Marine Pollution Bulletin 64 (2012) 2409–2420

Contents lists available at SciVerse ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors

Nabeel M. Gazzaz [a,*], Mohd Kamil Yusoff [a], Ahmad Zaharin Aris [b], Hafizan Juahir [b], Mohammad Firuz Ramli [a]

[a] Department of Environmental Sciences, Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 Serdang, Selangur Darul Ehsan, Malaysia
[b] Center of Excellence for Environmental Forensics, Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangur Darul Ehsan, Malaysia

ARTICLE INFO

Keywords:
Surface water
Kinta River
Water quality index
Artificial neural network
Three-layer perceptron
Quickprop algorithm

ABSTRACT

This article describes design and application of feed-forward, fully-connected, three-layer perceptron neural network model for computing the water quality index (WQI)[1] for Kinta River (Malaysia). The modeling efforts showed that the optimal network architecture was 23-34-1 and that the best WQI predictions were associated with the quick propagation (QP) training algorithm; a learning rate of 0.06; and a QP coefficient of 1.75. The WQI predictions of this model had significant, positive, very high correlation ($r = 0.977$, $p < 0.01$) with the measured WQI values, implying that the model predictions explain around 95.4% of the variation in the measured WQI values.

The approach presented in this article offers useful and powerful alternative to WQI computation and prediction, especially in the case of WQI calculation methods which involve lengthy computations and use of various sub-index formulae for each value, or range of values, of the constituent water quality variables.

1. Introduction

Water quality (WQ) is a description of biological, chemical, and physical characteristics of water in connection with intended use(s) and a set of standards (Boyacioglu, 2007; Khalil et al., 2011; Liou et al., 2004). Hence, water quality assessment can be defined as the evaluation of the biological, chemical, and physical properties of water in reference to natural quality, human health effects, and intended uses (Fernández et al., 2004; Pesce and Wunderlin, 2000). Nonetheless, the WQ can be evaluated by a single parameter for certain objective or by a number of critical parameters selected carefully to represent the pollution level of the water body of concern and reflect its overall WQ status. However, since no individual parameter can express the WQ sufficiently, the WQ is normally assessed by measuring a broad range of parameters (e.g., temperature; pH; electric conductivity (EC); turbidity; and the concentrations of a variety of pollutants, including pathogens, nutrients, organics, and metals). In consequence, a large amount of data is generated by the monitoring programs and these data require integration if the monitoring results are to be presented in a meaningful way to local planners and decision makers, watershed managers, and the general public. In view of this, water quality indices have been developed to integrate measurements of a set of parameters into a single index (Zandbergen and Hall, 1998). A quality index is a unitless number that assigns a quality value to an aggregate set of measured parameters (Pesce and Wunderlin, 2000). So, the water quality index (WQI) may be defined as a single numeric score that describes the WQ condition at a particular location in a specific time (Kaurish and Younos, 2007).

The WQIs have been designed to evaluate suitability of water for certain uses. The main idea of these indices is comparison of some water quality variables (WQVs) with WQ standards so that the indices will reveal the variable(s) exceeding the standards as well as the frequency and extent of exceedance. These indices offer several advantages including representation of measurements on many variables varying in measurement units in one metric, thus establishing a criterion for tracking changes in WQ over time and space and simplifying communication of the monitoring results (Fernández et al., 2004). Besides, when pollution is identified and remedial action is taken, the WQI can be used to track and follow-up any incremental WQ improvement trends to determine effectiveness of stream restoration efforts (Kaurish and Younos, 2007).

In 1974, the Department of Environment of Malaysia recommended adoption of WQ indexing to evaluate and rank the levels of pollution of the Malaysian rivers. Then, this department adopted

* Corresponding author. Tel.: +60 123 579 864; fax: +60 3 8946 7408.
E-mail address: NabeelMGazzaz@Yahoo.com (N.M. Gazzaz).

[1] *Abbreviations:* AAE, average absolute error; ANN, artificial neural network; FA, factor analysis; MAE, mean absolute error; MSE, mean squared error; N_h, number of hidden neurons; PFA, principal factor analysis; QP, quick propagation (Quickprop); r, correlation coefficient; R^2, coefficient of determination; SEM, standard error of the mean; SSE, sum of squared errors; WQ, water quality; WQI, water quality index; WQV, water quality variable.

0025-326X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.marpolbul.2012.08.005

the so-called 'Opinion Poll WQI (OP-WQI),' also known as the Department of Environment WQI (DoE-WQI), as the method of choice for calculation of an index of the WQ of local rivers. This index is a set of WQ guidelines that categorize the WQ of local rivers into five classes based on their suitability for a number of target uses including public water supplies, fish culture, and irrigation. However, as will be outlined in Section 2.4, the method used for computing the WQI in Malaysia entails lengthy calculations and transformations. Almost the same applies to the WQI calculation methods employed in some other countries like Argentina (Pesce and Wunderlin, 2000); Brazil (Abrahão et al., 2007); India (Sargaonkar and Deshpande, 2003); Iraq (Alobaidy et al., 2010); Portugal (Bordalo et al., 2006); South Korea (Song and Kim, 2009); Spain (Debels et al., 2005; Sánchez et al., 2007); Taiwan (Liou et al., 2004); Turkey (Boyacioglu, 2007); the USA (Cude, 2001); and Vietnam (Hanh et al., 2011). In addition, some of the transformations employ different formulae for the different values, or ranges of values, of almost each input WQ parameter. Accordingly, calculation of such WQIs takes time and effort and may be occasionally associated with unintentional errors during sub-index calculations. This argument is not in any way intended to undermine or undervalue these indices which are well-established and which, being founded on solid scientific grounds, proved to be highly successful and effective in practice. Rather, we suggest alternative, direct, and quick means of computing and forecasting WQI values based on artificial neural network (ANN) modeling that has the potential to reduce the computation time and effort and the possibility of errors in the calculation. Therefore, this study illustrates design of a neural network model for rapid, direct calculation of the WQI as an alternative to WQI computation methods involving sub-indexing and lengthy calculations.

The ANNs are popular tools for modeling highly complicated relationships, processes, and phenomena. They have been successfully employed in a multitude of applications including river flow forecasting (e.g., Shamseldin et al., 2002 and Teschl and Randeu, 2006); rainfall-runoff modeling (e.g., Riad et al., 2004 and Wu and Chau, 2011); and WQ forecasting (e.g., Khalil et al., 2011, Palani et al., 2008, and Singh et al., 2009). However, the specific use of ANNs to develop models predictive of the WQI for rivers is an application which has yet to be investigated as evidenced by search in the open sources of literature which lead to only two such publications: Juahir et al., 2004 and Khuan et al., 2002. In light of this, the main objectives of this study were to (i) demonstrate the potential of the ANN for producing models capable of efficient forecasting of the WQI; (ii) illustrate the general framework for ANN model design (e.g., selection of network type; determination of appropriate input variables and number of hidden neurons; and specification of the optimum settings of the network training parameters); and (iii) establish a neural network model that can be used to directly foresee the WQ status of the river and thus provide a reliable alternative to the WQI calculation method currently in use.

2. Materials and methods

2.1. Study area

Perak is one of the 14 states of Malaysia and is the second largest state in terms of land area (21,006 km^2). It is bordered by Kedah and the Thai state of Yala from the north; Penang from the northwest; the Strait of Malacca from the west; Selangor from the south; and Kelantan and Pahang from the east (Fig. 1). The capital of Perak is the city of Ipoh. It is situated in Kinta Valley which overlies a limestone bed that is overlooked by limestone outcrops and bounded by the Kledang and Titiwangsa granite ranges from the west and east, respectively (Ghani et al., 2007).

Kinta River (Fig. 2) is the largest of three rivers (the other two being Pari and Pinji) passing through Ipoh. It flows from Gunung Korbu in Ulu Kinta at an altitude of around 2000 m above sea level to Perak River (about 100 km to the south west of the headwater) at an altitude of nearly 7 m above sea level. Kinta River and its tributaries (Fig. 2) drain a basin that covers an area of about 2420 km^2. The topography of the catchment consists of steep, forest-covered limestone hills and mountains in the north and east of Ipoh and a valley (Kinta Valley) in the south. Subsequent to the headwater, Kinta River flows through heterogeneous, mixed-use lands where, besides an extensive forest cover, the major land uses in the basin are mining, rubber planting, oil palm planting, urban development, and logging.

Kinta River is the most important water resource in Ipoh and the second most important water resource in Perak. It is the main source of water for drinking and irrigation in Ipoh and is the major tributary of Perak River, which is the principal source of drinking and irrigation water in Perak. In the time being, there is one dam only on Kinta River. It was constructed in 2000 with the intention of raising the water supply of Perak by 25%. This dam can provide 639,000 m^3 of water daily and is expected to satisfy the water demand in Kinta Valley until 2020 (Ghani et al., 2007).

