Applying artificial neural network theory to exploring diatom abundance at tropical Putrajaya Lake, Malaysia

Sorayya Malek, Aishah Salleh, Pozi Milow, Mohd Sapiyan Baba & S.A. Sharifah

Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia
Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia


To cite this article: Sorayya Malek, Aishah Salleh, Pozi Milow, Mohd Sapiyan Baba & S.A. Sharifah (2012): Applying artificial neural network theory to exploring diatom abundance at tropical Putrajaya Lake, Malaysia, Journal of Freshwater Ecology, 27:2, 211-227

To link to this article: http://dx.doi.org/10.1080/02705060.2011.635883

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Applying artificial neural network theory to exploring diatom abundance at tropical Putrajaya Lake, Malaysia

Sorayya Maleka*, Aishah Salleha, Pozi Milowa, Mohd Sapiyan Babab and S.A. Sharifahc

aInstitute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia; bFaculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia; cFaculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia

(Received 12 April 2011; final version received 14 October 2011)

This article explores the relationship between diatom abundance and water quality variables in tropical Putrajaya Lake based on limnological data collected from 2001 to 2006, using supervised and unsupervised artificial neural networks (ANNs). Recurrent artificial neural network (RANN) was used for the supervised ANNs and Kohonen Self Organizing Feature Maps (SOMs) for the unsupervised ANNs. The RANN was developed for the prediction of diatom abundance using variables selected by sensitivity analysis (water temperature, pH, dissolved oxygen, and turbidity). The RANN model performance was measured using root mean squared error (19.0 cell/mL) and the $r$-value (0.7). SOM was used in this study for classification and clustering of diatom abundance in relation to selected water quality variables and was validated using a sensitivity curve of diatom abundance over the selected variable range generated from RANN. SOM has been employed in this study for pattern discovery of diatom abundance at Putrajaya Lake. The extracted patterns of diatom abundance in terms of propositional IF...else rules were tested and yielded an accuracy rate of 87%.

Keywords: recurrent artificial neural network; self-organizing maps; pattern discovery; tropical lakes; sensitivity analysis

Introduction

Diatoms can be important indicators of environmental conditions in lakes because they respond directly and sensitively to many physical, chemical, and biological changes in aquatic ecosystems, such as water temperature (Squires et al. 1979; Descy and Mouvet 1984), nutrient concentrations (Pringle and Bowers 1984; Pan and Stevenson 1996), and herbivory (Steinman et al. 1987; McCormick and Cairns 1994). Numerous studies have used diatoms to indicate or evaluate water quality of lakes and reservoirs. These include studies on the succession of diatom assemblages in Lake Washington (Stockner and Benson 1967), diatoms as indicators of the rate of lake acidification (Dickman et al. 1984), temperature effects on diatoms (Weckstrom et al. 1997), and on the relationships between diatoms and the environment in Spanish reservoirs (Negro and De Hoyos 2005). However, there is still a scarcity of
reports on trends in diatom abundance in tropical, human-made lakes, and reservoirs. The few examples that are closely related to this include a study on physico-chemical conditions of a reservoir in Kenya by Kotut et al. (1998) and a study on an oligotrophic reservoir in Brazil by Lopes et al. (2005). Human-made lakes and reservoirs are built to supply water for anthropogenic uses, for aesthetic and recreational purposes, to act as ecological buffers and to provide refuge for plants and animals. Thus, monitoring of water quality by managers is necessary to anticipate and mitigate any undesirable changes to the lakes and reservoirs.

Artificial neural networks (ANNs) are computational networks that attempt to simulate the networks of nerve cells (neurons) in an animal central nervous system (Graupe 2007). The use of ANN models to extract relationships between organisms and their environments has several advantages over conventional statistical approaches. Normally distributed data, which are usually a requirement in statistical studies, are not required in neural networks (Burke and Ignizio 1992). Furthermore, according to Zhang et al. (2001), ANNs are able to learn and respond to complex functional relationships within datasets, even when the underlying relationships are unknown or difficult to describe. They can generalize and infer an occurrence even if the sample input data are noisy. ANNs are efficient at estimating unknown non-linear, multivariate functions (Hornik 1991). This feature is important as the number of possible non-linear patterns in real world problems can be very large. Maier and Dandy (1998) have demonstrated that neural networks can be used when limited datasets are available and can be used with a wide diversity of datasets in medical, economics, finance, geography, and biology fields.

