Sains Malaysiana 41(8)(2012): 939–947

Supervised and Unsupervised Artificial Neural Networks for Analysis of Diatom Abundance in Tropical Putrajaya Lake, Malaysia

(Rangkaian Neural Buatan Diselia dan Tanpa Penyeliaan untuk Analisis Kelimpahan Diatom di Tasik Tropika Putrajaya, Malaysia)

M. Sorayya, S. Aishah & B. Mohd. Sapiyan*

ABSTRACT

Five years of data from 2001 until 2006 of warm unstratified shallow, oligotrophic to mesothropic tropical Putrajaya Lake, Malaysia were used to study pattern discovery and forecasting of the diatom abundance using supervised and unsupervised artificial neural networks. Recurrent artificial neural network (RANN) was used for the supervised artificial neural network and Kohonen Self Organizing Feature Maps (SOM) was used for unsupervised artificial neural network. RANN was applied for forecasting of diatom abundance. The RANN performance was measured in terms of root mean square error (RMSE) and the value reported was 29.12 cell/mL. Classification and clustering by SOM and sensitivity analysis from the RANN were used to reveal the relationship among water temperature, pH, nitrate nitrogen (NO_3-N) concentration, chemical oxygen demand (COD) concentration and diatom abundance. The results indicated that the combination of supervised and unsupervised artificial neural network is important not only for forecasting algae abundance but also in reasoning and understanding ecological relationships. This in return will assist in better management of lake water quality.

Keywords: ; Diatom; forecasting; recurrent artificial neural network; self organizing maps

ABSTRAK

Data selama lima tahun dari 2001 hingga 2006 bagi tasik tropika yang cetek dan panas, berstatus oligotrof ke mesotropi iaitu Tasik Putrajaya, Malaysia telah digunakan untuk mengkaji penemuan corak dan ramalan kuantiti diatom menggunakan rangkaian neural buatan yang diselia dan tidak diselia. Rangkaian neural buatan berulang (RANN) telah digunakan untuk rangkaian neural buatan diselia dan peta atur sendiri Kohonen (SOM) telah digunakan untuk rangkaian neural buatan tanpa pengawasan. RANN telah digunakan untuk ramalan kuantiti diatom. Prestasi RANN diukur daripada ralat min punca kuasa dua (RMSE) dan nilai yang dilaporkan adalah 29.12 sel/mL. Pengelasan dan kelompok oleh SOM dan analisis kepekaan daripada RANN digunakan untuk mendedahkan hubungan antara suhu air, pH, kepekatan nitrogen nitrat (NO_3-N), keperluan oksigen kimia (COD) dan kuantiti diatom. Keputusan menunjukkan bahawa gabungan rangkaian diselia dan tidak diselia neural buatan adalah penting bukan sahaja untuk ramalan pertumbuhan alga tetapi juga dalam analisis dan pemahaman hubungan ekologi. Ini akan membantu dalam pengurusan yang lebih baik bagi kualiti air tasik.

Kata kunci: Diatom; peta susun sendiri; ramalan; rangkaian neural buatan berulang

Introduction

This study investigated the application of supervised and unsupervised artificial neural networks (ANN) in the analysis of diatom abundance and relationship to the water quality parameters of tropical Putrajaya Lake, Malaysia. Diatom is commonly used as indicators of environmental conditions in lake and for its significant importance in ecosystems as indicators of environmental conditions. Indication of water quality status using diatoms provides accurate assessment as compared with using chemical and zoological assessment (Leclercq 1988). Diatom respond directly and sensitively to many physical, chemical and biological changes in water ecosystems. Shifts in abundance of algae species such as diatom can be used to detect environmental changes and can be use to indicate trophic status and nutrient problems in lake (Patrick 1994).

In Malaysia, diatoms were used to study pollution status of Linggi River Basin. It was reported that there was significance difference in diatom species variation between polluted and unpolluted stations (Nather Khan 1985, 1990, 1991). Another study conducted by Maznah and Mansor (1999) on diatom diversity and its relation to river pollution, concluded that diversity values can be related to changes in water quality. The study of the factor leading to algae succession, remains a difficult task as algal populations are determined by a compound collection of biotic and abiotic factors (Carrillo et al. 1995; Sommer 1989; Vanni & Temte 1990). Recknagel et al. (2005, 2006) reported that ANN are capable of predicting phytoplankton abundance and also provides enhanced understanding of processes and environmental circumstances that accelerates the growth and reaction of a particular algal populations to diverse management practices. This can be further explained by Zhang et al. (1998) that ANN are able to learn from examples and respond to complex functional relationships within the data, even when the underlying relationships are unknown or difficult to describe.

Another important feature of ANN is that they can generalize and infer or predict event or occurrence even if the sample input data is chaotic. Maier et al. (1998), reported that ANN can be used when limited dataset are available and when the data are highly diverse, and when the relationships between cause and effect are indistinguishable (Schultz et al. 2000). Therefore, ANN can be considered as a promising statistical approach that offers a non-linear statistical modeling method for discovering relationships in ecological data (Colasanti 1991; Edwards & Morse 1995; Lek et al. 1996).

ANN has been successfully applied to tropical lakes in Malaysia for water quality modeling (Talib et al. 2009; Najah et al. 2009; Elfithri et al. 2011; Gasim et al. 2006). However in Malaysia, few algal studies have been conducted in relation to water pollution. Moreover, most of the findings from water quality studies remains unpublished (Maznah 2010).

