International Journal of the Physical Sciences Vol. 6(22), pp. 5298-5308, 2 October, 2011
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS11.1180
ISSN 1992 - 1950 ©2011 Academic Journals

Full Length Research Paper

An application of different artificial intelligences techniques for water quality prediction

A. Najah[1]*, A. El-Shafie[1], O. A. Karim[1], O. Jaafar[1] and Amr H. El-Shafie[2]

[1]Department of Civil and Structural Engineering, Universiti Kebangsaan Malaysia, UKM, Malaysia.
[2]Faculty of Engineering, University of Garyounis, Banighazi, Lybia.

Accepted 5 September, 2011

Artificial Intelligence (AI) is a new technique with a flexible mathematical structure that is capable of identifying complex non-linear relationships between input and output data when compared with other classical modelling techniques. In this study, different techniques of AI have been investigated in prediction of water quality parameters including: multi-layer perceptron neural networks (MLP-ANN), ensemble neural networks (E-ANN) and support vector machine (SVM). The parameters were investigated in terms of the following: the dissolved oxygen (DO), biochemical oxygen demand (BOD) and chemical oxygen demand (COD). To assess the effect of input parameters on the model, the sensitivity analysis was adopted. To evaluate the performance of the proposed model, three statistical indexes were used, namely; correlation coefficient (CC), mean square error (MSE) and correlation of efficiency (CE). The principle aim of this study is to develop a computationally efficient and robust approach for predicting water quality parameters which could reduce the cost and labour for measuring these parameters. This research concentrates on the Johor river in Johor State, Malaysia where the dynamics of river water quality are significantly altered.

Key words: Artificial intelligence, water quality prediction model.

INTRODUCTION

Water quality impairment is often a trigger for conflict in a watershed, simply because degraded water quality means that desired uses are not possible or not safe (Heathcote, 1998). Malaysia is a developing country that moves towards the vision 2020. Unfortunately the development that had been carried throughout the country gives a bad impact to the environment, especially about water quality. This has become a sensitive issue, which not only affects human health, but also the entire environment. The development not only affects the water of quality, but also the aquatic lives that live in it. Most acceptable ecological and social decisions are difficult to make without careful modelling, prediction and analysis of river water quality for typical development scenarios. Water quality prediction enables a manager to choose an option that satisfies large number of identified conditions. For instance, water quality parameters, such as dissolved solids, electrical conductivity and turbidity in water describe a complex process governed by a considerable number of hydrologic, hydrodynamic and ecological controls that operate at a wide range of spatiotemporal scales. Sources of the admixtures often cannot be clearly identified, and the locally influenced complex mass exchange between the variables may not be known.

Recently, applications of artificial intelligence (AI) in the areas of water engineering, ecological sciences and environmental sciences have been reported since the beginning of the 1990s. In recent years, ANNs have been used intensively for prediction and forecasting in a number of engineering and water-related areas, including water resource study (Liong et al., 1999, 2001; Muttil and Chau, 2006; El-Shafie et al., 2008, 2011; Najah et al., 2009, 2010a, b, c; Razavi et al., 2011a, b), oceanography (Makarynskyy, 2004) and environmental science (Grubert, 2003). The use of data-driven techniques for modeling the quality of both freshwater (Chen and Mynett, 2003) and seawater (Lee et al., 2000, 2003) has met with success in the past decade. Reckhow (1999)

*Corresponding author. E-mail: ali_najah@ymail.com.