The uses of ANN models to study eutrophication and management of temperate lakes have been discussed by Recknagel et al. (1997), Maier and Dandy (1998), Wilson and Recknagel (2001), and Recknagel et al. (2006). However, there are uncertainties in applying such models to tropical lakes compared to temperate lakes due to differences in terms of water temperature, seasonality, nutrients, and phytoplankton productivity. Therefore, this study is an attempt to use ANN to predict diatom abundance in a tropical human-made lake. Diatoms have been reported to be one of the dominant algae found at Putrajaya Lake (Putrajaya Corporation 2006). The study involves the use of supervised (recurrent artificial neural network or RANN; Pineda 1987) and unsupervised ANN (self-organizing feature maps or SOMs; Kohonen 1982). SOM has been used in ecological modeling to find similarity between datasets and is an excellent tool in the visualization of multidimensional data (Vesanto 1993; Chon et al. 1996).

Materials and methods

Putrajaya Lake and its limnological data

The 400 ha of the human-made, freshwater Putrajaya Lake was created by inundating the valleys of Sungai Chua and Sungai Bisa. The lake was constructed to provide a landscape feature and varied recreational activities for the city population as well as wildlife habitats (Shutes 2001). Putrajaya Lake is warm polymictic, oligotrophic to mesotrophic, and is located in the southern portion of the densely inhabited Klang Valley, Malaysia. Major inflows from surrounding areas contain some pollutants derived from the use of agrochemicals and fertilizers, land clearing, and soil disturbance. Nutrient loading to the lake mainly comes from
non-point sources. Figure 1 shows Putrajaya Lake, and the principal features of Putrajaya Lake are listed in Table 1.

Water sampling for this study was carried out in the morning twice a month at 23 fixed sub-stations of 13 major sampling stations during the years 2001–2006. The water samples were collected near shore at a depth of 0.5 m. The sub-stations were divided arbitrarily into two sets (datasets A and B). Dataset A was used for training and dataset B for testing of ANN models. Water sampling for water quality and diatom abundance analyses was carried out according to American Public Health Association (APHA; Eaton et al. 1995) and World Health Organization (WHO 1987). Water samples for diatom identification were collected using a plankton net with mesh size of about 30 μm. Smaller mesh size plankton net was not used because of problems of clogging and reduced water flow during the transfer of water to vials (Bellinger and Sigee 2010). Each water sample for diatom analysis was gathered from several scoops of the site water to reduce the chance of missing less abundant diatoms.

Diatoms were preserved by adding several drops of 4% formaldehyde to the water samples that were subsequently kept in 50 mL vials. Identification of diatoms was carried out using an ordinary light microscope. Identification of diatom genera was based on the literature, such as Wehr and Sheath (2003) and Bellinger and Sigee (2010). Diatom abundance was calculated using the sedimentation technique described by Evans (1972). Diatoms that were found in Putrajaya Lake belonged to the following genera: *Amphora* Ehrenb., *Cyclotella* Kütz., *Cymbella* C. A. Agardh, *Gomphonema* Ehrenberg, *Meridion* C. A. Agardh, *Navicula* Bory, *Nitzschia* Hass., *Pinnularia* Ehrenberg, *Tabellaria* Ehrenb. ex Kützing, *Synedra* Ehrenberg, and *Surirella* Turp. Other information on the water characteristics of Putrajaya Lake are shown in Table 2.

**Data processing and input selection**

ANN is a non-parametric method, therefore assumptions of independent datasets and normally distributed data are not taken into consideration in this study. However, normalization of the raw data was carried out prior to the ANN model constructions to ensure that all values of the variables were within the same range. Input data used in this study was normalized to the range 0 to 1. Presenting a large number of inputs to ANN increases the network size, which leads to an increased amount of data used to estimate the connection weight and possible reduction of processing speed (Lachtermacher and Fuller 1994). Thus, the preferred method of selection of inputs should involve a combination of prior knowledge and analytical approaches (Maier and Dandy 1998, 2000).

Sensitivity analysis, an analytical approach, was used to select model inputs for RANN. A sensitivity analysis measures how much a small change in one of the independent variables affects the output in the training dataset (Principe et al. 1999). Inputs that have large sensitivities have more importance in the mapping and, therefore, are the ones that should be kept. The inputs with small sensitivities were discarded. Input variables that were included in the sensitivity analysis are shown in Table 2. Five different RANN models using all the variables were developed and trained using dataset A. A backward elimination method was used to eliminate less sensitive variables and the network was then retrained with the reduced number
of variables. Figure 2 illustrates the sensitivity analysis results from the best performing model. The model that performed best was retained and the effect of reducing the number of variables on model performance was investigated. This procedure was repeated until the discarding of variables did not improve model...
Table 1. Principal features of Putrajaya Lake (Putrajaya Corporation 1998).