2.2. Monitoring sites

Currently, eight WQ monitoring stations are located on Kinta River (Fig. 2). The coordinates and altitudes of these stations besides their distances from the headwater and from one the other are listed in Table 1. Six of these stations (2PK22, 2PK24, 2PK25, 2PK34, 2PK33, and 2PK19) were established in 1997 by the Department of Environment (Malaysia) for WQ monitoring purposes. In 2005, 2 stations (2PK59 and 2PK60) were added to the monitoring network for the same purpose. The eight stations were located in such away as to cover the river network comprehensively and to render data representative of the WQ of the whole river. In light of this, the present study divided Kinta River basin into 8 sub-basins each defined as the land areas contributing flow uniquely to 1 of the 8 WQ monitoring stations (Fig. 2).

2.3. The water quality parameters

The original WQ dataset was obtained from the Department of Environment (Malaysia) which conducts regular monitoring of the quality of Kinta River water during 7 months in every year (February, March, May, June, August, September, and November) since February 1997. This dataset comprised 9180 data points derived from 36 measurements on 255 samples. It reports the values of a set of water pollution indicators for the monitoring locations which cover the upper, middle, and lower sections of the river basin, starting from 15.4 km downstream of the river's headwater at Gunung Korbu and ending at 17.2 km upstream of the point of confluence of Kinta River with Perak River. These data constitute the first systematic study of this river and hence they represent valuable reference data against which to compare related future research as well as water management and conservation plans and efforts.

The 36 monitored parameters are statistically made up of sampling date; three categorical variables (water color, water level, and weather); and 32 continuous variables: longitude; latitude; temperature; turbidity; EC; salinity; pH; counts of the *Escherichia coli* bacteria; counts of the total coliform bacteria; and the concentrations of suspended solids (SS), dissolved solids (DS), total solids (TS), ammonia nitrogen (NH_3-N), dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), sodium (Na), potassium (K), calcium (Ca), magnesium (Mg), nitrate nitrogen (NO_3-N), chloride (Cl), phosphate phosphorous (PO_4-P),

Fig. 1. Map of peninsular Malaysia. The map shows the state of Perak and its capital, Ipoh.

arsenic (As), mercury (Hg), cadmium (Cd), chromium (Cr), lead (Pb), zinc (Zn), iron (Fe), oil and grease, and methylene blue active substances. However, not all WQVs were employed in this study. Specifically, latitude, longitude, and the three categorical variables were excluded from data processing and the remaining 30 continuous variables were examined for their effects on the WQ of Kinta River using principal factor analysis (Section 2.6) which reduced this number, based on the weights of the different WQVs in determining the river's WQ, to 23 variables (Table 2). Subsequently, only these 23 WQVs were employed in the ANN modeling. Values of some descriptive statistics for these variables are shown in Table 2.

2.4. The local water quality index

In 1974, the Department of Environment (Malaysia) adopted the so-called 'Opinion Poll WQI (OP WQI),' also known as the Department of Environment WQI (DoE-WQI), as the index of choice for assessing the WQ statuses of rivers in Malaysia. This WQI was developed in three steps. The first step entailed parameter selection. A panel of experts was consulted on the priority WQ parameters to select and on the weight to be assigned to each parameter. Those experts identified DO, BOD, COD, pH, NH_3-N, and SS as the WQVs of utmost concern (Khuan et al., 2002; Norhayati et al., 1997). The second step involved determination of quality function (curve), i.e., sub-index, for each selected parameter. Sub-indices are calculated by converting the value of each selected WQV into non-dimensional, scaled value through sub-index rating curve where each variable has its own rating curve on a scale of improving WQ, mostly from 0 to 100 (Kaurish and Younos, 2007; Liou et al., 2004). After computing the sub-index for each WQV using the related rating curve, the resultant sub-indices are averaged to give an overall WQI value. In the local case, the criteria and corresponding sub-index formulae needed for sub-index calculation are shown in Table 3 (Department of Environment, 2005). The third step corresponds to sub-index aggregation. Locally, the calculated sub-indices are combined to calculate the WQI according to the formula (Department of Environment, 2005; Khuan et al., 2002; Norhayati et al., 1997):

$$\mathrm{WQI} = 0.22^{*}\mathrm{SI}_{\mathrm{DO}} + 0.19^{*}\mathrm{SI}_{\mathrm{BOD}} + 0.16^{*}\mathrm{SI}_{\mathrm{COD}} + 0.15^{*}\mathrm{SI}_{\mathrm{AN}} + 0.16^{*}\mathrm{SI}_{\mathrm{SS}} + 0.12^{*}\mathrm{SI}_{\mathrm{pH}} \quad (1)$$

where all parameters, except pH, are expressed in mg/L, SI stands for sub-index, and $\mathrm{SI}_{\mathrm{DO}}$ refers to the DO percentage saturation sub-index which is calculated as illustrated in Table 3.

Based on the calculated WQI, a river may be classified into any of several classes, each reflecting the beneficial use(s) to which this river can be put. These classes are often based on standards or permissible limits of the selected pollution parameters. To this end, the Department of Environment (Malaysia) set values of the indicator WQVs and WQI that characterize each water quality class

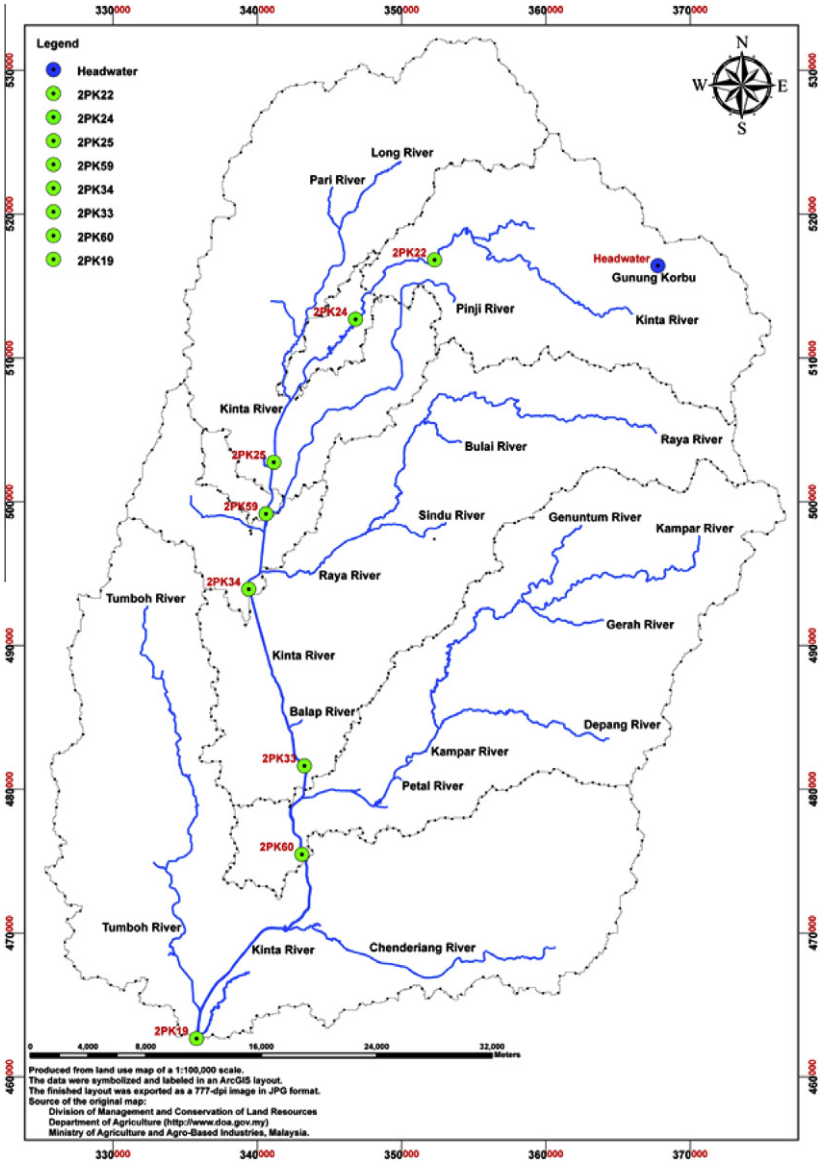

Fig. 2. Kinta River and its basin and sub-basins. The figure shows the headwater at Mount Korbu and the eight water quality monitoring stations along the river.

Table 1
Coordinates, distances, and altitudes of selected sites on Kinta River.

Site	Coordinates (RSO)[a]		Distance (km)		Altitude[b]
	Y	X	From source	Between stations	
Mount Korbu[c]	518256.14	367777.98	–	–	1599
2PK22[d]	516816.52	352365.48	15.5	15.5	264
2PK24[d]	512701.07	346869.20	22.4	6.9	50
2PK25[d]	502746.06	341721.84	33.9	11.5	34
2PK59[d]	499164.57	340907.99	37.5	3.6	28
2PK34[d]	492824.41	338308.55	43.9	6.4	23
2PK33[d]	479622.81	343084.89	57.6	13.7	11
2PK60[d]	472475.77	342025.07	64.8	7.2	9
2PK19[d]	462757.07	336011.65	77.1	12.3	7

[a] RSO: rectified skewed orthomorphic (The map projection system used for mapping in peninsular Malaysia).
[b] Height in meters above sea level.
[c] The headwater of Kinta River.
[d] Codes for the water quality monitoring stations.

and defined the uses and water treatment requirements of each class (Table 4).