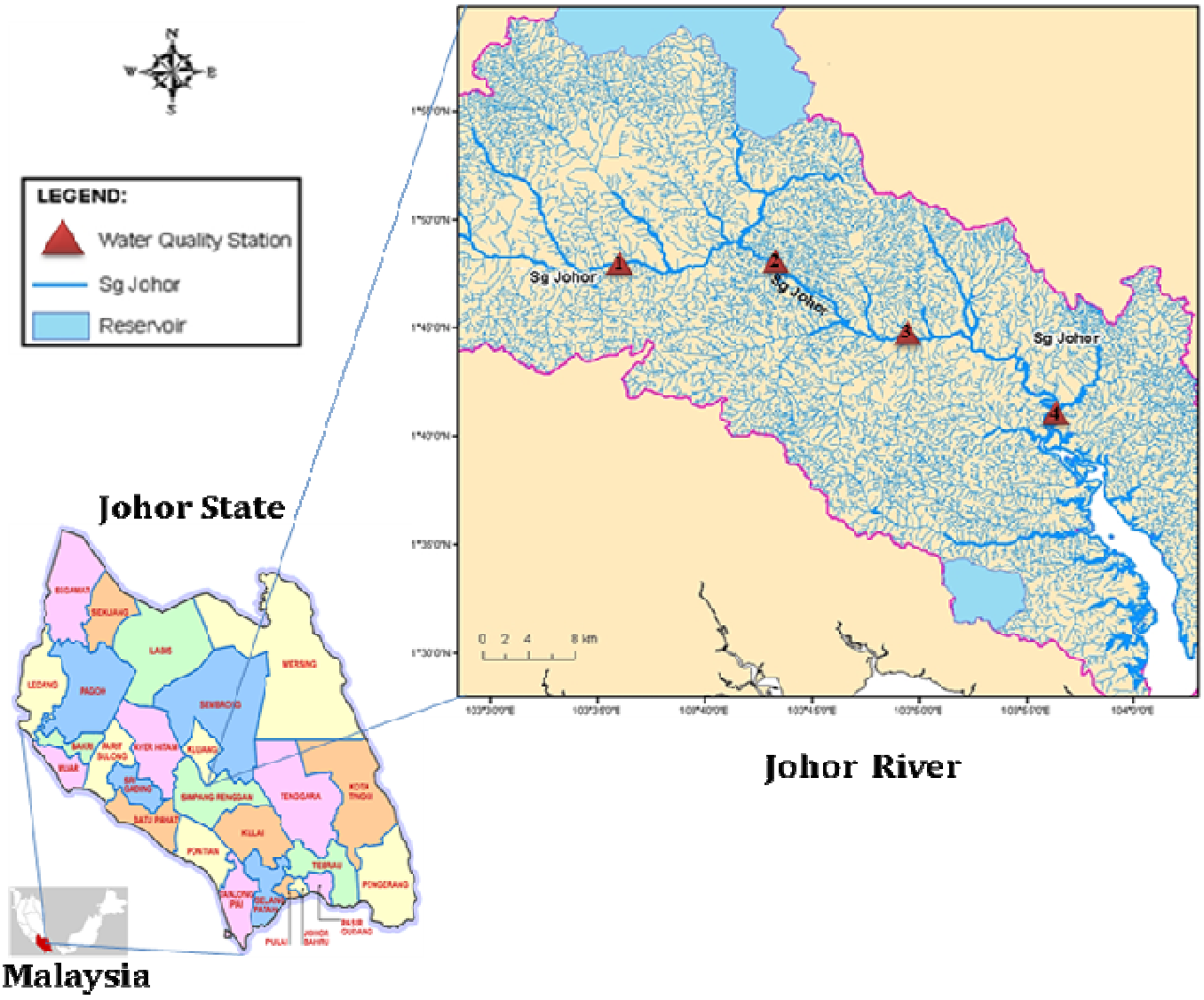

Figure 1. Map showing the geographical setting of the survey area with four field monitoring stations on the main stream.

studied Bayesian probability network models for guiding decision making regarding water quality in the Neuse river in north Carolina. Chau (2006) has reviewed the development and current progress of the integration of artificial intelligence (AI) into water quality modelling. Limited water quality data and the high cost of water quality monitoring often pose serious problems for process-based modelling approaches. AI provides a particularly good option, because they are computationally very fast and require many fewer input parameters and input conditions than deterministic models. AI does, however, require a large pool of representative data for training. Therefore, this paper demonstrates the application of different techniques of AI to model the values of selected river water quality parameters, having the dynamic and complex processes hidden in the monitored data itself. In addition, objective of this study is to investigate whether it is possible to predict the values of water quality parameters measured by a water quality monitoring program; this task is quite important for enabling selective monitoring of water quality parameters.

MATERIALS AND METHODS

Study area and water quality data

Johor is the second largest state in the Malaysia Peninsular, with an area of 18,941 km^2. The Johor river and its tributaries are important sources of water supply, not only for the state of Johor but also for Singapore. The river comprises 122.7 km long drains, covering an area of 2,636 km^2. The station's location map is provided in Figure 1. This station includes four locations along the main stream of the river, which are near to the mouth of the major tributaries and the two largest point sources.

The amount of dissolved oxygen present in a watercourse is one of the most important measures of the water quality. It is also commonly used as indicators of a river's health. The number of life forms that survive begins to decrease as the level of DO drops below 4 mg/L. In extreme cases, when anaerobic condition exists, most high forms of life are either killed or driven off. Eventually conditions like floating sludges, bubbling, odourous gasses and slimy fungal growths will subsist (Masters, 1993).

Most organic materials, such as those from waste water treatment plants, industrial effluents and agricultural run-off are biodegradable. The amount of oxygen used in the metabolism of biodegradable organics is termed biochemical oxygen demand. When organic matter decomposes, microorganisms such as bacteria and fungi feed upon it and eventually it becomes oxidized. Biochemical oxygen demand (BOD) is a measure of the quantity of oxygen used by these microorganisms in the aerobic oxidation of organic matter (Train, 1979).

Chemical oxygen demand (COD) is a water quality parameter to indicate the level of pollution in the water based on chemical characteristics and is a measure of the amount of oxygen required to oxidize the organic matter chemically by a strong oxidant known as dichromate and sulfuric acid. COD is therefore an estimate of the amount of organic and reduced matter present in the water or better known as the amount of oxygen needed to chemically decompose the organic matter in the water.

Artificial intelligence techniques

Multilayer perceptron neural network model MLP-ANN

Of the many types of ANNs, the most widely used are the feed forward neural network, multi-layer perceptron (MLP) or back-propagation network. The MLP is organized as layers of computing elements, known as neurons, which are connected between layers via weights. Apart from an input layer receiving inputs from the environment and an output layer generating the network's response, one or more intermediate hidden layers also exist. For brevity, we refrain from discussing the details of the neural network methodology and instead refer the reader to the papers of Lek et al. (1996) and Olden and Jackson (2001) for more comprehensive treatments. Generally, forecasting models can be divided into statistical and physically based approaches. Statistical approaches determine relationships between historical data sets, whereas physically based approaches model the underlying processes directly. MLP networks are closely related to statistical models and are the type of ANN most suited to forecasting applications (Rumelhart et al., 1986). When using ANNs for forecasting, the modeling philosophy employed is similar to that used in traditional statistical approaches. In both cases, the unknown model parameters (that is, the connection weights in the case of ANNs) are adjusted to obtain the best match between a historical set of model inputs and the corresponding outputs.

These neural networks are commonly used in ecological studies, because they are believed to be universal approximates of any continuous function (Hornik and White, 1989). A neural network consists of at least three or more layers, which comprise an input layer, an output layer and a number of hidden layers as shown in Figure 2. Each neuron in one layer is connected to the neurons in the next layer, but there are no connections between the units of the same layer (Kasabov, 1996). The number of neurons in each layer may vary depending on the problem. The weighted sum of the input components is calculated (Freeman and Skapura, 1991) as:

$$Net_j = \sum_{i=1}^{n} W_{ij} + \theta_j \quad (1)$$

where Net_j is the weighted sum of the jth neuron for the input data received from the preceding layer with n neurons, w_{ij} is the weight between the jth neuron and the ith neuron in the preceding layer and θj is the bias term of the jth neuron. The output of the jth neuron out_j is calculated with a sigmoid function as follows:

$$out_j = f(Net_j) = \frac{1}{1 + e^{-Net_j}} \quad (2)$$

The network is trained by adjusting the weights. The training process is carried out with a large number of training sets and training cycles (epochs). The main goal of the learning procedure is to find the optimal set of weights, which can ideally produce the correct output for the relative input. The output of the network is compared with the desired response to determine the error. The performance of the MLP is measured in terms of a desired signal and the criterion for convergence. For one sample, it is determined by the sum square error (SSE), which is expressed as:

$$SSE = \sum_{i=1}^{m} (T_i - out_i)^2 \quad (3)$$

where T i and out i are the desired (target) and the output of the neural network, respectively, for the ith output neuron and m is the number of neurons in the output layer.