<table>
<thead>
<tr>
<th>Catchment area (km²)</th>
<th>Water level (m)</th>
<th>Surface area (ha)</th>
<th>Storage volume (10⁶ m³)</th>
<th>Average depth (m)</th>
<th>Average catchment inflow (10⁶ L d⁻¹)</th>
<th>Average retention time (d)</th>
<th>Circulation type</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.9</td>
<td>21</td>
<td>400</td>
<td>26.5</td>
<td>6.6</td>
<td>200</td>
<td>132</td>
<td>Warm polymictic (non-stratified shallow lake)</td>
</tr>
</tbody>
</table>

Table 2. Mean limnological properties of Putrajaya Lake from 23 sampling stations collected twice a month from 2001 until 2006.

<table>
<thead>
<tr>
<th>Water quality variables</th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water temperature (°C)</td>
<td>30.35</td>
<td>26.76</td>
<td>34.54</td>
</tr>
<tr>
<td>pH</td>
<td>7.25</td>
<td>6.00</td>
<td>8.94</td>
</tr>
<tr>
<td>Secchi depth (m)</td>
<td>1.15</td>
<td>0.01</td>
<td>2.40</td>
</tr>
<tr>
<td>Turbidity (NTU)</td>
<td>14.64</td>
<td>0.00</td>
<td>416.0</td>
</tr>
<tr>
<td>DO (mg/L)</td>
<td>7.11</td>
<td>3.81</td>
<td>10.80</td>
</tr>
<tr>
<td>Conductivity (μS/m)</td>
<td>94.9</td>
<td>35.0</td>
<td>199.0</td>
</tr>
<tr>
<td>Salinity (ppt)</td>
<td>0.04</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>NH₃–N (mg/L)</td>
<td>0.11</td>
<td>0.00</td>
<td>2.70</td>
</tr>
<tr>
<td>NO₃–N (mg/L)</td>
<td>1.14</td>
<td>0.00</td>
<td>4.88</td>
</tr>
<tr>
<td>BOD (mg/L)</td>
<td>1.59</td>
<td>1.00</td>
<td>20.00</td>
</tr>
<tr>
<td>COD (mg/L)</td>
<td>16.82</td>
<td>1.00</td>
<td>79.00</td>
</tr>
<tr>
<td>Total suspended solids (mg/L)</td>
<td>6.50</td>
<td>0.00</td>
<td>326.00</td>
</tr>
<tr>
<td>Diatom abundance (cell/mL)</td>
<td>24</td>
<td>1</td>
<td>418</td>
</tr>
</tbody>
</table>

Figure 2. Sensitivity analysis results from RANN showing the relative sensitivity of diatom abundance to the factors.
performance any further. This process resulted in the inclusion of water temperature, pH, turbidity, and dissolved oxygen (DO). RANN was also used to generate a sensitivity curve of diatom abundance over a range of selected input variables. This was used to validate results of classification and clustering by SOM.

**Recurrent artificial neural network model development**

The RANN structure was built using Neurosolution 5.02 (NeuroDimension 2005). The geometry of the final model was $4 \times 4 - 1$ (number of input - number of hidden nodes – number of outputs). The learning algorithm used was back-propagation through time (BPTT; Rumelhart et al. 1986). The BPTT algorithm is based on changing the network from a feedback system to feed-forward system by folding the network over time. The network uses a momentum-learning algorithm to determine weights in the network. The equation to update weights for momentum learning was:

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

where $\alpha$ is the momentum constant and was set at 0.7. Weights were proportionally changed to the amount they were updated in the last iteration. A hyperbolic tangent function was used as an activation function at the hidden layer and at the output layer.

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

where $\alpha$, a slope parameter, was set to 0.7 and an iteration size of 100 was used. Network performance was measured in terms of root mean squared error (RMSE) and the $r$-value of the observed versus predicted diatom abundance.