Thereupon, the experimental WQI values were calculated in the current study as follows. For each water sample, the mathematical formulae presented in Table 3 were used to calculate the sub-index value for each of the six WQVs making up the local WQI (DO, BOD, COD, NH_3-N, SS, and pH). Then, the six sub-indices were aggregated using Eq. (1) to compute the value of the WQI, which served as the dependent variable in the WQV–WQI ANN model developed in this study.

2.5. Data pretreatment

Since the original dataset was large, corresponding to a matrix of 30 WQVs × 255 samples, the WQVs were first reduced in number using principal factor analysis (PFA) which aimed primarily at identifying the WQVs most responsible for the variations in the WQ of Kinta River. This analysis indicated that the latent structure of the WQ dataset was explained by 23 variables (Table 2). Then, only these variables were used in the modeling process after pretreatment as follows: (i) censored, missing, and inconsistent observations were imputed using either (a) ordinary least squares, multiple linear regression or non-linear regression (curve estima-

Table 2
Descriptive statistics of the variables most determinant of the water quality of Kinta River pooled over the monitoring years and stations.[a]

WQV	Min.[b]	Max.[c]	Range	Mean	SEM[d]	Median
Temperature (°C)	20.35	35.49	15.14	27.57	0.1300	27.58
pH	6.01	7.96	2.30	6.97	0.0200	6.96
Conductivity (μS/m)	6.00	22,170.00	22,164.00	352.62	125.0830	111.00
Turbidity (NTU)	2.0	3,306.0	3,304.0	313.2	34.0000	115.0
DS (mg/L)	0.50	7,560.00	7,559.50	145.66	43.4620	62.00
SS (mg/L)	9.00	16,500.0	16,491.0	393.89	81.7080	116.00
TS (mg/L)	29.00	16,517.0	16,488.0	538.72	91.1320	208.00
Na (mg/L)	0.370	2,150.0	2,149.6	27.29	11.6810	5.125
K (mg/L)	0.200	144.24	144.04	5.05	0.7510	3.200
Ca (mg/L)	0.003	105.00	105.00	13.06	0.7330	12.30
Mg (mg/L)	0.005	242.00	242.00	4.02	1.2400	1.90
Cl (mg/L)	0.440	4,400.0	4,399.6	51.39	25.7900	4.00
PO_4-P (mg/L)	0.0050	4.420	4.415	0.2213	0.0367	0.025
NO_3-N (mg/L)	0.0050	16.50	16.50	1.0102	0.1182	0.540
NH_3-N (mg/L)	0.0050	11.94	11.93	0.6358	0.0675	0.220
DO (mg/L)	0.000	9.930	9.930	4.317	0.1360	3.94
BOD (mg/L)	0.360	112.00	111.64	5.663	0.6560	3.00
COD (mg/L)	0.500	662.00	661.50	38.251	3.1560	28.00
As (mg/L)	0.0005	0.0404	0.0399	0.0073	0.0005	0.004
Zn (mg/L)	0.0002	0.2400	0.2398	0.0338	0.0020	0.030
Fe (mg/L)	0.0005	2.5340	2.5335	0.3728	0.0236	0.250
E. coli bacteria[e]	0	340,000	340,000	24,642	2,541	9,400
Total coliform bacteria[e]	7	4,800,000	4,799,993	105,883	25,738	41,000

[a] $N = 255$.
[b] Min.: minimum.
[c] Max.: maximum.
[d] SEM: standard error of the mean.
[e] Unit is count per 100 mL.

Table 3
The sub-index calculation formulae for the local WQI (Department of Environment (Malaysia), 2005).

Parameter	Value[a]	Sub-index equation
DO (%Saturation)[b]	$X \leqslant 8$	$SI_{DO} = 0$
	$8 < X < 92$	$SI_{DO} = -0.395 + 0.03 * X^2 - 0.0002 * X^3$
	$X \geqslant 92$	$SI_{DO} = 100$
BOD	$X \leqslant 5$	$SI_{BOD} = 100.4 - 4.23 * X$
	$X > 5$	$SI_{BOD} = (108 * e^{-0.055 * X}) - 0.1 * X$
COD	$X \leqslant 20$	$SI_{COD} = 99.1 - 1.33 * X$
	$X > 20$	$SI_{COD} = (103 * e^{-0.0157 * X}) - 0.04 * X$
NH_3-N	$X \leqslant 0.3$	$SI_{AN} = 100.5 - 105 * X$
	$0.3 < X < 4$	$SI_{AN} = (94 * e^{-0.573 * X}) - 5 * \lvert X - 2 \rvert$
	$X \geqslant 4$	$SI_{AN} = 0$
SS	$X \leqslant 100$	$SI_{SS} = (97.5 * e^{-0.00676 * X}) + 0.05 * X$
	$100 < X < 1000$	$SI_{SS} = (71 * e^{-0.0016 * X}) - 0.015$
	$X \geqslant 1000$	$SI_{SS} = 0$
pH	$X < 5.5$	$SI_{pH} = 17.2 - 17.2 * X + 5.02 * X^2$
	$5.5 \leqslant X < 7$	$SI_{pH} = -242 + 95.5 * X - 6.67 * X^2$
	$7 \leqslant X < 8.75$	$SI_{pH} = -181 + 82.4 * X - 6.05 * X^2$
	$X \geqslant 8.75$	$SI_{pH} = 536 - 77 * X + 2.76 * X^2$

[a] X is the concentration of the indicated parameter in mg/L, except for pH and DO. For DO, X refers to DO percentage saturation and for pH it refers to the pH value.
[b] DO (%Saturation) = [DO (mg/L) * 12.795] −0.05.

tion) modeling techniques, or (b) the corresponding station's mean for the WQV of interest pooled over the study years; (ii) statistical outliers and structural zeros were retained; and (iii) the variables input to the neural network (23 WQVs and the WQI) were standardized to the range of the logistic sigmoid transfer (activation) function, namely, (0,1). According to Srinivasan et al. (1994), the general formula for standardization to the range (a, b) is:

$$X_s = [(b-a)^*(X_0 - X_{min})/(X_{max} - X_{min})] + a \quad (2)$$

where x_s and x_0 express the normalized and raw observations of parameter x, respectively; a and b represent the lower and upper limits of the standardization range; and x_{min} and x_{max} refer to the minimum and maximum values of the parameter x, respectively. Owing to that in the present study $a = 0$ and $b = 1$, the formula becomes:

$$X_s = (X_0 - X_{min})/(X_{max} - X_{min}). \quad (3)$$

2.6. Principal factor analysis (PFA)

This study employed PFA to simply reduce a somewhat large set of WQVs to smaller, more manageable number while preserving as much of the overall variance as possible. The method comprises four steps: standardization of input data, extraction of factors, rotation of the factor axes, and interpretation of the extracted factors. Details on FA of the current WQ dataset have been presented in one of our earlier publications (Gazzaz et al., 2012) while elaborations on the theory and practice of factor analysis can be found in Chau and Muttil (2007), Kowalkowski et al. (2006), and Reimann et al. (2002), amongst others. However a briefing of the analysis is presented next.

In this study, factor analysis was applied to a correlation matrix having the dimensions of 255 water samples and 30 WQVs. The factor extraction method was principal component analysis (hence the notion 'Principal Factor Analysis') and the rotation method was varimax rotation. So as to ensure that the different WQVs would have equal weights in the analysis, the raw WQVs were standardized through z-scale transformation to a mean of 0.0 and variance of 1.0 by using the equation (Güler et al., 2002; Kowalkowski et al., 2006):

$$Z_{ij} = (X_{ij} - \mu_j)/\sigma_j \quad (4)$$

where z_{ij} is the standard score of the jth value of the measured variable i; x_{ij} is the jth observation of variable i; and μ_i and σ_i are the mean and standard deviation values, respectively, of the population of variable i.

In the beginning of this analysis, the Kaiser-Meyer-Olkin (KMO) and Barlett's tests were performed. The KMO test predicts if the

Table 4
Water quality class, index, condition, and use (Department of Environment (Malaysia), 2005).