Understanding of algae dynamics using ANN in tropical water bodies are not given much attention. Therefore, by using data of tropical Putrajaya Lake the study aimed at forecasting diatom abundance in Putrajaya Lakes and determination of the population sensitivities against physical and chemical lake properties by means of supervised ANN; analyzing complex interaction among diatom abundance and water temperature, pH, nitrate nitrogen (NO_3-N) and chemical oxygen demand (COD) by means of unsupervised ANN. Results from the sensitivity analysis by means of supervised ANN were taken into a context with data classification and clustering by unsupervised ANN in order to test hypotheses on complex interactions of algae population with environmental conditions as postulated in literature findings. Putrajaya Lake has been selected in this study as diatoms comprises majority of the algal population and high volume of data is available that is required for conducting studies related to ANN.

Material and Methods

Study Site and Data

The diatom data used in the study are the monthly data collected from Putrajaya Lake, Malaysia. Putrajaya Lake and Wetlands is an oligotrophic to mesotrophic lake located at the south of the densely inhabited Klang Valley. It is a warm polymictic lake which is shallow to develop thermal stratification. Putrajaya Lakes was developed in 1997. The ultimate purpose of Putrajaya Lake is for primary contact. The Putrajaya Lake is divided into a few zones to cater for different types of activities at designated zones. Therefore it is important to ensure the lake water quality meets the requirements of Putrajaya Lake Water Quality Standards (Putrajaya Corporation 1998).

Limnological parameter data used was compiled over a period of five years (August 2001 – May 2006). The input parameters selected for the present study was determined using correlation analysis method. The parameters selected were those that showed higher importance in affecting diatom abundance. Once the importance of each variable has been determined, backward elimination method was used to eliminate less important variables and the network was retrained with reduced number of variables. This procedure was repeated, until the discarding of any extra variables did not improve the model performance. This process resulted in the inclusion of water temperature, pH, chemical oxygen demand (COD) and nitrate nitrogen for the model development.

Data samples were collected from 23 sampling monitoring stations. Sixteen sampling stations were used for training and the 7 stations were used for testing. The data was categorized into two different sets arbitrarily as there was no significance difference among the stations and to avoid producing biased results. Cross validation subset is not being utilized as it is more suitable to be used when data sets are too few and not reasonable to be divided into training and testing (Maindonald & Braun 2007). Furthermore, application of cross validation dataset in time series analysis does not seems to improve the results.

Sampling procedures, including preservation for water quality parameters were carried out in accordance with WHO-GEMS (1987) and APHA (1995). The analytical methods for the measured parameters were adopted from manual published by the American Public Health Association (APHA 1995). Algae samples (cell mL^{-1}) were collected using 50 mL vials and preserved in four percent formalin. Net samples were obtained using plankton net with mesh size of about 30 um. Species of algal were then identified by means of a Nikon light microscope (× 1000) (Salleh 1996) and algae counts were made using the sedimentation inverted microscope technique. Summary statistics of all the variables used in this study are given in Table 1.

Supervised ANN

Recurrent artificial neural network (RANN) was used for the supervised ANN. RANN structure was adopted in this study as it has been reported to success fully predict algae growth at temperate lake (Gurbuz et al. 2003; Jeong et al. 2001; Recknagel 2001). RANN modeling approach are explained as the system state at time (t) is calculated by system state at time (t-1) and the duplicated weights of time (t-1) is used as response input to establish weights at time t (Pineda 1987). One hidden layer RANN with back-propagation through time learning was employed in this study. It has been reported that only one hidden layer is needed to approximate any continuous function (Cybenko 1989). The Back-propagation through time (BPTT) algorithm was based on changing the network from a feedback system to feed-forward system by folding the network over time. The network uses momentum learning

TABLE 1. Limnological properties of variables from Putrajaya Lake and Wetlands (2001 – 2006)

Limnological Variables	Mean	Minimum	Maximum	Standard Deviation
Input Variables				
Water Temperature °C	30.35	26.76	34.54	1.09
pH	7.25	6.00	8.94	0.48
NO_3-N (mg/L)	1.14	0.00	4.88	0.93
COD (mg/L)	16.82	1.00	79.00	10.11
Output Variable(cell/ml)				
Diatom	24	1	418	37.6 5

algorithm to determine the weights in the network. This algorithm is considered as an improvement to the gradient-descent search, where previous increment to the weight is used to speed up and stabilize convergence. The equation to update the weights is as follows:

$$w_{ij}(n +1) = w_{ij} (n) + \eta\partial_i(n)x_j(n) + \alpha(w_{ij} (n) - w_{ij}(n-1),$$

where α is the momentum constant, which should take on values between 0.5 and 0.9. Weights adaptation using momentum learning is changed proportionally based on how much they were updated in the last iteration. Momentum learning is considered as a robust method to speed up learning. It is suitable as the default search rule for networks with nonlinearities (Principe et al. 1999). Hyperbolic tangent function is used in this study as an activation function at the hidden layer and at the output layer. This is to introduce nonlinearity to the system as the growth of algae is known to be nonlinear process (Thomann & Mueller 1987):