**Kohonen SOM development**

Kohonen SOM, an unsupervised ANN, was used to classify and cluster diatom abundance with respect to key water quality variables selected from the sensitivity analysis. The clustered map is validated against the sensitivity curve generated from RANN. The classification criteria used for clustering were based on the Interim National Water Quality Standards for Malaysia (Department of Environment; DOE 2000) and statistical distribution of data at Putrajaya Lake. Table 3 lists the clustering criteria used in this study. SOM was also used for pattern discovery of diatom abundance in the lake. SOM produces a two-dimensional graphical image

Table 3. Criteria for classification and clustering of limnological data of Putrajaya Lake (DOE 2000).

<table>
<thead>
<tr>
<th>Measured parameters</th>
<th>Classification criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water temperature ($^\circ$C)</td>
<td>$&lt; 30.5$ (low); $30.5 \geq$ and $&lt; 31.5$ (medium); $\geq 31.5$ (high)</td>
</tr>
<tr>
<td>pH</td>
<td>$&lt; 7$ (low); $7 \geq$ and $&lt; 8.0$ (medium); $\geq 8.0$ (high)</td>
</tr>
<tr>
<td>DO (mg/L)</td>
<td>$\leq 7.0$ (low); $7.0 &lt;$ and $\leq 8.0$ (medium); $&gt; 8.0$ (high)</td>
</tr>
<tr>
<td>Turbidity (NTU)</td>
<td>$\leq 20$ (low); $20 &lt;$ and $\geq 50$ (medium); $&gt; 50$ (high)</td>
</tr>
</tbody>
</table>
Figure 3. Forecast of diatom abundance at Putrajaya Lake using selected water quality variables.

Figure 4. Clusters of diatom abundance and component plane of diatom abundance by SOM (top), sensitivity curve of diatom abundance over the water temperature range by RANN (bottom) of Putrajaya Lake.
that describes feature-based relationships of data within each cluster which assists in rules extraction.

The SOM procedure was adopted from Kohonen (1995) and developed using the SOM toolbox in MATLAB (Matlab 2006). Euclidean distance between the inputs were calculated and visualized as a distance matrix (U-matrix) and a partition map (K-means). Data clusters were generated using the Davies–Bouldin index (Davies and Bouldin 1979) of the K-means algorithm. Diatom abundance patterns were extracted by mapping the clusters generated from the clustering map with the input variables’ component planes. The rules extracted were based on the propositional IF ... THEN ... ELSE type of rules. Prior to the rules extraction, each variable range was dichotomized. Diatom abundance was dichotomized to $<45$ and $\geq 45$ cell/mL; pH was dichotomized to $<7.32$ and $\geq 7.32$; DO was dichotomized to $<6.79$ and

Figure 5. Clusters of diatom abundance and component plane of diatom abundance by SOM (top), sensitivity curve of diatom abundance over the pH range by RANN (bottom) of Putrajaya Lake.
Results

Figure 3 shows the forecasting results of diatom abundance based on dataset B. The predicted timing and abundance of diatoms corresponded reasonably well with the measured data ($r = 0.7$, $n = 180$) and RMSE value (19.0 cell/mL). The SOM results in Figures 4–7 were used to study the relationship between diatom abundance and selected water quality variables (water temperature, pH, DO, and turbidity). Diatom abundance is indicated by the shade intensity in the ‘honeycomb panel’ of the SOM component plane. Figure 4 illustrates the relationship between diatom

\[ \text{Diatom abundance} = \begin{cases} \text{cell/mL} \\ 7.0 \text{ mg/L} < \text{DO} \\ \text{Dissolved Oxygen} \leq 8.0 \text{ mg/L} \\ \text{Dissolved Oxygen} \leq 7.0 \text{ mg/L} \end{cases} \]

Figure 6. Clusters of diatom abundance and component plane of diatom abundance by SOM (top), sensitivity curve of diatom abundance over the DO range by RANN (bottom) of Putrajaya Lake.
abundance and water temperature. It shows that diatom abundance reached its highest when water temperature was below 30.3°C. Figure 5 shows that high diatom abundance coincided with pH values above 8. Figure 6 shows high diatom abundance coincided with DO concentration of 7 mg/L and above. Figure 7 shows high diatom abundance coincided with turbidity between 20 and 50 NTU.

Pattern discovery of diatom abundance using SOM, carried out by mapping clusters (Figure 9) with component planes in Figure 8, produced eight rules. The labels in Figure 9 refer to the extracted rule for diatom pattern abundance for each cluster shown in Table 4. Both rules 8 and 7 (Table 4) were merged together as turbidity was the only criteria that distinguished the two rules. The accuracy rates of each rule representing diatom pattern abundance are shown in Figure 10. The strongest rule is rule 8 with an accuracy rate of 98% and the weakest is rule 1.
with an accuracy rate of 14%. The combined extracted rules produced accuracy rates of 87% and 82% when tested on datasets A and B, respectively.