WQC[a]	Parameter							Water status	Water use
	pH	TSS[b]	DO[b]	BOD[b]	COD[b]	NH-N[b]	WQI[c]		
I	>7	<25	>7	<1	<01	<1.0	>92.7	Very Good	Conservation of natural environment Water supply 1 – Practically no treatment is necessary Fishery 1 – Very sensitive aquatic species
II	6–7	25–50	5–7	1–3	10–25	0.1–0.3	76.5–92.7	Good	IIA: Water Supply II – Conventional treatment required Fishery II – Sensitive aquatic species IIB: Recreational use with body contact
III	5–6	50–150	3–5	3–6	25–50	0.3–0.9	51.9–76.5	Average	Water Supply III – Extensive treatment required Fishery III: Common, of economic value, and tolerant species Livestock drinking
IV	<5	150–300	1–3	6–12	50–100	0.9–2.7	31.0–51.9	Polluted	Irrigation
V	>5	>300	<1	>12	>100	>2.7	<31	Very Polluted	None of the above

[a] WQC: water quality class.
[b] Unit: mg/L.
[c] WQI: water quality index.

data of interest are, or are not, likely to factor well. The KMO test results were evaluated following Kaiser's rules (Kaiser, 1974). Bartlett's test of sphericity examines the null hypothesis that the elements of the correlation matrix are non-correlated. The desired result is rejection of this hypothesis (McNeil et al., 2005) to confirm that the variables used in FA are correlated. Afterwards, a correlation matrix was produced to disclose the extent of variability shared between the individual WQV pairs. The subsequent step was estimation of the eigenvalues and factor loadings (or regression weight estimates) for the correlation matrix. The factor loadings provided by the rotated component matrix were utilized in assessment of the relationships between the WQVs and the extracted factors. The primary criterion followed for specifying the number of extractable factors was the eigenvalues greater than one criterion (Kaiser, 1960), also known as Kaiser's rule. On the other side, because many WQVs will potentially load on each factor and some variables may load concurrently on two or more factors (i.e., cross-loading) the WQVs loading on each factor were sorted by the proportion of variance in the data which they explain. As a rule-of-thumb, a loading less than 0.4 is categorized as weak while a loading $\geqslant 0.6$ is considered as strong. So, only the WQVs with factor loadings $\geqslant 0.4$ were selected and included in subsequent analysis.

2.7. Development of the ANN model

The ANNs represent an innovative and attractive solution to the problem of relating output variables to input ones in complex systems (Dawson and Wilby, 2001) and prediction is a common reason for employment of the neural network technology. The major steps for development of ANN models include defining the suitable model inputs, specifying network type, pre-processing and partitioning of the available data; determining network architecture; defining model performance criteria; training (optimization of connection weights); and validating the model (Dawson and Wilby, 2001; Govindaraju, 2000; Maier and Dandy, 2000). These steps are outlined below with some detail.

2.7.1. Optimization of network architecture

This study employed a parallel, feed-forward, fully-connected, multilayer perceptron (MLP) neural network in order to establish a non-linear regression model that can be used to directly foresee the WQ status of Kinta River, in terms of the local WQI, using WQ monitoring data. Construction of this model was carried out in two major steps; determination of the network architecture and specification of the network structure. The first step aimed at determining the numbers of input, hidden, and output layers; the numbers of input, hidden, and output neurons; and the optimal data splitting plan while the second phase involved specifying the training algorithm, learning rate, number of iterations, number of retrains, and training stopping criteria.

The hidden layers provide the network with its ability to generalize. In theory, a network with a hidden layer and adequate number of hidden neurons can simulate any continuous function and represents a rich and flexible class of universal approximators (Dawson and Wilby, 2001; Fischer, 2006; Palani et al., 2008). Thereupon, this study employed a neural network with one input layer, one hidden layer, and the WQI as the output layer, thus producing a three-layer perceptron (TLP) network.

Empirical datasets usually have variables of different measurement units and are quite often burdened with some measurement errors, noise, or interference. These factors may exert negative impacts on operation of some ANN training algorithms. In order to avoid such impacts it is necessary during the initial data preparation stage to standardize the data, that is, to convert the data into a non-dimensional form of uniform range of variability (Dawson and Wilby, 2001; Özesmi et al., 2006). This prevents any attribute from arbitrarily dominating the neural network modeling outputs. Hence, the input WQ data were pre-processed by standardization within the limits of the logistic sigmoid function, i.e., 0–1 (Section 2.5).

The number of water samples (or examples) available for modeling was 255 and the number of input WQVs (neurons) was 23. The 255 samples were divided into training, cross-validation (or over-fitting), and testing sets. The testing subset should include data never used in the training and cross-validation sets and this data should constitute approximately 10–40% of the size of the training set (Palani et al., 2008).

The number of input and output units is usually fixed, depending on the number of input predictors and output variables (Bruzzone et al., 2004). However, determining the number of hidden nodes is usually a trial and error task in ANN modeling (Özesmi et al., 2006; Palani et al., 2008; Singh et al., 2009). There is no magic formula for the selection but there are some rules of thumb. As an example, Fletcher and Goss (1993) stated that the appropriate number of neurons in the hidden layer (N_h) ranges from $2I^{1/2} + O$ to $2I + 1$, where I and O represent the numbers of input and output nodes, respectively. The Alyuda Research

Company (2003) however maintains that N_h should fall within the range of $I/2$ to $4I$. More recently, Palani et al. (2008) supported that N_h can lie between I and $2I + 1$ and that it should not in any way be less than the maximum of $I/3$ and O. However, networks with few hidden nodes are generally preferable to networks with many hidden nodes because the former usually have better generalization capabilities and fewer over-fitting problems than the latter (Khalil et al., 2011; Özesmi et al., 2006; Palani et al., 2008). In the current case, I is 23 and O is 1. Therefore, the typical number of hidden nodes is expected to be $\geqslant 23/3 \approx 8$ neurons and $\leqslant 23 \times 4 = 92$ neurons ($8 \leqslant N_h \leqslant 92$). Within this range, the most suitable value of N_h was determined following the trial and error approach and the whole spectrum of numbers from 8 to 92 was scanned using the NeuroIntelligence v2.2 (577) software (Alyuda Research Inc., Los Altos, CA, USA).

Thus, the researchers utilized the 'Heuristic Search' option of the NeuroIntelligence ANN software in an effort to identify (i) the optimum architecture for the WQV-WQI model and (ii) the most appropriate data partitioning scheme (Table 5). During the search process, subset selection was random and the search accuracy was 1. And since this modeling problem is a non-linear regression-type problem, the criterion for selecting the best network architecture was the amount of variance in the experimental WQI values that is explained by the model predictions, namely, model's R^2.

As to activation of the neurons, the sigmoidal-type (logistic sigmoid and hyperbolic tangent) functions and the Gaussian function are non-linear activation functions that can be used in the MLP neural networks. However, the logistic sigmoid and hyperbolic tangent functions are the functions most commonly used with the MLP neural networks (Dawson and Wilby, 2001; Maier and Dandy, 2000) while the Gaussian function is the one most commonly used with the RBF neural networks (Corsini et al., 2003; Dawson and Wilby, 2001). The logistic sigmoid function is particularly appealing when the raw data have outliers because this function reduces the effects of extreme input values on the performance of the network and hence extreme values will have no extreme effects on the network outputs (Hill et al., 1994). Consequently, this study used a linear transfer function in the input layer and a logistic sigmoid activation function in the hidden and output layers.

Eventually, this search process showed that the optimal network architecture was 23-34-1 and that the optimum associated partitioning scheme was 80%-10%-10% (i.e., the proportions of the samples allocated to the training, cross-validation, and testing sets were 80%, 10%, and 10%, respectively). This network had a training error (sum squared error, SSE) of 0.6087 and the WQI predictions it produced had a significant, very high, positive correlation with the measured WQI values ($r = 0.998$, $p < 0.01$), suggesting that these predictions explain almost 99.5% of the variations in the measured WQI values.

Table 5
Common data splitting schemes for ANN modeling purposes.

Splitting scheme[a]	Training set (%)	Selection set (%)	Testing set (%)
I	80.0[b]	10.0	10.0
II	75.0	10.0	15.0
III	70.0	10.0	20.0
IV	70.0	15.0	15.0
V	65.0	15.0	20.0
VI	60.0	20.0	20.0
VII	60.0	15.0	25.0
VIII	50.0	25.0	25.0

[a] Examples on studies where these partitioning schemes were applied include: I, Zeghal and Khogali (2007); II, A splitting scheme suggested by this study; III, May and Sivakumar (2009); IV, Lucio et al. (2007), Shanthi et al. (2009), and Banerjee et al. (2011); V, Mas and Ahlfeld (2007) ; VI, Singh et al. (2009); VII, Amiri and Nakane (2009); and VIII, Lischeid (2001), Olszewski et al. (2008), and da Costa et al. (2009).

[b] The listed figures represent the percentage equivalent of the number of observations (or samples) assigned to the indicated subset.

2.7.2. Optimization of network structure

In light of the findings related to the best network architecture, the WQ data were divided into a training subset made-up of 203 samples (79.60% of the samples) and cross-validation and testing subsets comprising 26 samples each (10.20% of the samples). Next to this, an investigation of the network structure capable of producing the most accurate WQI predictions was implemented using the 'Expert' mode of the Forecaster v1.6 (778) ANN software (Alyuda Research Inc., Los Altos, CA, USA). This step aimed mainly at identifying the optimal training algorithm, learning rate, number of iterations, number of retrains, and training stopping conditions. The optimum network structure was determined by calibrating these variables one at a time. All model parameters were calibrated by varying the values of one of the foregoing parameters while fixing the values of the others until the best WQI predictions had been achieved. Then, the value of the examined parameter was set at the optimum value and fixed and another parameter was calibrated much in the same way until the finest values of all model parameters had been determined.