Assembly neural network model E-ANN

It is proposed to find a technique based on ensemble neural networks (Chiewchanwattana et al., 2002; Drucker et al., 1994; Cheni et al., 2005) that by using over-fitted neural networks leads to generalization. In order to achieve this goal, we use a sequence of the previous behaviour of the system as the training data then generate a sequence of inputs with proper length and their corresponding outputs from first, the 90% of training data and with respect to the size of the best period regarded to the previously discussed. Subsequently, we construct a series of networks with an initial guess for the number of hidden layers neurons and initialize their parameters randomly. Finally, for every network, the parameters vector will stop on a local minimum of its performance surface. Up to this point all of the networks are over fitted on the training set. Afterward, a simulated annealing process is applied on each network. To do this, the model is modified to generate a set of vectors named the noise vectors. Length of each noise vector is equal to the length of each network parameters vector and its components are random numbers with uniform distribution between -0.05 and +0.05. By adding noise vectors to the network parameter vectors, a new set of network parameters are obtained. This action makes relatively minor changes in the location of each network in its state space. Networks are trained with these noisy parameters until another local minimum is achieved. Making noise vectors and training are repeated for a number of times and the outputs of these networks are compared on the followed 10% of the data that are not used during training steps. The winner has the best generalization amongst all and is selected as the first member of an ensemble of neural networks.

Support vector machine model SVM

SVM possess great potential and superior performance as it

appeared in many previous researches. This is largely due to the structural risk minimization (SRM) principle in SVM that has greater generalization ability and is superior to the empirical risk minimization (ERM) principle as adopted in neural networks. In SVM, the results guarantee global minima whereas ERM can only locate local minima. For example, the training process in neural networks, the results give out any number of local minima that are not promised to include global minima. Furthermore, SVM is adaptive to complex systems and robust in dealing with corrupted data. This feature offers SVM a greater generalization ability that is the bottleneck of its predecessor, the neural network approach.

Considering a set of training data $\{(x_1, y_1),....,(x_\ell, y_\ell)\}$, where each $x_i \subset R^n$ denotes the input space of the sample and has a corresponding target value $y_i \subset R$ for i=1,..., l where l corresponds to the size of the training data (Vapnik, 1995; Muller et al., 2000). The idea of the regression problem is to determine a function that can approximate future values accurately.

The generic SVR estimating function takes the form:

$$f(x) = (w \cdot \Phi(x)) + b \quad (4)$$

Where $w \subset R^n, b \subset R$ and Φ denotes a non-linear transformation from R^n to high dimensional space. Our goal is to find the value of w and b such that values of x can be determined by minimizing the regression risk:

$$R_{reg}(f) = C\sum_{i=0}^{\ell}\Gamma(f(xi) - yi) + \frac{1}{2}\|w\|^2 \quad (5)$$

where $\Gamma(\cdot)$ is a cost function, C is a constant and vector w can be written in terms of data points as:

$$w = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)\Phi(x_i) \quad (6)$$

By substituting Equation 3 into Equation 1, the generic equation can be rewritten as:

$$f(x) = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)(\Phi(x_i) \cdot \Phi(x)) + b$$

$$= \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)k(x_i, x) + b \quad (7)$$

In Equation 7, the dot product can be replaced with function $k(x_i, x)$, known as the kernel function. Kernel functions enable dot product to be performed in high-dimensional feature space using low dimensional space data input without knowing the transformation Φ. All kernel functions must satisfy Mercer's condition that corresponds to the inner product of some feature space. The radial basis function (RBF) is commonly used as the kernel for regression.

Performance criteria

Due to the fact that water parameters had been truthfully monitored over the 5-year period, the performances of the proposed models could be examined and evaluated. The performances of the models were evaluated according to three statistical indexes. Coefficient of efficiency (CE) is often used to evaluate the model performance, introduced (Nash and Sutcliffe, 1970).

$$CE = 1 - \frac{\sum_{i=1}^{n}(X_m - X_p)^2}{\sum_{i=1}^{n}(X_m - \overline{X}_m)^2} \quad (8)$$

where n is the number of observations, X_p and X_m are the predicted and measured parameter, respectively and $\overline{X}_m$ is the average of measured parameter. The mean square error (MSE) can be used to determine how well the network output fits the desired output. The smaller values of MSE ensure better performance. It is defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_m - X_p)^2 \quad (9)$$

The coefficient of correlation (CC) is often used to evaluate the linear relationship between the predicted and measured parameter. It is defined as follows:

$$CC = \frac{\sum_{i=1}^{n}(X_m - \overline{X}_m)(X_p - \overline{X}_p)}{\sqrt{\sum_{i=1}^{n}(X_m - \overline{X}_m)^2 \sum_{i=1}^{n}(X_p - \overline{X}_p)^2}} \quad (10)$$

RESULTS AND DISCUSSION

One of the most important characteristics of MLP-ANN technique is the number of neurons in the hidden layer. If an insufficient number of neurons are used, the network will be unable to model the complex data and the resulting fit will be poor. On the contrary, if too many neurons are used, the training time may become excessively long and the network may over fit the data. In this study, different MLP-ANN architectures were used to examine the best performance. In fact, there is no formal and/or mathematical method for determining the appropriate "optimal set" of the key parameters of neural network. Therefore, it was decided to perform this task utilizing the trial and error method. The neurons of the hidden layer were randomized from N = 1 to 20 neurons and the best number of nodes in the hidden layer is the one that gives the lowest error (Lek et al., 1996).