$$f(x) = tanh(\alpha x),$$

where α is a slope parameter and normally is set to 1. The hyperbolic tangent produces values between [-1 and 1]. The desired output requires to be normalized to the range of (0 and 1) at the output layer. This enables the network output to be compared to the desired output. Normalization method applied was zero mean and unit variance transformation. To summarize the RANN for the current study were designed with one hidden layer using following network parameters which were identified via trial and error; learning rate = 0.01, momentum term = 0.7, epoch = 100, number of nodes 4-4-1(number of inputs-number of hidden nodes-number of outputs). The number of nodes used in this study was based on trial and error where further increasing the number of nodes does not significantly improves the results obtained. The stopping criteria on was based on MSE value 0.01. Performance validation criteria for the RANN developed are based on root mean square error (RMSE). Data set A would be used for training meanwhile dataset B for testing the network. Sensitivity analysis was carried out using trained RANN to understand the relationships between the input variables and the diatom population. The trained RANN weights was fixed prior to performing sensitivity analysis. Sensitivity analysis measures how much a small change in one of the independent variables affects the functional value. It effectively measures change in a given input affects the output across the training data set. The sensitivity analysis for input k used in this study based on:

$$s_k = \frac{\sum_{p=1}^{p}\sum_{i=1}^{o}\left(y_{ip} - \overline{y}_{ip}\right)^2}{\sigma_k^2},$$

where $\overline{y}_{ip}$ is the ith output obtained with the fixed weights for the pth pattern, o is the number of network outputs, P is the number of patters, and σ_k^2 is the variance of the input perturbation (Principe et al. 1999). The result from sensitivity analysis was validated against results from unsupervised ANN (SOM).

UNSUPERVISED ANN

Type of unsupervised ANN, Kohonen Self Organizing feature Maps (SOM) as introduced by Kohonen (1989) was applied to ordinate, cluster and map diatom population data with respect to water quality input variables. SOM was developed using Matlab and SOM toolbox Matlab (2006). SOM reduces data dimensions by producing a map of 1 or 2 dimensions which plot the similarities of the data by grouping similar data items together. Thus SOM reduce dimensions and display similarities. This enables the discovery or identification of features or patterns of most relevance through data reduction and projection. SOM has been used in ecological modeling to find similarity between dataset (Chon 1996; Foody 1999). In the Kohonen network, every node in the input layer are represented as vector $\mathbf{x}_i$ and are connected to each neuron, j. This connectivity constituted as weights, $w_{ij}(t)$, adaptively varying at each iteration of t. The weights were arbitrarily assigned in a small value at first. As the input vector was sent through the network, each neuron computes the summed distance between the weight and input as follows:

$$d_j(t) = \sum_{i=0}^{N-1} \left(x_i - w_{ij}(t) \right)^2 .$$

The winning neuron was selected based on neuron that responds greatly to a given input vector. The winning neuron has the weight vector which has the shortest distance to the input vector. The winning neuron and maybe its neighboring neurons were allowed to learn by altering the weights in a way to additionally decrease the Euclidian distance among the weight and the input vector via the following equation:

$$w_{ij}(t+1) = w_{ij}(t) + \eta\,(t)(x_i - w_{ij}(t))Z_j.$$

where Z_j is assigned 1 for the winning and neighboring neurons while 0 is assigned for the other neurons, and $\eta(t)$ represents the fractional increase of the alteration (Kohonen 1989). The present study utilized the Kohonen Self Organizing Feature Maps (SOM) to ordinate algae data with respect to ranges of nitrate nitrogen (NO_3-N), pH, water temperature and chemical oxygen demand (COD) concentrations. As a result of the training of the unsupervised ANN, the euclidian distance between the inputs are calculated and visualized as distance matrix (U-matrix) and a partition map (K-means). Classification criteria used for the cluster map are based on the Interim National Water Quality Standards for Malaysia (INWQS) and statistical distribution of data at Putrajaya Lake and Wetlands. Table 2 illustrates the clustering criteria used in this study.

Results

Forecasting Diatom Abundance

The RANN was trained using Dataset A to predict diatom abundance. Dataset B was then used for testing purposes to avoid biasness in result. Figure 1 illustrates forecasting result for diatom using testing dataset B. Root mean square error (RMSE) cell/mL was used to quantitatively evaluate the model performance. RMSE is a measure of the average level of prediction error. It indicated the absolute fit of the model to the data or how close the observed data points are to the model's predicted values. It was shown in the following formula where y is the observed value, $\tilde{y}$ is the predicted value, n is the number of readings used, and j is the individual reading of the value:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j - y_j \right)^2} .$$

The r value is a measure of correlation between the predicted and observed values of the independent variable. The r value indicates agreement between predicted and observed values but it does not indicate the performance of the models. The r value for training data is 0.7 and testing data is 0.4. The ideal values for RMSE are 0, however it is impossible to achieve ideal values when real environmental models are in concern. The RMSE value reported is approximately 0.13 cell/mL for training dataset A and 29.12 cell/mL for testing dataset B. Anova test has been conducted to determine any significant difference between observed and predicted diatom abundance for Dataset B. A value of $p > 0.005$ is achieved. RMSE however is not considered as an ideal fitness measure but as a suitable error measure. The RMSE places greater emphasis on larger forecasting error then on the pattern of the predicted algae abundance. Therefore, a visual inspection is required of the actual and predicted results in addition to calculating the RMSE (Bowden et al. 2006). Visual inspection of the graph produced for diatom revealed that high abundance is not predicted well by the RANN. Visual inspection of the graphs produced for diatom (Figure 1) reveals that year 2001 is not being predicted well by the RANN model. High abundance of diatom is only reported in year 2001 and its abundance reduced thereafter. This explains the inability of RANN model developed to predict well for the year 2001.

