T.G. Toschi *et al.*, 2011. Classification of Pecorino cheeses using electronic nose combined with artificial neural network and comparison with GC-MS analysis of volatile compounds. Food Chem., 129: 1315-1319.

Chaves, P. and F.J. Chang, 2008. Intelligent reservoir operation system based on evolving artificial neural networks. Adv. Water Resour., 31: 926-936.

Chaves, P. and T. Kojiri, 2007. Deriving reservoir operational strategies considering water quantity and quality objectives by stochastic fuzzy neural networks. Adv. Water Resour., 30: 1329-1341.

Cude, C.G., 2001. Oregon water quality index a tool for evaluating water quality management effectiveness1. JAWRA. J. Am. Water Resources Assoc., 37: 125-137.

DoE., 2014. Malaysia environmental quality report 2013. Department of Environment, Ministry of Natural Resources and Environment, Kuala Lumpur, Malaysia.

Dogan, E., B. Sengoru and R. Koklu, 2009. Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. J. Environ. Manage., 90: 1229-1235.

Gazzaz, N.M., M.K. Yusoff, A.Z. Aris, H. Juahir and M.F. Ramli, 2012. Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Mar. Pollut. Bull., 64: 2409-2420.

Glezakos, T.J., T.A. Tsiligiridis, L.S. Iliadis, C.P. Yialouris and F.P. Maris *et al.*, 2009. Feature extraction for time-series data: An artificial neural network evolutionary training model for the management of mountainous watersheds. Neurocomputing, 73: 49-59.

Huang, M., S. Lee Barbour, A. Elshorbagy, J.D. Zettl and B. Cheng Si, 2011. Infiltration and drainage processes in multi-layered coarse soils. Can. J. Soil Sci., 91: 169-183.

Huo, Z., S. Feng, S. Kang, G. Huang and F. Wang *et al.*, 2012. Integrated neural networks for monthly river flow estimation in arid inland basin of Northwest China. J. Hydrol., 420: 159-170.

Iliadis, L.S. and F. Maris, 2007. An Artificial Neural Network model for mountainous water-resources management: The case of Cyprus mountainous watersheds. Environ. Model. Software, 22: 1066-1072.

Julien, P.Y., M. Asce, A.A. Ghani, N.A. Zakaria and R. Abdullah, 2009. Case study: Flood mitigation of the Muda River, Malaysia. J. Hydraul. Eng., 136: 251-261.

Kentel, E., 2009. Estimation of river flow by artificial neural networks and identification of input vectors susceptible to producing unreliable flow estimates. J. Hydrol., 375: 481-488.

Khalil, B., T.B.M.J. Ouarda and A. St-Hilaire, 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. J. Hydrol., 405: 277-287.

Khashei, M. and M. Bijari, 2011. Which methodology is better for combining linear and nonlinear models for time series forecasting? J. Ind. Syst. Eng., 4: 265-285.

Kim, J.W. and Y.A. Pachepsky, 2010. Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. J. Hydrol., 394: 305-314.

Kruzlicova, D., J. Mocak, B. Balla, J. Petka, M. Farkova and J. Havel, 2009. Classification of Slovak white wines using artificial neural networks and discriminant techniques. Food Chem., 112: 1046-1052.

Liou, S.M., S.L. Lo and S.H. Wang, 2004. A generalized water quality index for Taiwan. Environ. Monit. Assess., 96: 35-52.

Mandal, A. and P. Roy, 2006. Modeling the compressive strength of molasses-cement sand system using design of experiments and back propagation neural network. J. Mater. Processing Technol., 80: 167-173.

Melesse, A.M., S. Ahmad, M.E. McClain, X. Wang and Y.H. Lim, 2011. Suspended sediment load prediction of river systems: An artificial neural network approach. Agric. Water Manage., 98: 855-866.

Rajaee, T., 2011. Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers. Sci. Total Environ., 409: 2917-2928.

Sanchez, E., M.F. Colmenarejo, J. Vicente, A. Rubio, M.G. Garcia, L. Travieso and R. Borja, 2007. Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. Ecol. Indicators, 7: 315-328.

Santos Simoes, F., A. Moreira, M. Bisinoti, S. Gimenez and M. Santos Yabe, 2008. Water quality index as a simple indicator of aquaculture effects on aquatic bodies. Ecol. Indic., 8: 476-484.

Santos-Roman, D.M., G.S., Warner and F. Scatena, 2007. Multivariate analysis of water quality and physical characteristics of selected watersheds in Puerto Rico. J. Am. Water Resour. Assoc., 39: 829-839.

Semiromi, F.B., A.H. Hassani, A. Torabian, A.R. Karbassi and F.H. Lotfi, 2011. Water quality index development using fuzzy logic: A case study of the Karoon River of Iran. Afr. J. Biotechnol., 10: 10125-10133.

Singh, K.P., A. Bansant, A. Malik and G. Jain, 2009. Artificial neural network modeling of the river water quality-A case study. Ecol. Model., 220: 888-895.

Singh, K.P., A. Malik and S. Sinha, 2005. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques-a case study. Analytica Chimica Acta, 538: 355-374.

Singh, K.P., A. Malik, D. Mohan and S. Sinha, 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)-A case study. Water Res., 38: 3980-3992.

Su, Y.S., A.E. Gelman, J. Hill and M. Yajima, 2011. Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. J. Stat. Software, 45: 1-31.

Suratman, S., M. Awang, L.A. Ling and N.M. Tahir, 2009. Water quality index study in paka river basin, terengganu (in Malay). Sains Malaysiana, 38: 125-131.

Thi Minh Hanh, P., S. Sthiannopkao, D.T. Ba and K.W. Kim, 2010. Development of water quality indexes to identify pollutants in Vietnam's surface water. J. Environ. Eng., 137: 273-283.

Tyagi, S., B. Sharma, P. Singh and R. Dobhal, 2013. Water quality assessment in terms of water quality index. Am. J. Water Res., 1: 34-38.

Wang, K.W., L.C. Chang and F.J. Chang, 2011. Multi-tier interactive genetic algorithms for the optimization of long-term reservoir operation. Adv. Water Resour., 34: 1343-1351.

Wu, C.L. and K.W. Chau, 2010. Data-driven models for monthly streamflow time series prediction. Eng. Applic. Artif. Intell., 23: 1350-1367.

Yoon, H., S.C. Jun, Y. Hyun, G.O. Bae and K.K. Lee, 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. J. Hydrol., 396: 128-138.

Zainudin, Z., 2010. Benchmarking river water quality in Malaysia. Inst. Eng. Malaysia IEM. Jurutera, 1: 12-15.

Zampella, R.A., J.F. Bunnell, K.J. Laidig and N.A. Procopio, 2006. Using multiple indicators to evaluate the ecological integrity of a coastal plain stream system. Ecol. Indic., 6: 644-663.

Zhang, Z. and K. Friedrich, 2003. Artificial neural networks applied to polymer composites: A review. Compos. Sci. Technol., 63: 2029-2044.

Zhou, F., Y. Liu and H. Guo, 2007. Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern new territories, Hong Kong. Environ. Monit. Assess., 132: 1-13.