The optimum number of neurons was determined based on two performance indices. The first index is the root-mean-square error (RMSE) value of the prediction error and the second index is the value of the maximum error. Both indices were obtained while examining the

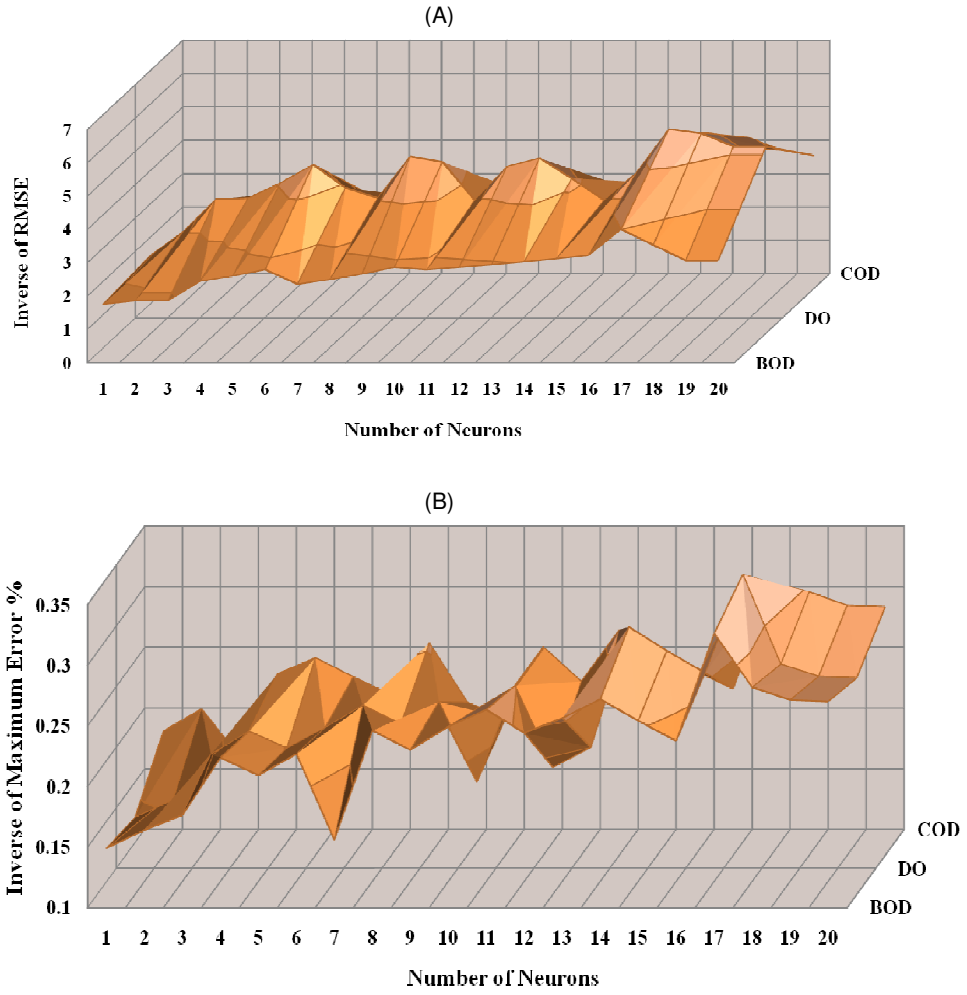

Figure 2. Neural network performance utilizing different number of neurons, (A) inverse of RMSE and (B) inverse of maximum error%.

ANN model with the WQP data between 1998 and 2007. In fact, in developing such a predicting model using neural network, the model could perform well during the training period and might provide a higher level of error when evaluating during either the validation or testing period. In the context of this study, these performance indices were used to make sure that the proposed model could provide consistent levels of accuracy during all periods.

The advantages of utilizing these two statistical indices as a performance indicator of the proposed model are first, to make sure that the highest error while evaluating the performance is within the acceptable error for such a forecasting model. This is done while utilizing the RMSE to ensure that the summation of the error distribution within the validation period is not high. Consequently, by using both indices, it guarantees the consistent level of errors by providing a great potential for having the same level error while examining the model for unseen data in the testing period. In order to show how the trial and error procedure for selecting the best number of neurons of certain ANN architecture was performed; the relationship between the numbers of neurons versus RMSE and maximum error are presented in Figure 2 for better visualization.

The inverse value of both RMSE and maximum error were used as seen in Figure 2a and b instead of the real values. It is interesting to observe the large number of local minima that exist in both domains. Changing the number of hidden neurons to the network clearly affects the prediction performance to a significant degree. It clearly shows that the prediction performance increases as the number of hidden neurons were increased (from 1 to 18), with a corresponding decrease in RMSE and maximum error for all parameters. However, adding-up the hidden neurons further (19 to 20) to the network resulted in a drop of prediction performance. For example, it can be observed that the best combination of the proposed statistical indices for evaluating the predicting model for the DO when the ANN architecture

Table 1. The ANN architecture for each parameter.

Parameter	Number of neuron	RMSE	Maximum error (%)	TFHL	TFOL	TA
DO	17	0.17	2.92	TS	PL	LMA
BOD	17	0.28	3.07	TS	PL	LMA
COD	18	0.24	3.29	LS	PL	LMA

TFHL: Transfer function between input layer and hidden layer; TFOL: transfer function between hidden layer and output layer; TA: training algorithm; LS: log sigmoid; TS: tan sigmoid; PL: pure-line and LMA: Levenberg-Marquardt algorithm.

Table 2. The optimal SVM parameters for training and testing data sets.

Parameter	Input	SVM parameter			Training		Testing	
		C	ε	γ	CC	CE	CC	CE
DO	5	10	0.3	0.35	0.983	0.966	0.964	0.929
BOD	5	8	0.3	0.16	0.971	0.942	0.956	0.913
COD	5	10	0.3	0.32	0.965	0.931	0.973	0.946

has 17 neurons, achieving RMSE 0.17 and maximum error 2.92%. While the best combination of the proposed statistical indices for evaluating the predicting model for the BOD when the ANN architecture has 18 neurons, achieving RMSE 0.24 and maximum error 3.29%. The optimal numbers of neurons for the rest parameters are presented in Table 1.