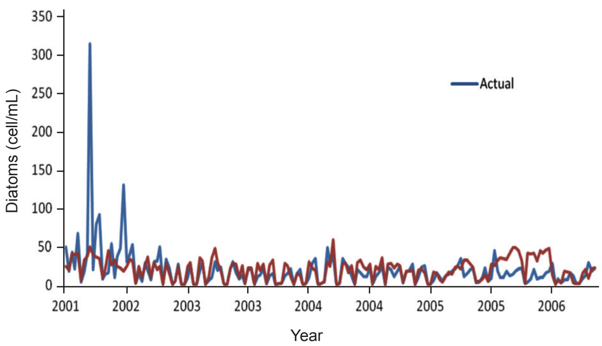


FIGURE 1. Forecasting of diatom using RANN for Putrajaya Lake

TABLE 2. Classification criteria for ordination and clustering of time-series data of Putrajaya Lakes

Measured Parameters	Classification Criteria
Water temperature °C	< 30.5 ; 30.5 ≥ and < 31.5 ; ≥ 31.5
pH	< 7; 7 ≥ and < 8.0 ; ≥ 8.0
NO_3-N (mg/L)	< 1.0 ; 1.0 ≤ and < 1.5 ; ≥ 1.5
COD (mg/L)	≤ 5.0 ; 5 < and ≥ 20 ; > 20

Relationships Between Diatom Abundance and Water Quality Conditions

Sensitivity analysis by means of RANN as well as classification and clustering by means of SOM was carried out to study the relationship between diatom abundance and water quality conditions. Figure 2 illustrates the relationships between water temperature and diatom. Diatoms were abundant at the water temperature below 30.5°C namely between 29.3°C and 29.5°C. Diatom abundance was reduced at water temperature 30.5°C and higher, and their abundance is also low at water temperature below 29.0°C. In Figure 3, relationship between pH and diatom are represented. Diatoms are mapped by SOM at pH range of 7 to 8, slightly neutral to alkaline conditions. High abundance of the diatom was observed at pH higher than 7.3 from the sensitivity curve. Figure 4 reflects the preferences of relatively low concentrations of NO_3-N. The relationship of diatom to COD concentrations showed similar trends represented in Figure 5, where preferences were for the low concentration of COD.

Discussion

Forecasting Diatom Abundances and Succession

The results (Figure 1) demonstrated that supervised RANN have the capacity for forecasting abundance of diatom at tropical lake. However more data on higher abundance of diatom are required to enable better forecasting results. The RMSE values reported are within the range of 29.12 cell/mL with a value of $p > (0.05)$. This value conforms well to the value cited in literature for similar ANN models developed for temperate lake. The RMSE value reported for temperate lake is between 256 and 577.9 cell/mL (Bowden et al. 2006). Meanwhile a value of $p > (0.05)$ indicated that there is no significance difference between the observed and predicted value of diatom abundance by the RANN model. The large RMSE error for Putrajaya Lake is due to the fact that the lake is an oligotrophic lake that is still maturing. The number of incidence for high abundance of algae for all division is less frequently reported and thus the trend is not been captured well by the RANN. The number of algae increased from 2001 and 2006 as the lake was maturing and thus there was no clear trend developed for high abundance of algae which resulted in poor prediction for high abundance of algae. However, from the inspection of graph generated the low to medium abundance of diatom were predicted well. Therefore the capability of RANN model as a predictive tool for tropical lake cannot be disregarded at all.

Relationships Between Diatom Abundances and Water Quality

The results in Figures 2 to 5 show that ordination and clustering by unsupervised ANN – SOM and sensitivity analysis by supervised RANN can be integrated as a powerful tool for analysing complex ecological relationships. The tropical Putrajaya lake was unstratified shallow, oligotrophic, diatom are abundance at the water temperature below 30.5°C (Figure 2). Diatom abundance reduced at water temperature 30.5°C and higher unlike other tropical alge such as Cyanobacteria, Chlorophyta and Chrysophyta which are also found at Putrajaya Lake. However water temperature recorded for diatom abundance was much higher than at temperate lake. Diatom was recorded to be abundant at water temperature 16°C at unstratified shallow, eutrophic Lake Kasumigaura (Japan) (Recknagel et al. 2006). Reynolds (1984) reported that diatom species especially pennate forms are dominant at low temperature at the temperate lakes. It was further reported that low water temperature favors growth of pennate type of diatom species (*Asterionella, Diatoma, Fraglilaria and Tabellaria*) due to their efficient photosynthetic ability, cell size, surface area and inactivity of filter feeder. The pennate forms are also found in Putrajaya Lakes such as *Diatoma, Fragilaria* and *Tabellaria* which are abundant at higher temperature at tropical water bodies. Darley (1982) reported that optimum water temperature for many algae lie in the range of 18°C to 25°C. However, many literature works are carried out in temperate lakes.