**Discussion**

Sensitivity analysis of input variables generated using RANN indicates that there are differences in sensitivity of diatom abundance to the various water quality variables. The more sensitive variables purportedly have a more immediate effect on diatom abundance. Our findings indicate that water temperature, pH, DO, and turbidity had the most immediate effects on diatom abundance. Secchi depth, conductivity, nitrate, ammonia, salinity, biochemical oxygen demand (BOD), and chemical oxygen demand (COD) had negligible effect on diatom abundance. The results suggest an optimum of four input variables (water temperature, DO, pH, and turbidity) for the neural network model for Putrajaya Lake. The RMSE value was 19.0 cell/mL and the $r$-value was 0.7. A higher RMSE value (256–578 cell/mL) was obtained in a similar study on cyanobacteria by Bowden et al. (2006).

The relationship between water temperature and diatom abundance at Putrajaya Lake is apparent in Table 4. The criterion that distinguishes rule 1 (high abundance)
and rule 2 (low abundance) is water temperature. Rules 1 and 2 explain high diatom abundance at water temperature lower than 30.3°C and, conversely, low diatom abundance at water temperature higher than 30.3°C. This result also suggests that pH, DO, and turbidity did not influence diatom abundance as much as water temperature. The yearly average water temperatures of Putrajaya Lake for the years 2001–2006 were 29.5°C, 30.3°C, 30.10°C, 30.3°C, 30.6°C, and 30.8°C, respectively. However, Recknagel et al. (2006) reported high diatom abundance below 16°C for a temperate lake in Japan and Reynolds (1984) reported that diatoms are dominant at low temperatures in temperate lakes. He further added that low water temperature favors growth of pinnate-type diatoms (e.g. Asterionella, Diatoma, Fragilaria, and Tabellaria) due to their efficient photosynthetic ability, cell size and surface area, and the inactivity of filter feeders. Pennate forms are also found in Putrajaya Lake. Shen (2002) reported that the most favorable temperature for phytoplankton growth was 30°C.

The coincidence of high abundance of diatoms with high pH values (8–9) in Putrajaya Lake has been observed at the temperate lake Kasumigaura as well, where the diatom abundance for Cyclotella and Asterionella concided with high water pH of 8–9 (Recknagel et al. 2006). This study shows that the relationship between diatom abundance and pH conforms to previous findings. High diatom abundance is

Figure 9. Cluster map of diatom abundance at Putrajaya Lake. The labels on the cluster map correspond to the extracted rules in Table 4.
generally related to relatively high conductivity and pH (Eloranta 1995). Tailing (1971) reported from a series of experiments that photosynthesis in diatoms appears to be tolerant of high pH and low CO₂ concentration. The removal of CO₂ from water by phytoplankton increases the pH as bicarbonate and carbonate levels are depleted to replenish the removed CO₂.

Rules 3–8 (Table 4 and Figure 4) show that low diatom abundance coincides with low pH (below 7). Rules 3–8 also show that turbidity, water temperature, and DO