In case of SVM, it is essential to determine approximate values of optimal hyper parameters C, ε and γ. In fact, there is no formal and/or mathematical method for determining the appropriate “optimal set” of the key parameters of neural network. Therefore, it was decided to perform this task utilizing the trial and error method. For each parameter, different ranges of kernel hyper parameters c, ε and γ were adopted. For instance, the range of C was set to [1-10] at increment of 1.0 and [0.1 - 0.5] at increment of 0.1 for ε and γ. The optimal values of hyper parameters are selected based on 10-fold cross-validation repeated ten times until it reached the optimal result. Table 2 demonstrates the best result for the training and prediction data sets. Apparently, the results show that the performance of SVM is sensitive to the hyper parameters, where SVM model that is used to predict the DO reached the best results when C equal to 10, ε equal to 0.3 and γ equal to 0.35. While the C, ε and γ are equal to 8, 0.3 and 0.16, respectively for BOD and 10, 0.3 and 0.32 for COD, respectively.

In the SVM theory, there are different types of kernel functions that can be used, such as linear, polynomial, RBF and sigmoid. Different base functions are applicable for dealing with different types of data. The RBF kernel non-linearly maps the samples into the high-dimensional space, so it can handle non-linear problem. Furthermore, the linear kernel is a special case of the RBF. The sigmoid kernel behaves like the RBF for certain parameter; however, it is not valid under some parameters. The second reason is the number of hyper parameters which influences the complexity of model selection. The polynomial has more parameters than the RBF kernel. In this paper, we study the result of the use of compactly supported RBF kernels. RBF kernels are frequently used in non-linear function estimation problems (Evgeniou et al., 2000).

After optimizing the proposed models, the next step is to test the models with the test sequence data. Figure 3 depicts the error distribution of the three models that were adopted to predict the DO, BOD and COD, respectively. The results showed that the use of ensemble model led to more accurate results than MLP-ANN, where the percentage of error distribution of ensemble did not exceed 10% for DO, while the percentage of error distribution of MLP-ANN exceeded the 15%. Same results were obtained for BOD and COD, where the percentage of error distribution of ensemble did not exceed 10%, while the percentage of error distribution of MLP-ANN exceeded the 15%, respectively. These results show one advantage of utilizing ensemble over the MLP-ANN which is the ability to perceive the input-output mapping in the historical data for better prediction.

It can also be observed from Figure 3 that the SVM based model outperformed the ensemble and was able to provide improvement in prediction accuracy of water quality parameters, where percentage of error distribution of SVM did not exceed 5% for DO, BOD and COD. The significant observed that the ranges of error for SVM based model at all, are narrow, which means that the proposed model can provide a reliable and consistent level of accuracy. This result showed that the SVM models could be considered as the appropriate modelling technique for prediction of such water quality parameters

Plots of residuals versus predicted water parameters could be more useful regarding model fitting to a data set. If the residuals appear to behave randomly, it suggests

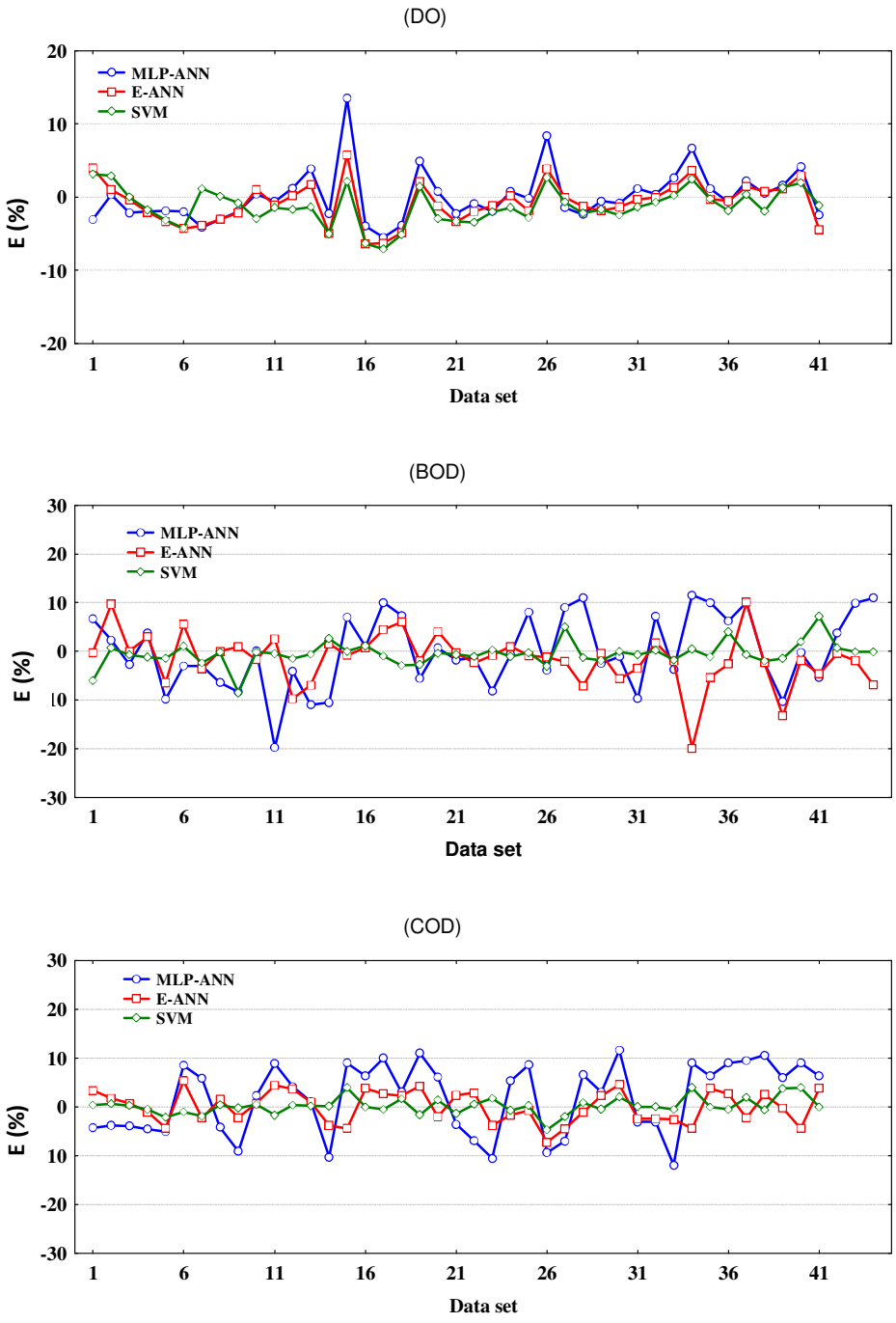

Figure 3. Depicts the error distribution of the three models that adopted to predict the DO, BOD and COD.