To conform the finding for diatom relationship to water temperature in this study, data from Kutty et al. (2001) Lake Chini, Pahang, Malaysia was compared. The author reported that the water temperature recorded at Lake Chini is between 27°C and 31°C, where diatom are being observed. With regards to pH value, diatom are mapped by SOM at pH ranges of 7 to 8. High abundance of the diatoms is observed at pH higher than 7 from the sensitivity curve. Diatom abundance at eutrophic Lake Kasumigaura coincide with high pH values greater than 9 (Recknagel et al. 2006). This observation conforms with the literature and results from the sensitivity curve, where diatom abundance is related to relatively high water conductivity and pH. However, some species are acidophilous and prefer soft water (Kingston 1982; Round 1984). Some of these acidophilous species are common in Putrajaya lakes, such as *Frustulia rhomboides* and *Tabellaria flocculosa*. This was explained by the distribution of some of diatom species at lower pH (Figure 3).

In relation to pH and CO_2 concentration, Talling (1976) reported from a series of experiments that diatom appear to be photosynthesis tolerant regarding high pH and low CO_2 concentration. The extraction of CO_2 from water increases the pH as bicarbonate and carbonate levels are depleted to replenish the extracted CO_2 (Sawyer et al. 1994). The elucidation of diatom with regards to NO_3-N concentration (Figure 4) gave evidence that that the diatoms reached the highest abundance when NO_3-N concentration was low in concentration. These findings were supported by corresponding sensitivity curve showing that diatom peaked at low NO_3-N concentration about 0.2 mg/L. Diatoms at eutrophic Lake Kasumigaura peaked at medium NO_3-N concentrations 0.9 mg/L (Recknagel et al. 2006). Werner (1977) reported diatoms of oligotrophic lake (e.g. Putrajaya Lake) prefers lower amount of nitrates as compared with diatoms at eutrophic lake (e.g. Lake Kasumigaura). The ordination and clustering using SOM showed that COD values in the ranges of 5 mg/L to 20 mg/L were related to higher abundance of diatom at Putrajaya Lake (Figure 5). COD value is less than 20 mg/L in unpolluted water and more than 200 mg/L in polluted water (Gray 1999). The abundance of diatom at COD value of 5 mg/L to 20 mg/L shows diatom preferences of unpolluted

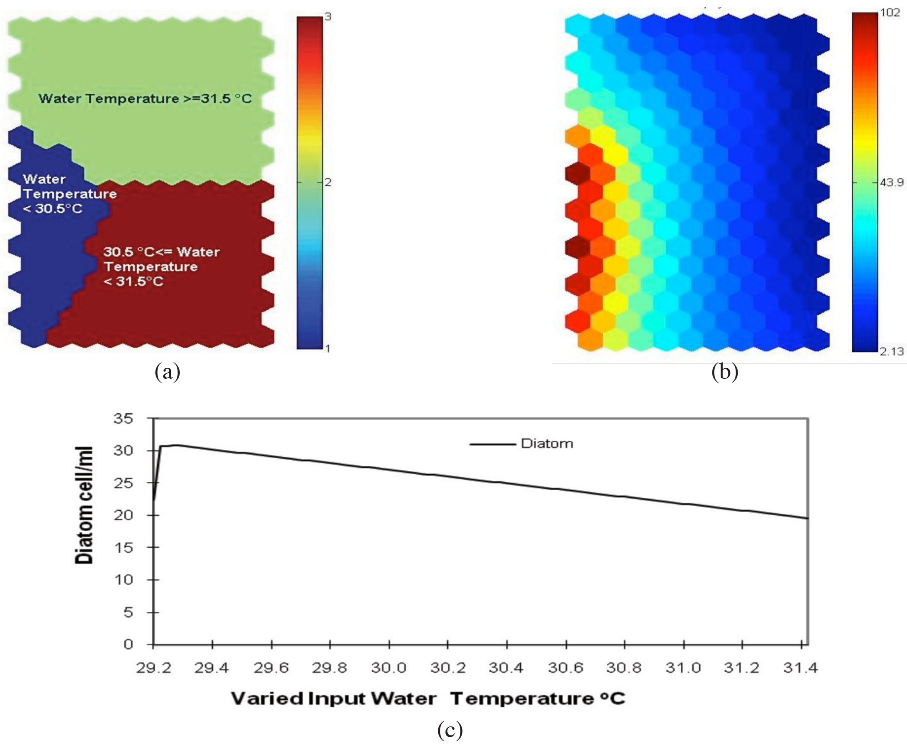


FIGURE 2. Cluster map of diatom over water temperature (°C) (a) diatom component plane, (b) using SOM and sensitivity curves of diatom abundance over water temperature (°C) range using RANN and (c) of Putrajaya Lake