Network training is the process by which the synaptic weights and bias levels of a neural network are adapted through a continuous process of stimulation by the environment in which the network is set. The different networks tested in the current study used the quick propagation (Quickprop or QP) algorithm to optimize the connection weights. In the beginning, both the QP and the batch propagation (BP) training algorithms were evaluated and their performances compared. Details about the QP algorithm have been given by Fahlman (1988) and Wilamowski (2003) while details about the BP algorithm have been provided by Eisenstein and Kanter (1993), Rumelhart et al. (1986), and Yu et al. (2002). Results of initial testing in the present study indicated that the QP algorithm always outperformed the batch BP one. The former algorithm frequently produced lower testing set errors and higher R^2 values than the latter. This finding agrees with empirical comparisons in the literature which evidenced that the QP training algorithm is one of the fastest, most reliable algorithms and that it outperforms the majority of the rest heuristic variants of the BP algorithm on a broad range of modeling problems (Fahlman, 1988; Shanthi et al., 2009). Consequently, all networks were trained with the standard QP algorithm due to its fast convergence and stability besides its better performance on the studied data than the BP algorithm.

To efficiently find a reasonable learning rate, networks are trained and the outputs of the validation dataset are evaluated (Bruzzone et al., 2004; Corsini et al., 2003). During this process, the ANN model is trained and continually optimized against the cross-validation dataset so that the network structure can be assessed by using the error values produced during training on the cross-validation data. On the other hand, the testing data set is used to assess the ANN model on unseen data once the model has been developed. The testing set error measures the network's generalization ability in a consistent fashion. For good generalization, training should be stopped when the testing set error reaches its lowest point and the topology with the minimum testing error is considered as the optimal network topology (Liao and Fildes, 2005; Olszewski et al., 2008; Qi and Zhang, 2001; Turney, 1993; Twomey and Smith, 1998). At this stage of model development, the authors also took into consideration agreement between the training and testing error values since the closer these values are to one the other, the better is the model.

Subsequently, this study determined the optimum learning rate using the 'Expert' mode of the Forecaster v1.6 software. This mode

allows the user full control over pre-processing and analysis of data, selection of the network structure, and training the neural network. In practical sense, the goal of network training is to find a good approximation to the regression by minimization of a sum-of-squares error defined over a finite training set. To this end, early stopping procedure(s) should be implemented to avoid overtraining and to improve generalization. In general, the commonly applied early stopping criteria can be classified into three categories: cross-validation; predefined number of training iterations; and predefined error value (Özesmi et al., 2006; Singh et al., 2009; Wang and Wang, 2010). This study employed four stopping criteria that fall within these three general categories to prevent network over-fitting: (i) the cross-validation technique. Kovalishyn et al. (1998) accentuated that the over-fitting problem does not have any effect on the predictive capability of the ANN when overtraining is precluded by the cross-validation technique. Training according to this measure is stopped if the error on the cross-validation subset stops changing or begins increasing (Atiya and Ji, 1997; May and Sivakumar, 2009; Tiron and Gosav, 2010); (ii) a mean-squared error (MSE) value on the training set of 0.01; (iii) a minimum improvement in error of 0.0000001; and (iv) a maximum of 10,000 iterations. On the other hand, each model was retrained 10 times, following the recommendation of the ANN software developer who specifically suggested retraining each network architecture within the range of three to, preferably, 10 times with weight randomization and initial weight adjustment (Alyuda Research Company, 2003). Accordingly, training was stopped when any of these stopping criteria was satisfied and the weights were then reset to the weights which resulted in the lowest error on the cross-validation set before testing for generalization. After training was stopped, the model with the lowest validation set error was used to forecast the WQI for the testing data and the testing set error was calculated. Then, the optimum network was selected based on the lowest testing set error.

With respect to the learning rate, the synaptic weights of the networks may be initialized with random learning rates distributed uniformly in the range of 0–1. Accordingly, the researchers tested a wide range of learning rates ranging from 0.01 to 0.09 at 0.005 increments and from 0.1 to 0.9 at 0.05 increments. Since the weights were changed according to the QP rule, then the QP coefficient rather than a learning momentum needed to be determined (Atiya and Ji, 1997). Within this context, the QP coefficient was maintained in this study at the default value of 1.75 (Alyuda Research Company, 2003; Lucio et al., 2007; Wang and Wang, 2010), complying with the recommendation of the algorithm developer (Fahlman, 1988).

2.7.3. Sensitivity analysis

A sensitivity analysis was carried out to evaluate the relative importance of each of the 23 WQVs in prediction of the WQI. Sensitivity analysis in data mining and model building, or fitting, generally refers to assessment of the importance of predictors in the fitted models. This analysis ranks the predictor variables according to the deterioration in model performance that occurs if a variable is removed from the model. Ultimately, it identifies the variables that can be ignored safely in subsequent analyses as well as the essential variables which must be retained (Olszewski et al., 2008). The analysis results may be used for purely informative purposes or for input variable pruning.

In this study, sensitivity analysis was based on examining the effects of the input variables (23 WQVs) on the dependent variable (WQI) following the so-called 'leave-one-out' method which corresponds to assessing changes in the network error that will be obtained if each input variable is removed at a time (Ha and Stenstrom, 2003; Pastor-Bárcenas et al., 2005). This method applies two indicators: ratio of network errors and rank. The ratio of network errors specifies the effects on network functioning upon removal of individual variables. It is the ratio of the error obtained after individual variable removal to the error obtained using all variables, i.e., error of the reduced model against that of the full model. The higher the ratio, the more important is the individual predictor to the model, and vice versa. When the ratio is $\leqslant 1.0$, the variable can be rejected and deleted from the model. Then, based on the ratios, ranks are assigned to the independent variables such that the variable having the highest significance for the network is assigned the rank of 1 (Olszewski et al., 2008).

2.7.4. Model selection and performance evaluation

Usually either of two broad types of model selection approaches is followed in ANN modeling. The first is the cross-validation-based approach whereas the second is the in-sample model selection method. The cross-validation-based approach divides the available data into three sets: training, validation (or cross-validation), and testing sets. The training set is employed in training the network while the cross-validation set is used for deciding on when to stop training before over-fitting takes place and it is assumed that a good model is a model that minimizes the cross-validation error (Liao and Fildes, 2005; Qi and Zhang, 2001; Turney, 1993). On the other hand, the testing set is utilized for genuine, out-of-sample evaluation (Qi and Zhang, 2001), i.e., for estimating the network performance after training has finished, and the true network error is then estimated as the testing set error (Prechelt, 1998; Qi and Zhang, 2001; Twomey and Smith, 1998). If representative training data is used, the testing set error is an optimal estimation of the actual network performance (Liao and Fildes, 2005; Prechelt, 1998; Turney, 1993). So, the training set is used for parameter estimation for a number of alternative neural network specifications (e.g., networks of different numbers of inputs and different numbers of hidden layer units). Then, the trained network is evaluated with the validation set and the network model that performs the best on the validation set is selected as the final forecasting model. Thereafter, the validity, usefulness, and generalization performance of the model is evaluated on the testing set (Prechelt, 1998; Qi and Zhang, 2001) using a suitable performance measure like the SSE or the mean (or average) absolute error (MAE or AAE) (Prechelt, 1998; Qi and Zhang, 2001; Twomey and Smith, 1998). The MAE (or AAE) is one of the best overall measures of model performance (Twomey and Smith, 1998; Willmott, 1982; Willmott and Matsuura, 2005). It has the advantage of being less sensitive to extreme values (more robust to outliers) than other error measures. It does not assign high weights to large errors but rather weighs all error sizes equally (Fox, 1981; Liao and Fildes, 2005; Willmott, 1982; Willmott and Matsuura, 2005).

In addition to the foregoing error measures, Flavelle (1992) supported that linear regression analysis of the model predictions and the measured data can be used to evaluate the results of a validation in an objective and quantitative manner. According to this approach, the coefficient of determination (R^2) or adjusted R^2 (R^2_{Adj}) is a measure of goodness-of-fit of the model to the data that represents the model's predictive capacity. Consequently, selection of the best of a number of potential models can be based on the largest value of R^2 or R^2_{Adj} where the higher the R^2 value and closer to 1.0, the better.

Therefore, this study differentiated between the different potential ANN models based on the (i) cross-validation and testing set errors (SSE and AAE, respectively); (ii) correlation between the predicted and observed WQI values (Fox, 1981; Miao et al., 2006); and (iii) the amount of variance in the measured WQI values which the model predictions explain (R^2 value). Thus, the ANN model reported here is the model which exhibited the lowest cross-validation and testing subset errors of close agreement;

highest correlation between the predicted and observed WQI values; and highest R^2 value.