that the model fit the data well. In contrast, if non-random distribution is evident in the residuals, the model does not fit the data effectively (Singh et al., 2009). Residuals versus predicted water parameters of the proposed model (SVM) as shown in Figure 4. The observed relationships between residuals and model predicted for all three parameters show complete independence and random distribution. It is obvious from the figure that the markers are well distributed on both sides of the horizontal line of zero coordinates which represents the mean of the residuals and the respective correlation for DO, BOD and COD equal to 0.01, 0.019 and 0.00016, respectively which are slight small.

In the systematic analysis, the performance evaluation of various possible combinations of the parameters was investigated utilizing coefficient of efficiency (CE) and

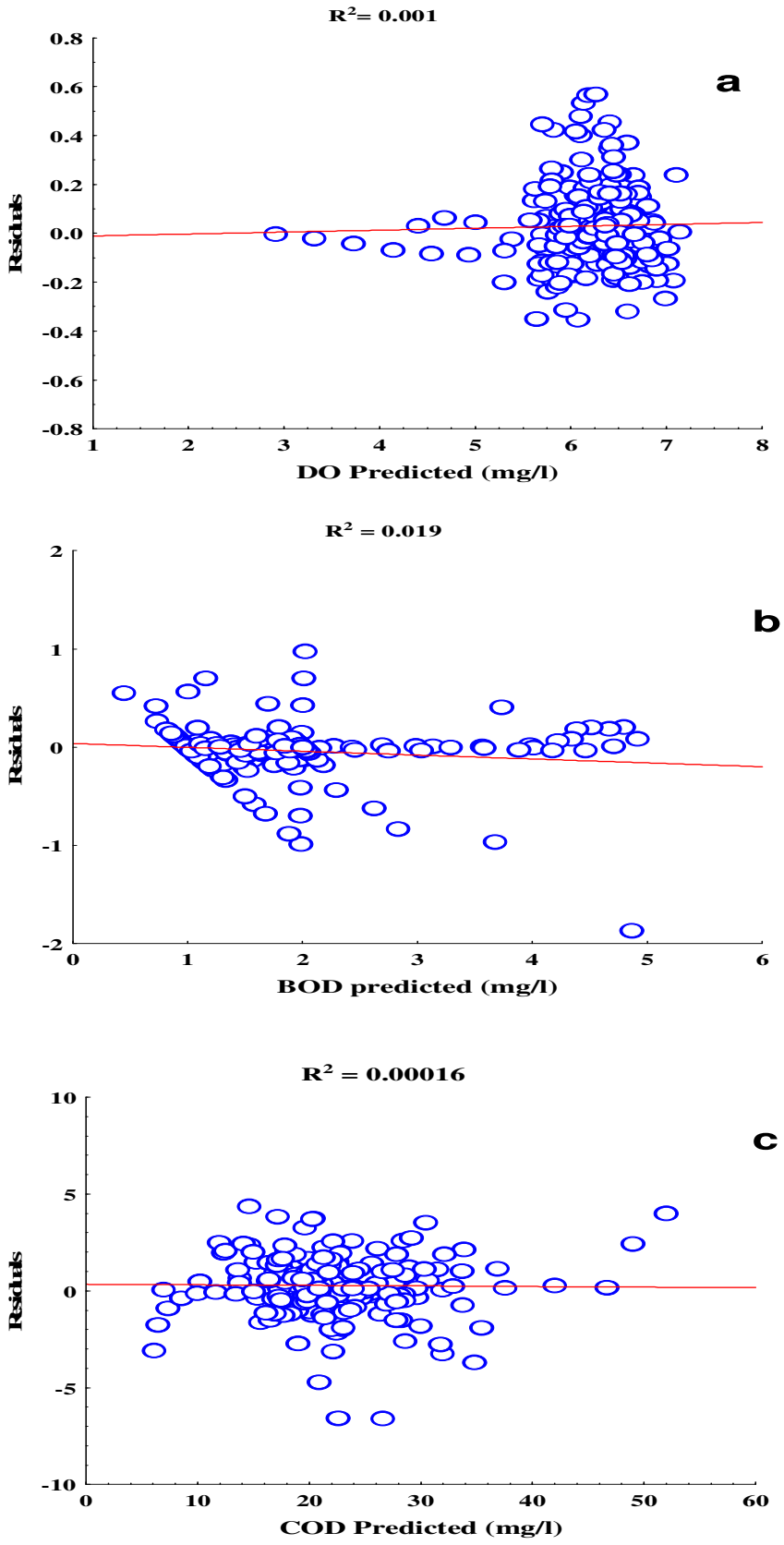

Figure 4. Demonstration of the residuals versus predicted water quality parameters: (a) DO, (b) BOD and (c) COD.

Table 3. Evaluation of possible combinations of input variables.

Combination					DO		BOD		COD	
					MSE	**CE**	**MSE**	**CE**	**MSE**	**CE**
COND					45	0.02	20	0.32	0.82	0.58
COND	TEMP				0.92	0.45	0.69	0.69	0.69	0.60
COND	TEMP	NO3			0.62	0.39	0.22	0.71	0.72	0.65
COND	TEMP	NO3	CI		0.21	0.74	0.24	0.76	0.21	0.75
COND	TEMP	NO3	CI	Na	**0.12**	**0.93**	**0.07**	**0.91**	**0.07**	**0.95**

mean square error (MSE) approaches to determine the most effective parameters on the output. In the analysis, performance evaluation of various possible combinations of variables was investigated. Generally, all networks for each parameter were compared, as shown in Table 3. Findings of the sensitivity analysis showed that nitrate was found to be the effective parameter when add to the model for all parameters. The MSE value becomes smaller when more combination variables were used together. On the basis of the performance evaluation of combinations of input variables, best group performances according to number of parameters was equal to five. The respective MSE values, as given in Table 3 show that MSE values decrease as the number of variables in the group increases.