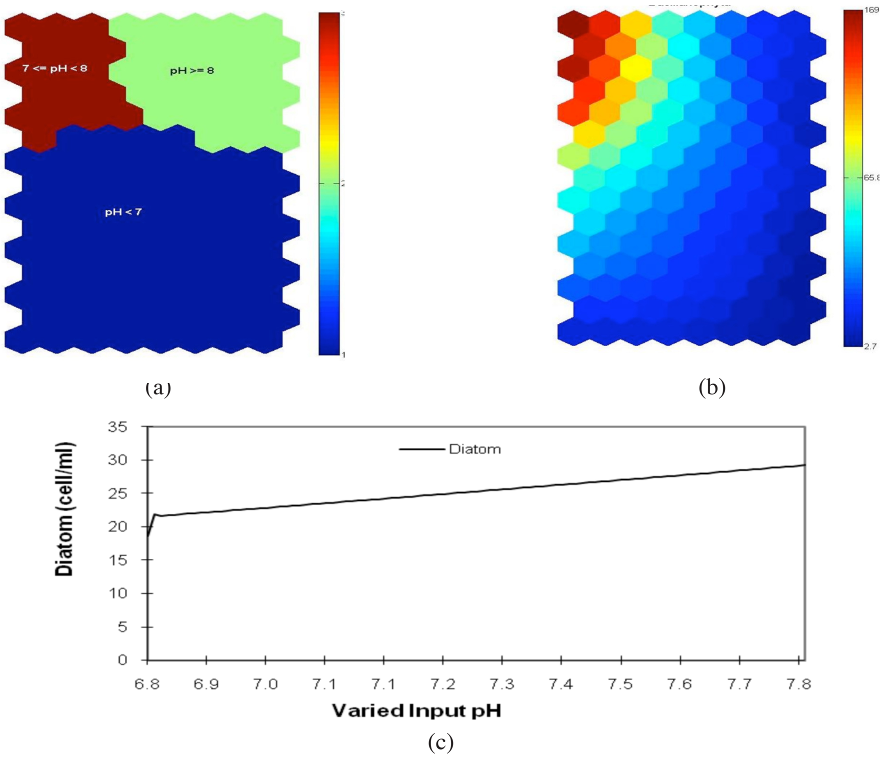
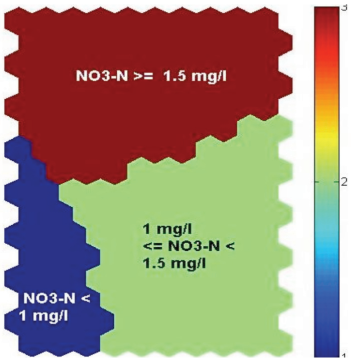
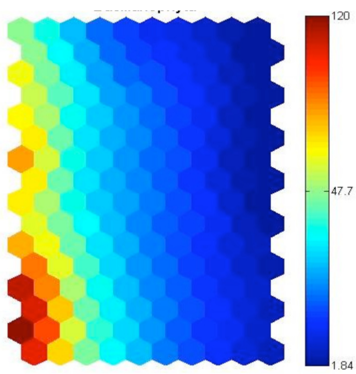


FIGURE 3. Cluster map of diatom abundance over pH (a) diatom component plane, (b) using SOM and sensitivity curve of diatom abundance over varied pH range using RANN and (c) of Putrajaya Lake

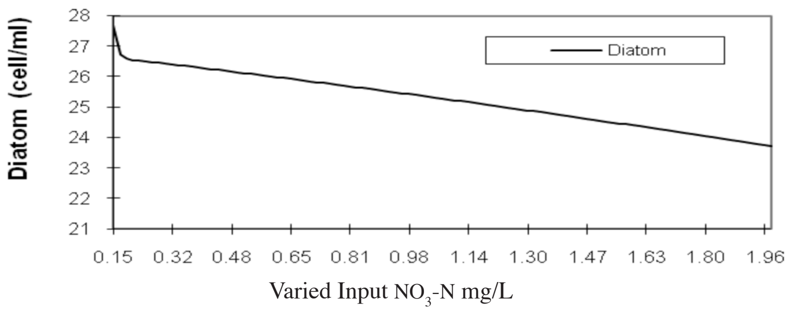
water. Furthermore the sensitivity curve indicates that high diatom abundance is noted at COD at about 8 mg/L after which the abundance reduces. However from the sensitivity curve, there is no significant reduction in diatom abundance with increase in COD concentration. This may be due to the fact that COD concentration level recorded at Putrajaya Lake falls under the category of unpolluted water, which explains diatom abundance pattern.



(a)

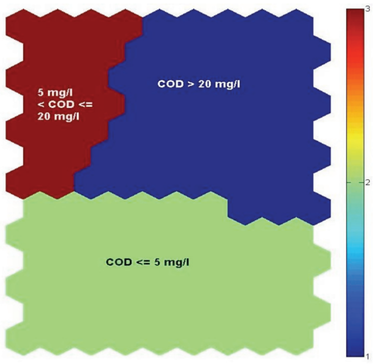


(b)

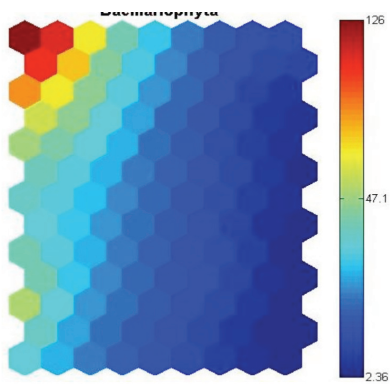


(c)