3. Results and discussion

This study employed a parallel, fully-connected, feed-forward MLP network with one input layer, one hidden layer, and the WQI as the output layer. The number of examples available for modeling was 255. The numbers of neurons in the input and output layers were fixed to 23 (the number of input WQVs) and one (the WQI), respectively. The input WQ data were pre-processed by standardization to the range of (0,1). Afterwards, a search for the optimal network architecture and best data partitioning scheme was conducted using the NeuroIntelligence 2.2 (577) ANN software which serves as an initial tool for identifying the optimum network architecture for subsequent, in-depth inspection. During this step, guiding rules (Alyuda Research Company, 2003; Maier and Dandy, 2000; Palani et al., 2008) indicated that the potential N_h lies in the range 8–92; that is, higher than one third the number of input WQVs and less than four times this number. Within this range, the optimum value of N_h was determined following the trial and error approach. The search results demonstrated that the optimum network architecture was 23-34-1 and that it was obtained with the 80%-10%-10% partitioning scheme. The model predictions of the WQI produced by this network had a significant, positive, very high relationship ($r = 0.998$, $p < 0.01$) with the experimental WQI values. This r value implies that the ANN model having this architecture accounts for around 99.5% of the observed variations in the experimental WQI values ($R^2 \cong 0.995$) and underscores that the model was well-specified. Thereafter, this network architecture was subjected to further analysis in order to specify the optimal associated network structure by means of the 'Expert' mode of the Forecaster 1.6 (778) ANN software.

In consequence, the 255 WQ samples were divided into a training set made-up of 203 samples and cross-validation and testing subsets comprising 26 samples each. The network was trained using the QP algorithm and different learning rates. Differentiation between the tested learning rates was based on the testing set AAE value. On this account, the learning rate conducive to the lowest AAE (1.663) was 0.06. On the other side, Flavelle (1992) suggested use of linear regression analysis of calculated results and measured data to evaluate the results of a validation in an objective and quantitative manner and Flavelle (1992) and Schenker and Agarwal (1996) articulated that when comparing between models selection of the best model should be based on the largest R^2 or R^2_{Adj} value. For that, the ordinary least square multiple linear regressions were an additional measure which the researchers followed in evaluating performance of the final model. The WQI predictions of the 23-34-1 network trained using the QP algorithm and a learning rate of 0.06 had a significant, positive, very high correlation ($r = 0.977$, $p < 0.01$) with the measured WQI values. This means that the ANN predictions of the WQI explain around 95.4% of the variations in the measured WQI values. This conclusion is reinforced by Figs. 3 and 4. Fig. 3 compares between the ANN-predicted and the measured WQI values for each single observation while Fig. 4, which is a scatter plot of the model's WQI outputs versus the observed values, shows the validation results for the network. Both figures show that the overall agreement between the observed and simulated WQI values was satisfactory.

Lastly, results of sensitivity analysis (Table 6) point that the six most important WQVs for WQI prediction, in descending order, are DO, BOD, NH_3-N, pH, COD, and turbidity. This suggests that for a selective chemical WQ monitoring plan, priority consideration should be given to these six WQVs. On the other hand, the NO_3-N, As, and PO_4-P concentrations and total coliform bacteria count almost had the same ratios and ranked as the least important WQVs for the model. In light of these findings, it may be argued that the importance ranking of the DO, BOD, and NH_3-N concentrations to WQI computation are almost the same in the ANN and the current WQI computation methods. However, the SS concentration ranked 18th in the ANN model while it almost ranked 3rd in the WQI formula in use and pH had higher weight in the ANN model than in the WQI formula currently in use. The ANN model on the other hand highlights a higher importance of turbidity than of SS concentration for WQI prediction. Accordingly, the results of sensitivity analysis briefed here on importance ranking of the WQVs may lay grounds for re-establishment of the present WQI formula if the ANN approach to WQI forecasting developed in this paper is not to be adopted by the WQ monitoring agencies.

The ANN method for WQI calculation and forecasting offers some advantages over the traditional method. The method used

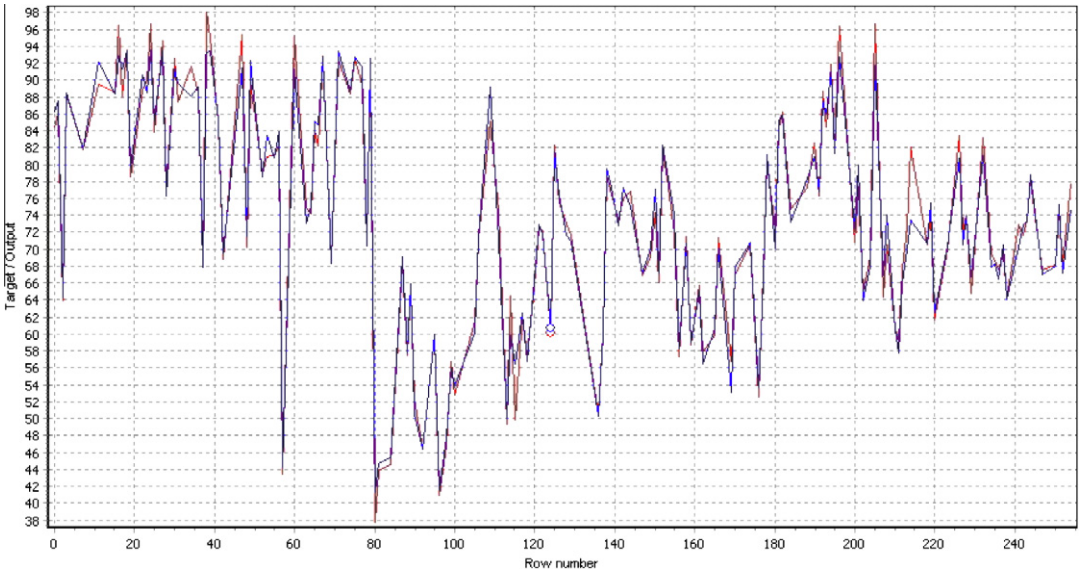

Fig. 3. A graphical representation of the results of ANN model validation. This figure compares between the ANN-predicted and the measured WQI values for each individual observation.

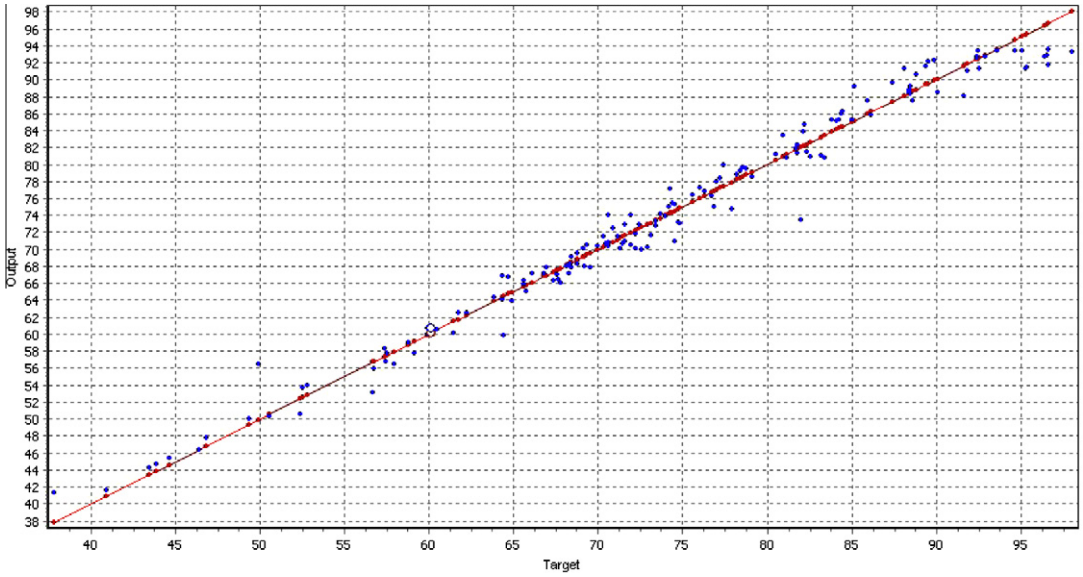

Fig. 4. Results of ANN model validation. This figure is a scatter plot of the relationship between the measured WQI values and their corresponding ANN model predictions. It demonstrates that a reasonable approximation was made by the ANN model across the spectrum of the measured WQI values. The overall agreement between the measured and simulated WQI values was very satisfactory ($r = 0.977$, $p < 0.01$; $R^2 = 95.4\%$; AAE = 1.663, $N = 255$).

Table 6
Sensitivity analysis results.