For further visualization of the proposed SVM model performance, a demonstration of the comparison of SVM model versus the measured water quality parameter concentrations during testing data set is shown in Figure 5. It is obvious that the proposed SVM model was able to mimic the pattern (dynamics) in the measured values, in addition, for those extreme and low values experienced during this set. The proposed model showed efficiency in predicting the concentration of water quality parameters in the Johor river, and it was compatible with the results of other researchers/authors. The results also indicated that the proposed model was basically an attractive alternative, offering a relatively fast algorithm with good theoretical properties to predict the water quality parameters and can be extended to predict different water quality parameters.

However, it should be emphasized that there are no structured methods today to identify what network structure can be best to approximate the function, mapping the inputs to outputs. In addition, pre-processing for the data is an essential step for water quality prediction model and required more survey and analysis that could lead to better accuracy in their application. Moreover, the optimal selection of the key parameter still required to be achieved by augmenting the AI model with other optimization model such as genetic algorithm or particle swarm optimization methods. On the other hand, the variable selection (input pattern) in AI model is always a challenging task due to the complexity of the hydrologic process.

Conclusion

Classical process-based modeling approaches can provide relatively accurate predictions for water quality parameters; however, these models rely on data sets that require a long time to process and a large amount of input data that is often unknown.

The artificial intelligence (AI) is a new technique with a flexible mathematical structure that is capable of identifying complex non-linear relationships between input and output data when compared with other classical modelling techniques. In this study, three different model techniques, including: Multi-layer perceptron neural networks (MLP-ANN), ensemble neural networks (E-ANN) and support vector machine (SVM), were performed to identify the optimal prediction for water quality parameters along the Johor river basin. Both MLP-ANN and E-ANN model, did not reach high level of accuracy in term of water quality prediction. This result shows that it is difficult to produce a reliable model with ANN modelling approaches due to the high variance and inherent non-linear relationship of the water quality parameters due to their stochastic nature and the chemical processes. E-ANN models experienced slow convergence during training, because they required a relatively large number of hidden neurons and hidden layers. Accordingly, SVM was adopted to overcome this drawback.

Further enhancement of water quality parameters prediction was achieved when SVM was utilized. We observed that the SVM converges to a solution faster than the E-ANN and requires fewer computational complexities to train the network. In addition, the SVM improved the precision of water quality parameters prediction with minimal computation.

ACKNOWLEDGEMENTS

The authors wish to thank the Department of Environment (DOE) for providing the required data for developing this research. This research was supported by the research grant for the second and third authors from University Kebangsaan Malaysia UKM-DLP-2011-002, UKM-GUP-PLW-08-13-308 respectively.

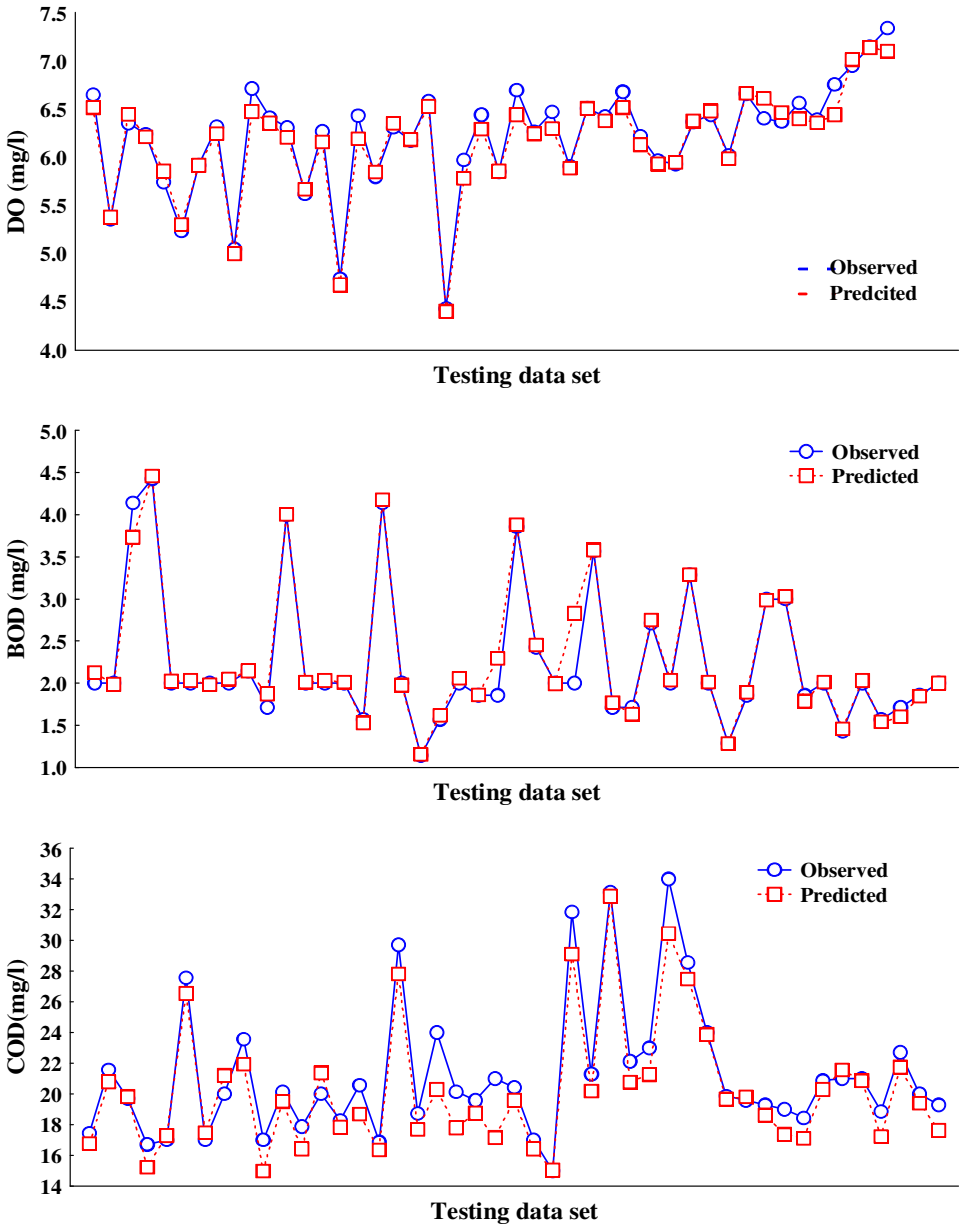

Figure 5. Demonstration of the comparison of SVM model versus the measured water quality concentrations during testing data set.