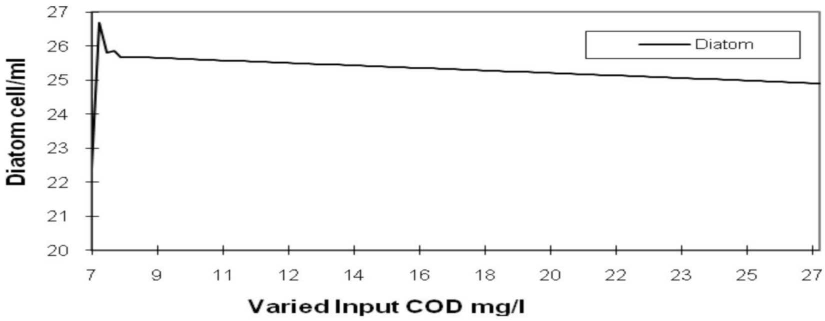
FIGURE 4. Cluster map of diatom abundance over NO_3-N (a) diatom component plane, (b) using SOM and sensitivity curves of diatom abundance over varied input range of NO_3-N using RANN and (c) of Putrajaya Lake



(a)



(b)



(c)

FIGURE 5. Cluster map of diatom abundance of COD (a) diatom component map, (b) using SOM and sensitivity curves of diatom abundance over varied range of COD using RANN and (c) for Putrajaya Lake

Conclusion

The study has demonstrated that complex limnological time series data can be processed by ANN model for tropical lakes in order to provide prediction of algal succession and growth by means of supervised RANN, clusters to unravel ecological relationship with algae abundances and water quality parameters by unsupervised ANN-SOM. In conclusion there is a potential in ANN approach to be used as a modeling tool to estimate major parameter of eutrophication due to its inherit property of being able to be trained for complex and non-linear system.

ACKNOWLEDGEMENT

Data used for this study are courtesy of Mr. Akashah Hj. Majizat, Environmental, Lake and Wetland Division, City Planning Department, Perbadanan Putrajaya, Putrajaya, Malaysia.

REFERENCES

APHA. 1995. *Standard Methods for the Examination of Water and Waste Water.* 19th Ed. Washington, D.C: American Public Health Association.

Bowden, G.J., Dandy, G.C. & Maier, H.R. 2006. An evaluation of methods for the selection of inputs for an artificial neural network based river model. *Ecological Informatics* Recknagel, F. (ed.) 2nd ed. Berlin: Springer-Verlag. pp. 275-292.

Carrillo, P., Reche, I., Sánchez-Castillo, P. & Cruz-Pizarro, L. 1995. Direct and indirect effects of grazing on the Phytoplankton seasonal succession in an oligotrophic lake. *Journal of Plankton Research* 17: 1363-1379.

Chon, T.S., Park, Y.S., Moon, K.H. & Cha, E.Y. 1996. Patternizing communities by using an artificial neural network. *Ecological Modelling* 90: 67-78.

Colasanti, R.L. 1991. Discussions of the possible use of neural network algorithms in ecologicalmodelling. *Binary* 3: 13-15.

Cybenko, G. 1989. Approximations by superpositions of sigmoidal functions. *Mathematics of Control Signals and Systems* 2: 303-314.

Darley, W.M. 1982. *Alga Biology: A Physiological Approach*. Oxford London: Blackwell Scientific Publication.

Edwards, M. & Morse, D.R. 1995. The potential for computer-aided identification in biodiversity research. *Trends in Ecology and Evoluation* 10: 153-158.

Elfithri, R., Toriman, M.E., Mokhtar, M.B. & Juahir, H. 2011. Perspectives and initiatives on integrated river basin management in Malaysia: A review. *The Social Sciences* 6(2): 169-176.

Foody, G.M. 1999. Application of the self-organizing feature map neural network in community data analysis. *Ecological Modelling* 120: 97-107.

Gasim, M.B., Toriman, M.E., Rahim, S.A., Islam, M.S., Chek, T.C. & Juahir, H. 2006. Hydrology and water quality assessment of tasik chini's feeder rivers, Pahang Malaysia *Geografia* 3: 1-16.

Gray, N.F. 2010. *Water Technology, An Introduction For Environmental Scientists and Engineers* (3rd ed.). Oxford: Elseiver: 747.

Gurbuz, H., Kivrak, E., Soyupak, S. & Yerli, S.V. 2003. Predicting dominant phytoplankton quantities in reservoir by using neural networks. *Hydrobiologia* 504: 133-141.

Jeong, K.S., Joo, G.J., Kim, H.W., Ha, K. & Recknagel, F. 2001. Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificialneural network. *Ecological Modelling* 146: 115-129.

Kingston, J.C. 1982. Association and distribution of common diatoms in surface samples from Northern Minnesota peatlands. *Nova Hedwigia* 73: 333-346.

Kohonen, T. 1989. *Self-Organization and Associative Memory* (3rd ed.) Berlin: Springer-Verlag.

Kutty, A.A., Ismail, A. & Fong, C.S. 2001. A preliminary Study of Phytoplankton at Lake Chini Pahang, *Pakistan Journal of Biological Sciences* 4(3): 309-313.

Leclercq L. 1988. Utilization de trios indices, chimique, diatomique et biocénotique, pour l'évaluation de la qualité de l'eau de la Joncquiere, riviére calcaire polluée par le village de Doische (Belgique, Prov. Namur). *Mém. Soc. Roy. Bot. Belg*. 10: 26-34.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulagnier, S. 1996. Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling* 90: 39-52.

Matlab 2006. The Math Works Inc. Matlab (Version 6.5.1).