Variable	Ratio	Rank
DO	2.3443	1
BOD	1.4669	2
NH_3-N	1.3615	3
pH	1.3458	4
COD	1.2350	5
Turbidity	1.2015	6
Mg	1.1547	7
Ca	1.0752	8
K	1.0585	9
Cl	1.0473	10
Temperature	1.0422	11
E. coli bacteria	1.0404	12
Zn	1.0333	13
DS	1.0329	14
Fe	1.0285	15
TS	1.0231	16
Na	1.0188	17
SS	1.0185	18
EC	1.0174	19
NO_3-N	1.0130	20
As	1.0096	21
PO_4-P	0.9932	22
Total coliform bacteria	0.9917	23

by the Department of Environment (Malaysia) for calculation of the WQI uses manual calculations whereby the raw data of six WQVs (DO, BOD, COD, SS, NH_3-N, and pH) have to be converted into sub-indices (Table 3) before the WQI can be calculated. The calculations are not performed on the parameters themselves but rather on their sub-indices whose values are obtained from a set of equations (Table 3 and Eq. (1)). The equations listed in Table 3 are actually best-fit equations obtained from rating curves. To the contrary, the ANN approach utilized archived data to establish a model that can be used for direct calculation of the WQI from raw WQVs without need for sub-indexing. The suggested ANN approach is therefore a more direct, rapid, and convenient means of calculation of the WQI than the traditional method. Accordingly, this study accentuates that the ANN constitutes an effective tool for assessment of the river WQ that simplifies the computation of the WQI and that saves substantial efforts and time by optimizing the calculation. Moreover, the ANN approach is more comprehensive means of WQI calculation, than the WQI formula presently in use, as it uses the 23 WQVs which were identified by PFA as explaining the latent construct of the water quality data. Therefore, the ANN model has an extra advantage of addressing nutrients, microbial pollution indicators, and heavy metals, which the current WQI formula does not cover.

In other respects, we can still compare between the outputs of the ANN model and those of the traditional method for WQI computation despite the different numbers of predictors in both cases. The ANN model presented in this paper uses 17 WQVs in addition to the six WQVs which the traditional method uses to calculate the WQI. The WQI values calculated using the traditional method were set as reference values for the ANN method. Actually, the traditional method is a statistical method where the WQI is calculated based on curve estimation models/equations (Section 2.4 and Table 3). Both this method and the ANN approach are non-linear modeling techniques. However, for the ANN approach to successfully produce highly accurate estimates of the WQI relative to the WQI values calculated using the traditional method, measurements on additional 17 WQVs need to be included in the model. And since the WQI calculated using the traditional method provided reference WQI values for the ANN model, i.e., it was set as the target for the corresponding neural network model, then the performance of the ANN model can be evaluated by comparing its WQI outputs with those of the traditional method through correlation or regression analyses and/or graphical methods, e.g., Figs. 3 and 4 (Section 2.7). The results (e.g., Figs. 3 and 4) demonstrate that a reasonable approximation was made by the ANN model across the spectrum of the measured WQI values. The overall agreement between the measured and simulated WQI values was very satisfactory ($r = 0.977$, $p < 0.01$; $R^2 = 95.4\%$; AAE = 1.663, $N = 255$).

4. Conclusions

This study described the application of ANN to a prediction (or function approximation) problem entailing use of archival measurements on WQVs of a surface water body for construction of a

model capable of calculating and forecasting the WQI. It discussed common problems concomitant to design of ANN models, with example application on Kinta River (Malaysia), and demonstrated effectiveness of the ANN approach in this particular field. The power of a neural solution in rendering satisfactory models based on a reduced set of predictors has also been illustrated. Eventually, a model based on the three-layer perceptron neural network was developed for computation of the WQI. The different potential models were trained and tested on monthly data of 23 WQVs measured over a period of 10 years using a parallel, fully-connected, feed-forward network trained using the QP learning algorithm. The ANN could uncover and utilize the latent relationships in the historical WQ data efficiently and generated a model that is capable of providing highly accurate forecasts of the WQI for any monitoring site and period of time for which WQ data are available, hence facilitating and speeding-up prediction of the WQ status of the river.

Findings from this study emphasize that the ANN enables easy modeling of the WQI and allows for identification of the comparative importance and contribution of input WQVs to the model predictions. Accordingly, this study accentuates that the ANN constitutes an effective tool for assessment of the river WQ that simplifies computation of the WQI and that saves substantial efforts and time by optimizing the calculation. Thereupon, the ANN approach presented in this article constitutes a useful, powerful alternative to traditional (or statistic) WQI calculation methods, especially those methods which involve lengthy computations and use of various sub-index formulae for each value or range of values of the constituent WQVs. This approach can be commonly used and it can apply equally successfully to any aquatic system worldwide. The study results should therefore encourage WQ monitoring authorities and water resource managers to adopt ANN models as comprehensive and highly reliable alternatives to such WQI calculation methods. Therefore, empirical data analysis techniques such as the ANNs are recommended for analysis of long-term environmental monitoring records. The authors hope that this study and its outcomes provide a protocol for application of ANN models to WQI calculation.

References

Abrahão, R., Carvalho, M., da Silva Júnior, W.R., Machado, T.T.V., Gadelha, C.L.M., Hernandez, M.I.M., 2007. Use of index analysis to evaluate the water quality of a stream receiving industrial effluents. South African Water Research Commission Water SA 33 (4), 459–466.

Alobaidy, A.M.J., Abid, H.S., Maulood, B.K., 2010. Application of water quality index for assessment of Dokan Lake ecosystem, Kurdistan region, Iraq. Journal of Water Resource and Protection 2 (9), 792–798.

Alyuda Research Company, 2003. Alyuda NeuroIntelligence User Manual. Version 2.1. Alyuda Research Inc., Los Altos, CA, USA.

Amiri, B.J., Nakane, K., 2009. Comparative prediction of stream water total nitrogen from land cover using artificial neural network and multiple linear regression approaches. Polish Journal of Environmental Studies 18 (2), 151–160.

Atiya, A., Ji, C., 1997. How initial conditions affect generalization performance in large networks. IEEE Transactions on Neural Networks 8 (2), 448–451.

Banerjee, P., Singh, V.S., Chatttopadhyay, K., Chandra, P.C., Singh, B., 2011. Artificial neural network model as a potential alternative for groundwater salinity forecasting. Journal of Hydrology 398, 212–220.

Bordalo, A.A., Teixeira, R., Wiebe, W.J., 2006. A water quality index applied to an international Shared River basin: the case of the Douro River. Environmental Management 38, 910–920.

Boyacioglu, H., 2007. Development of a water quality Index based on a European classification scheme. Water SA 33 (1), 101–106.

Bruzzone, L., Cossu, R., Vernazza, G., 2004. Detection of land-cover transitions by combining multidate classifiers. Pattern Recognition Letters 25 (13), 1491–1500.

Chau, K.W., Muttil, N., 2007. Data mining and multivariate statistical analysis for ecological system in coastal waters. Journal of Hydroinformatics 9 (4), 305–317.

Corsini, G., Diani, M., Grasso, R., de Martino, M., Mantero, P., Serpico, S.B., 2003. Radial basis function and multilayer perceptron neural networks for sea water optically active parameter estimation in case II waters, a comparison. International Journal of Remote Sensing 24 (20), 3917–3932.

Cude, C.G., 2001. Oregon water quality index a tool for evaluating water quality management effectiveness. Journal of the American Water Resources Association 37 (1), 125–137.

da Costa, A.O.S., Silva, P.F., Sabará, M.G., da Costa, E.F., 2009. Use of neural networks for monitoring surface water quality changes in a neotropical urban stream. Environmental Monitoring and Assessment 155 (4), 527–538.

Dawson, C.W., Wilby, R.L., 2001. Hydrological modelling using artificial neural networks. Progress in Physical Geography 25 (1), 80–108.

Debels, P., Figueroa, R., Urrutia, R., Barra, R., Niell, X., 2005. Evaluation of water quality in the Chillan River (Central Chile) using physicochemical parameters and a modified water quality index. Environmental Monitoring and Assessment 110 (1–3), 301–322.

Department of Environment. Malaysia Environmental Quality Report 2005. Department of Environment, Ministry of Natural Resources and Environment, Petaling Jaya, Malaysia. 2007.

Eisenstein, E., Kanter, I., 1993. Numerical study of back-propagation learning algorithms for multilayer networks. Europhysics Letters 21 (4), 501–506.

Fahlman, S.E., 1988. An Empirical Study of Learning Speed in Back-Propagation Networks. School of Computer Science. Carnegie Mellon University, Pittsburgh, USA.

Fernández, N., Ramírez, A., Solano, F., 2004. Physico-chemical Water Quality indices a comparative review. Bistua: Revista de la Facultad de Ciencias Básicas 2 (1), 19–30.

Fischer, M.M., 2006. Neural networks: a general framework for non-linear function approximation. Transactions in GIS 10 (4), 521–533.

Flavelle, P., 1992. A quantitative measure of model validation and its potential use for regulatory purposes. Advances in Water Resources 15 (1), 5–13.

Fletcher, D., Goss, E., 1993. Forecasting with neural networks: an application using bankruptcy data. Information Management 24 (3), 159–167.

Fox, D.G., 1981. Judging air quality model performance, a summary of the AMS workshop on dispersion model performance. Bulletin of the American Meteorological Society 62 (5), 599–609.

Gazzaz, N.M., Kamil, M., Firuz, M., Zaharin, A., Juahir, H., 2012. Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. Marine Pollution Bulletin 64 (4), 688–698.

Ghani, A., Zakaria, N.A., Kiat, C.C., Ariffin, J., Abu Hasan, Z., Abdul Gaffar, A.B., 2007. Revised equations for manning's coefficient for sand-bed rivers. International Journal of River Basin Management 5 (4), 329–346.

Govindaraju, R.S., 2000. Artificial neural networks in hydrology. I, preliminary concepts. Journal of Hydrologic Engineering 5 (2), 115–123.