<table>
<thead>
<tr>
<th>Rule number</th>
<th>Cluster number</th>
<th>Extracted rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1)</td>
<td>IF water temperature ( \leq 30.3 ) and IF pH &gt; 7.32 and IF DO &gt; 6.79 and IF turbidity ( \leq 31 ) Then diatom abundance medium</td>
</tr>
<tr>
<td>2</td>
<td>(2, 3)</td>
<td>IF water temperature &gt; 30.3 and IF pH &gt; 7.32 and IF DO &gt; 6.79 and IF turbidity ( \leq 31 ) Then diatom abundance low</td>
</tr>
<tr>
<td>3</td>
<td>(4, 8)</td>
<td>IF water temperature ( \leq 30.3 ) and IF pH ( \leq 7.32 ) and IF DO ( \leq 6.79 ) and IF turbidity ( \leq 31 ) Then diatom abundance low</td>
</tr>
<tr>
<td>4</td>
<td>(5)</td>
<td>IF water temperature ( \leq 30.3 ) and IF pH ( \leq 7.32 ) and IF DO &gt; 6.79 and IF turbidity ( \leq 31 ) Then diatom abundance low</td>
</tr>
<tr>
<td>5</td>
<td>(6)</td>
<td>IF water temperature &gt; 30.3 and IF pH ( \leq 7.32 ) and IF DO &gt; 6.79 and IF turbidity ( \leq 31 ) Then diatom abundance low</td>
</tr>
<tr>
<td>6</td>
<td>(7)</td>
<td>IF water temperature ( \leq 30.3 ) and IF pH ( \leq 7.32 ) and IF DO ( \leq 6.79 ) and IF turbidity &gt; 31 Then diatom abundance low</td>
</tr>
<tr>
<td>7 and 8</td>
<td>(9)</td>
<td>IF water temperature &gt; 30.3 and IF pH ( \leq 7.32 ) and IF DO ( \leq 6.79 ) and IF turbidity &gt; 31</td>
</tr>
<tr>
<td>7 and 8</td>
<td>(9)</td>
<td>IF water temperature &gt; 30.3 and IF pH ( \leq 7.32 ) and IF DO ( \leq 6.79 ) and IF turbidity &gt; 31</td>
</tr>
</tbody>
</table>
have negligible effect on diatom abundance at low pH (7 and below). The rules complement earlier findings by Battarbee et al. (1986), who stated that diatoms are good indicators of lake water pH. Therefore, fluctuations in lake water pH could be used by lake managers to predict diatom abundance and anticipate the possibility of blooms in Putrajaya Lake.

The rule strength in terms of accuracy rate (Figure 10) shows that rule 1 is the weakest with an accuracy rate of 14%. Other rules exhibited strengths of 83–97%. Rule 1, although weak, described the optimum water temperature for high diatom abundance. The rest of the rules generally describe the water condition that prevails in Putrajaya Lake. It conforms to the characteristics of an oligotrophic lake. Based on this study, the optimum water condition for high diatom abundance at Putrajaya Lake occurs at water temperature lower than 30.3°C, pH 8, or above, DO concentration more than 8 mg/L, and turbidity between 20 and 50 NTU. This is inferred from results in Figures 4–7. The optimum condition for high diatom growth has been discussed by Tailoring (1971) and Reynolds (1984). Diatoms are able to grow under turbid conditions as they are adapted to low light and equipped with effective photosynthetic capabilities under such conditions (Reynolds 1984). The main source contributing to DO in lakes is photosynthesis carried out by algae. High photosynthesis rates more oxygen. This conforms to findings from sensitivity curve graphs that the amount of DO increases with diatoms abundance as result of photosynthesis carried out by the diatoms.

Conclusion
This study shows that trends in diatom abundance in Putrajaya Lake are distinguished by water temperature, pH, DO, and turbidity with diatom abundance being most sensitive to water temperature and pH. This study demonstrates potential uses of ANN to detect trends and construct models that describe the abundance of phytoplankton that can be subsequently used to monitor water quality of
tropical lakes. This study also shows that supervised and unsupervised ANN can be
used together to generate models and extract trends in diatom abundance with
respect to their environment. The importance of performing variable selection
prior to modeling in order to achieve optimum forecasting capability cannot be
disregarded.

Acknowledgments
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
Bellinger EG, Sigee DC. 2010. Freshwater algae identification and use as bioindicators.
Chichester (UK): Wiley-Blackwell.
Bowden GJ, Dandy GC, Maier HR. 2006. An evaluation of methods for the selection
Chon TS, Park YS, Moon KH, Cha EY. 1996. Patternizing communities by using an artificial
Descy JP, Mouvet C. 1984. Impact of the Tihange nuclear power plant on the periphyton
Dickmann M, Dixit S, Fortescue J, Barlow B, Terasmae J. 1984. Diatoms as indicators of the
Technology and Environment Malaysia: Annual Reports.
and waste water. 19th ed. Washington (DC): APHA.
Eloranta P. 1995. Type and quality of river waters in central Finland described using diatom
indices. In: Marino D, Montresor M, editors. Proceedings of the 13th International
Diatom Symposium; 1994 Sep 1–7; Maratea, Italy. Bristol (UK): Biopress, p. 271–280.
Evans JH. 1972. A modified sedimentation system for counting algae with an inverted
systems. 2nd ed. Singapore: World Scientific.
Networks. 4:251–257.
Cybernetics. 43:59–69.
cyanoprocaryotes and algae from Turkwel Gorge Reservoir, Kenya. Archives of
In: Hipel KW, MacLeod AI, Panu US, Singh VP, editors. Vol. 3. Stochastic and