Maier, H.R., Dandy, G.C. & Burch, M.D. 1998. Use of artificial neural networks for modeling cyanobacteria Anabaena spp. in the River Murray, South Australia. *Ecological Modelling* 105: 257-272.

Maindonald, J.H. & Braun, W.J. 2007. *Data Analysis and Graphics Using R . An Example-Based Approach*. 2nd ed. Cambridge: Cambridge University Press.

Maznah, W. 2010. Perspectives on the use of algae as biological indicators for monitoring and protecting aquatic environments, with special reference to Malaysian freshwater ecosystems. *Tropical Life Sciences Research* 21(2): 51-67.

Maznah, W. & Mansor, M. 1999. Benthic diatoms in the Pinang River (Malaysia) and its tributaries with emphasis on species diversity and water quality. *International Journal on Algae* 1(4): 103-118.

Najah, A., Elshafie, A., Karim, O.A. & Jaffar, O. 2009. Prediction of Johor river water quality parameters using artificial neural networks. *European Journal of Scientific Research* 28(3): 422-435.

Nather Khan, I.S.A. 1985. Studies on the water quality and periphyton community in Linggi River Basin, Malaysia. PhD Diss., University of Malaya (unpublished)

Nather Khan, I.S.A. 1990. Assessment of water pollution using diatom community structure and species distribution: A case study in a tropical river basin. *Internationae Revue der gesamten Hydrobiologie and Hydrographie* 75(3): 317-338.

Nather Khan, I.S.A. 1991. Effect of urban and industrial wastes on species diversity of the diatom community in a tropical river, Malaysia. *Hydrobiologia* 224: 175-184.

Patrick, R. 1994. What are the requirements for an effective biomonitor? In: *Biological Monitoring of Aquatic Systems*. Loeb, Stanford L and Anne Spacie, eds. pp. 23-29.

Pineda, F. 1987. Generalization of backpropagation to recurrent neural networks. *Physical Review Letters* 19(59): 2229-2232.

Principe, J.C., Euliano, N.R. & Lefebvre, W.C. 1999. *Neural and Adaptive Systems: Fundamentals Through Simulations*. N.Y: John Wiley & Sons.

Putrajaya Corporation. 1998. *Putrajaya Lake Management Guide*. Putrajaya Corporation, Selangor Darul Ehsan.

Recknagel, F., Welk, A., Kim, B. & Takamura, N. 2006. Artificial Neural Network Approach to Unravel and Forecast Algal Population Dynamics of Two Lakes Different in Morphometry and Eutrophication. In: *Recknagel, F. Ecological Informatics* (2^{nd} ed.) NY: Springer-Verlag.

Recknagel, F., Kim, B. & Welk, A. 2005. Unravelling and prediction of ecosystem behaviours of Lake Soyang (South Korea) in response to changing seasons and management by means of artificial neural networks. V*erh. Internat. Verein. Limnol* 29(3): 1497-1502.

Recknagel, F. 2001. Applications of machine learning to ecological modelling. *Ecological Modelling* 146: 303-310.

Reynolds, C.S. 1984. *The Ecology of Freshwater Phytoplankton*. Cambridge: Cambridge University Press.

Round, F.E. 1984. *The Ecology of Algae*. Cambridge : Cambridge University Press. pp 653.

Salleh, A. 1996. *Panduan Mengenali Alga Air Tawar*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Sawyer, C.N., McCarty, P.L. & Parkin, G.F. 1994. *Chemistry for Fnvironmental Engineering*. New York: McGraw-Hill.

Schultz, R., Wieland & Lutze, G. 2000. Neural networks in argroecological modelling—stylish application or helpful tool?. *Computers and Electronics in Agriculture* 29: 73-97.

Sommer, U. 1989. Nutrient status and nutrient competition of phytoplankton in a shallow, hypertrophic lake. *Limnology and Oceanography* 34(7): 1162-1173.

Talib, A., Abu Hasan, Y. & Abdul Rahman, N.N. 2009. Predicting biochemical oxygen demand as indicator of river pollution using artificial neural networks. *18th World IMACS / MODSIM Congress*, Cairns, Australia 13-17 July 2009

Thomann, R.V. & Mueller, J.A. 1987. *Principles of Surface Water Quality Modeling and Control* New York: Harrper and Row.

Vanni, M.J. & Temte, J. 1990. Seasonal patterns of grazing and nutrient limitation of phytoplankton in a eutrophic lake. *Limnology and Oceanography* 35: 697-709.

WHO. 1987. WHO, *UNEP/WHO/UNESCO/WMO Project on Global Environmental Monitoring*. GEM Water Operational Guide.

Werner, D. 1977. *The Biology of Diatoms*. Bot. Monogr., V. 13. Berkeley and New York: University of California Press,

Zhang, G.P., Patuwo, B.E. & Hu, M.Y. 1998. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting* 14: 35-62.

M. Sorayya & S. Aishah
Institute of Biological Sciences (ISB)
University of Malaya
50603 Kuala Lumpur
Malaysia

B. Mohd. Sapiyan*
Faculty of Science Computer and Information Technology
University of Malaya
50603 Kuala Lumpur
Malaysia

*Corresponding author; email: pian@um.edu.my

Received: 29 June 2011
Accepted: 31 January 2012