Güler, C., Thyne, G.D., McCray, J.E., Turner, A.K., 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeology Journal 10 (4), 455–474.

Ha, H., Stenstrom, M.K., 2003. Identification of land use with water quality data in stormwater using a neural network. Water Research 37 (17), 4222–4230.

Hanh, P.T.M., Sthiannopkao, S., Ba, D.T., Kim, K., 2011. Development of water quality indexes to identify pollutants in Vietnam's surface water. Journal of Environmental Engineering 137 (4), 273–283.

Hill, T., Marquez, L., O'Connor, M., Remus, W., 1994. Artificial neural network models for forecasting and decision making. International Journal of Forecasting 10 (1), 5–15.

Juahir, H., Zain, S.M., Toriman, M.E., Mokhtar, M., Man, H.S., 2004. Application of artificial neural network models for predicting water quality index. Jurnal Kejuruteraan Awam 16 (2), 42–55.

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educational and Psychological Measurement 20 (1), 141–151.

Kaiser, H.F., 1974. An index of factorial simplicity. Psychometrika 39 (1), 31–36.

Kaurish, F.W., Younos, T., 2007. Developing a standardized water quality index for evaluating surface water quality. Journal of the American Water Resources Association 43 (2), 533–545.

Khalil, B., Ouarda, T.B.M.J., St-Hilaire, A., 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. Journal of Hydrology 405 (3–4), 277–287.

Khuan, L.Y., Hamzah, N., Jailani, R., 2002. Prediction of Water Quality Index (WQI) Based on Artificial Neural Network (ANN). In: Proceedings of the Student Conference on Research and Development, Shah Alam, Malaysia.

Kovalishyn, V.V., Tetko, I.V., Luik, A.I., Kholodovych, V.V., Villa, A.E.P., Livingstone, D.J., 1998. Neural network studies. 3. variable selection in the cascade-correlation learning architecture. Journal of Chemical Information and Computer Sciences 38 (4), 651–659.

Kowalkowski, T., Zbytniewski, R., Szpejna, J., Buszewski, B., 2006. Application of chemometrics in river water classification. Water Research 40 (4), 744–752.

Liao, K., Fildes, R., 2005. The accuracy of a procedural approach to specifying feedforward neural networks for forecasting. Computers and Operations Research 32 (8), 2151–2169.

Liou, S.M., Lo, S.L., Wang, S.H., 2004. A generalized water quality index for Taiwan. Environmental Monitoring and Assessment 96, 32–35.

Lischeid, G., 2001. Investigating short-term dynamics and long-term trends of SO_4 in the runoff of a forested catchment using artificial neural networks. Journal of Hydrology 243 (1–2), 31–42.

Lucio, P.S., Conde, F.C., Cavalcanti, I.F.A., Serrano, A.I., Ramos, A.M., Cardoso, A.O., 2007. Spatiotemporal monthly rainfall reconstruction via artificial neural network – case study: south of Brazil. Advances in Geosciences 10, 67–76.

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modeling and Software 15 (1), 101–124.

Mas, D.M.L., Ahlfeld, D.P., 2007. Comparing artificial neural networks and regression models for predicting faecal coliform concentrations. Hydrological Science 52 (4), 713–731.

May, D.B., Sivakumar, M., 2009. Prediction of urban stormwater quality using artificial neural networks. Environmental Modeling and Software 24 (2), 296–302.

McNeil, V.H., Cox, M.E., Preda, M., 2005. Assessment of chemical water types and their spatial variation using multi-stage cluster analysis, Queensland Australia. Journal of Hydrology 310 (1–4), 181–200.

Miao, Y., Mulla, D.J., Robert, P.C., 2006. Identifying important factors influencing corn yield and grain quality variability using artificial neural networks. Precision Agriculture 7 (2), 117–135.

Norhayati, M.T., Goh, S.H., Tong, S.L., Wang, C.W., Abdul Halim, S., 1997. Water quality studies for the classification of Sungai Bernam and Sungai Selangor. Journal of Ensearch 10, 27–36.

Olszewski, T., Ryniecki, A., Boniecki, P., 2008. Neural network development for automatic identification of the endpoint of drying barley in bulk. Journal of Research and Applications in Agricultural Engineering 53 (1), 26–31.

Özesmi, S.L., Tan, C.O., Özesmi, U., 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecological Modelling 195 (1–2), 83–93.

Palani, S., Liong, S., Tkalich, P., 2008. An ANN application for water quality forecasting. Marine Pollution Bulletin 56 (9), 1586–1597.

Pastor-Bárcenas, O., Soria-Olivas, E., Martín-Guerrero, J.D., Camps-Valls, G., Carrasco-Rodriguez, J.L., del Valle-Tascón, S., 2005. Unbiased sensitivity analysis and pruning techniques in neural networks for surface ozone modelling. Ecological Modelling 182 (2), 149–158.

Pesce, S.F., Wunderlin, D.A., 2000. Use of water quality indices to verify the impact of Cordóba City (Argentina) on Suquía River. Water Research 34 (11), 2915–2926.

Prechelt, L., 1998. Automatic early stopping using cross validation. Quantifying the Criteria. Neural Networks 11 (4), 761–767.

Qi, M., Zhang, G.P., 2001. An investigation of model selection criteria for neural network time series forecasting. European Journal of Operational Research 132 (3), 666–680.

Reimann, C., Filzmoser, P., Garrett, R.G., 2002. Factor analysis applied to regional geochemical data: problems and possibilities. Applied Geochemistry 17 (3), 185–206.

Riad, S., Mania, J., Bouchaou, L., Najjar, Y., 2004. Rainfall-runoff model using an artificial neural network approach. Mathematical and Computer Modelling 40 (7–8), 839–846.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back propagating errors. Nature 323 (6088), 533–538.

Sánchez, E., Colmenarejo, M.F., Vicente, J., Rubio, A., García, M.G., Travieso, L., et al., 2007. Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. Ecological Indicators 7 (2), 315–328.

Sargaonkar, A., Deshpande, V., 2003. Development of an overall index of pollution for surface water based on a general classification scheme in indian context. Environmental Monitoring and Assessment 89 (1), 43–67.

Schenker, B., Agarwal, M., 1996. Cross-validated structure selection for neural networks. Computers and Chemical Engineering 20 (2), 175–186.

Shamseldin, A.Y., Nasr, A.E., O'Connor, K.M., 2002. Comparison of different forms of the multi-layer feed-forward neural network method used for river flow forecast combination. Hydrology and Earth System Sciences 6 (4), 671–684.

Shanthi, D., Sahoo, G., Saravanan, N., 2009. Comparison of neural network training algorithms for the prediction of the patient's post-operative recovery area. Journal of Convergence Information Technology 4 (1), 24–32.

Singh, K.P., Basant, A., Malik, A., Jain, G., 2009. Artificial neural network modeling of the river water quality. A case study. Ecological Modeling 220 (6), 888–895.

Srinivasan, D., Liew, A.C., Chang, C.S., 1994. A neural network short-term load forecaster. Electric Power Systems Research 28 (3), 227–234.

Song, T., Kim, K., 2009. Development of a water quality loading index based on water quality modeling. Journal of Environmental Management 90 (3), 1534–1543.

Teschl, R., Randeu, W.L., 2006. A neural network model for short term river flow prediction. Natural Hazards and Earth System Sciences 6 (4), 629–635.

Tiron, G., Gosav, S., 2010. The July 2008 rainfall estimation from Barnova WSR-98 D radar using artificial neural network. Romanian Reports in Physics 62 (2), 405–413.

Turney, P., 1993. A theory of cross-validation error. Journal of Experimental and Theoretical Artificial Intelligence 6 (4), 361–391.

Twomey, J.M., Smith, A.E., 1998. Bias and variance of validation methods for function approximation neural networks under conditions of sparse data. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 28 (3), 417–430.

Wang, H., Wang, S., 2010. Mining incomplete survey data through classification. Knowledge and Information Systems 24 (2), 221–233.

Wilamowski, B.M., 2003. Neural Network Architectures and Learning. In: Proceedings of the IEEE International Conference on Industrial Technology. 10–12 December 2003 (pp. TU1–T12 Vol. 1). Maribor: IEEE Industrial Electronic Society, University of Maribor.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bulletin American Meteorological Society 63 (11), 1309–1313.

Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research 30 (1), 79–82.

Wu, C.L., Chau, K.W., 2011. Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. Journal of Hydrology 399 (3–4), 394–409.

Yu, X., Efe, M.O., Kaynak, O., 2002. A general backpropagation algorithm for feedforward neural networks learning. IEEE Transactions on Neural Networks 13 (1), 251–254.

Zandbergen, P.A., Hall, K.G., 1998. Analysis of the British Columbia water quality index for watershed managers: a case study of two small watersheds. Water Quality Research Journal of Canada 33 (4), 519–549.

Zeghal, M., Khogali, W.E.I., 2007. On the use of artificial neural networks as quality control tool. In: Proceedings of the CANCAM 2007. NRC Institute for Research in Construction; National Research Council Canada, Toronto, Ontario, Canada, pp. 594–595.