REFERENCES

Chen Q, Mynett AE (2003). Integration of data mining techniques and heuristic knowledge in fuzzy logic modelling of eutrophication in Taihu Lake. Ecol. Modell. 162 (1/2), 55-67.

Chiewchanwattana S, Lursinsap C, Chu CH (2002). Time series Data Prediction Based on Reconstruction of Missing Samples and Selective Ensembling of FIR Neural Networks, Proceedings Of The 9th International Conference on Neural Information Processing.

Drucker H, Cortes C, Jackel D (1994). Boosting and Other Ensemble Methods. Neural Computation, pp. 1289-1301.

El-shafie AE, Noureldin M, Taha R, Basri H (2008). Neural network model for Nile River inflow forecasting analysis of historical inflow data. J. Appl. Sci., 8 (24): 4487-4499.

El-shafie A, Mukhlisin M, Najah AA, Taha MR (2011). Performance of artificial neural network and regression techniques for rainfall-runoff prediction. Int. J. Phys. Sci. 6(8):1997-2003, 18 April,

Evgeniou T, Pontil M, Poggio T (2000). Regularization networks and support vector machines", Advances in Computational Mathematics, 13(1): 1-50.

Freeman JA, Skapura (1991). Neural networks algorithms, applications, and programming techniques. Addison-Wesley DM.

Grubert JP (2003). Acid deposition in the eastern United States and neural network predictions for the future. J. Environ. Eng. Sci. 2 (2): 99-109.

Heathcote IW (1998). Integrated Watershed Management: Principles and Practices, John Wiley & Sons, Inc. New York.

Hornik K, White H (1989). Multilayer feed forward networks are universal approximators. Neural Netw., 2:359-366.

Kasabov NK (1996). Foundations of neural networks, fuzzy systems,

and knowledge engineering MIT Press Cambridge.
Lee JHW, Huang Y, Dickmen M, Jayawardena AW (2003). Neural network modelling of coastal algal blooms. Ecological Modelling 159:179-201.
Lee JHW, Wong KTM, Huang Y, Jayawardena AW (2000). A real time early warning and modeling system for red tides in Hong Kong. In: Wang, Z.Y. (Ed.), Proceedings of the Eighth International Symposium on Stochastic Hydraulics. Balkema, Beijing, pp. 659-669.
Lek S, Delacoste M, Baran P , Dimopoulos I , Lauga J, Aulagnier S (1996). Application of neural networks to modeling nonlinear relationships in ecology. Ecol. Model 90:39-52.
Liong SY, Lim WH, Paudyal G (1999). Real time river stage forecasting for flood Bangladesh: neural network approach. J. Comput. Civil Eng. ASCE 14 (1): 1-8.
Najah A, Elshafie A, Karim OA, Jaffar O (2009). Prediction of Johor River water quality parameters using artificial neural networks. Eur. J. Sci. Res., 28(3):422-435 ISSN 1450-216X
Najah, A., El-Shafie A, Karim OA, Jaffer O (2010a). Water Quality Prediction Model Utilizing Integrated Wavelet-ANFIS Model with Cross Validation. Neural Computing and Applications J. DOI: 10.1007/s00521-010-0486-1
Najah AA, El-Shafie A, Karim OA, Jaafar O (2010b). Water Quality Prediction Model Utilizing Integrated Wavelet-ANFIS Model with Cross Validation, Neural. Comput. Appl., in press, doi: 10.1007/s00521-010-0486-1.
Najah A, Elshafie A, Karim OA, Jaffar O (2010c). Evaluation the efficiency of radial basis function neural network for prediction of water quality parameters. Eng. Int. Syst. 4: 221-231.
Nash JE, Sutcliffe JV (1970). River flow forecasting through conceptual models. Part 1: a discussion of principles. J. Hydrol. 10(3): 282-290.
Muller KR, Smola A, Ratch G, Scholkopf B, Kohlmorgen J, Vapnik V (2000). Using Support Vector Support Machines for Time Series Prediction, Image Processing Services Research Lab, AT&T Labs.
Muttil, N, Chau, KW (2006). Neural network and genetic programming for modelling coastal algal blooms. Int. J. Environ. Pollut., 28 (3/4): 223-238.
Makarynskyy O (2004). Improving wave predictions with artificial neural networks. Ocean Eng., 31: 709-724.
Masters T (1993). Practical Neural Network Recipes in C ++. San Diego: Academic Press.
Olden JD, Jackson DA (2001). Fish-habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks. Trans. Am. Fish. Soc., 130: 878-897.
Razavi SV, Jumaat MZ, Ahmed H El-Shafie, Mohammadi P (2011a). General regression neural network (GRNN) for the first crack analysis prediction of strengthened RC one-way slab by CFRP Int. J. Phys. Sci., 6(10): 2439-2446
Razavi SV, Jumaat MZ, Ahmed H El-Shafie (2011b). Using feed-forward back propagation (FFBP) neural networks for compressive strength prediction of lightweight concrete made with different percentage of scoria instead of sand. Int. J. Phys. Sci., 6(6): 1325-1331
Rumelhart D, Hinton GE, Williams RJ (1986). Learning internal representations by error propagation, in parallel Distributed Processing vol. 1 chap. 8, ed by Rumelhart DE JL, McCelland, MIT Press, Cambridge, Mass.
Reckhow KH (1999). Water quality prediction and probability network models. Can. J. Fish. Aquat. Sci., 56: 1150-1158.
Vapnik VN (1995). The Nature of Statistical Learning Theory. Springer, New York.