Marine Pollution Bulletin 98 (2015) 171–178

Contents lists available at ScienceDirect

Marine Pollution Bulletin

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

Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean

Mohamad Javad Alizadeh *, Mohamad Reza Kavianpour

Faculty of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran

A R T I C L E   I N F O

Article history:
Received 18 March 2015
Revised 26 June 2015
Accepted 28 June 2015
Available online 2 July 2015

Keywords:
Water quality
Neural networks
Wavelet transform
Daily prediction
Ocean parameters

A B S T R A C T

The main objective of this study is to apply artificial neural network (ANN) and wavelet-neural network (WNN) models for predicting a variety of ocean water quality parameters. In this regard, several water quality parameters in Hilo Bay, Pacific Ocean, are taken under consideration. Different combinations of water quality parameters are applied as input variables to predict daily values of salinity, temperature and DO as well as hourly values of DO. The results demonstrate that the WNN models are superior to the ANN models. Also, the hourly models developed for DO prediction outperform the daily models of DO. For the daily models, the most accurate model has R equal to 0.96, while for the hourly model it reaches up to 0.98. Overall, the results show the ability of the model to monitor the ocean parameters, in condition with missing data, or when regular measurement and monitoring are impossible.

1. Introduction

Ocean water quality parameters play an important role in aquatic life. A relief map of the ocean bottom reveals two major areas: the continental shelf and the deep oceans. The continental shelf, especially near major estuaries, is the most productive in terms of food supply. Because of its proximity to human activity, it receives the greatest pollution load. Many estuaries have become so badly polluted that they are closed to commercial fishing (Weiner and Matthews, 2003). Therefore, assessments of water quality parameters in oceans especially in estuaries are of great importance.

Water quality covers a wide range of physical, chemical, and biological parameters such as dissolved oxygen (DO), temperature, electrical conductivity (EC), salinity, turbidity, alkalinity, ammonia, total dissolved solids (TDS), nitrate, sulfate and phosphate. Some of these parameters are mutually interrelated. However, only a few of them are taken under consideration in this study. DO which refers to the level of free, non-compound oxygen in water or other liquids, is an important parameter in assessing water quality because it is essential to many forms of life including fish, invertebrates, bacteria and plants. Prediction of such parameter for one or more upcoming time steps in an ecosystem is one of the challenging issues. Since a large number of factors affecting water quality that are mainly complicated and there is a non-linear relationship among the variables, the traditional methods which are based on linear relationships are not good enough for solving these types of problems (Ravansalar et al., 2015). Moreover, in oceans and estuaries, the presence of tidal waves, wind induced wave actions and other oceanic phenomena increase the complexity of the problem.

Over the past few years, artificial intelligence techniques have been frequently used to predict the nonlinear time series and achieved good results (Kisi, 2008; Nourani et al., 2011). Recently, wavelet transform, which is a data-preprocessing technique, has shown excellent performance in hydrological modelling due to its ability to analyze a signal in both time and frequency (Okkan, 2012). This approach overcomes the basic drawbacks of conventional Fourier transform. Nourani et al. (2009), showed that the wavelet transform provided effective decompositions of time series so that decomposed data increased the performance of hydrological prediction model by capturing useful information on different resolution level. Hence a wavelet neural network model which uses multi-scale signals as input data can present more suitable prediction performance rather than a single pattern input (Alizadeh et al., 2015; Nourani et al., 2009; Okkan, 2012; Rajaee et al., 2010). Generally, using soft computing techniques such as ANN, ANFIS and WNN has the potential to reduce the computation time and effort and the possibility of errors in the calculation.

Gazzaz et al. (2012) developed an ANN model for the prediction of water quality index for Kinta River (Malaysia). They applied more than 20 monitored parameters for their model development.

* Corresponding author.
E-mail addresses: mjalizadeh@mail.kntu.ac.ir (M.J. Alizadeh), kavianpour@kntu.ac.ir (M.R. Kavianpour).

http://dx.doi.org/10.1016/j.marpolbul.2015.06.052
0025-326X/

The results demonstrated that the proposed model provides an appropriate prediction of water quality index. The ASCE task committee presented different studies on water quality using the ANNs (Govindaraju, 2000). Bahaa et al. (2012) examined ability of ANN models for monthly COD (chemical oxygen demand) concentration prediction in Nile Delta. Tryland et al. (2014) investigated the impact of rainfall on the hygienic quality of blue mussels and water in urban areas in the Inner Oslofjord, Norway. Wu et al. (2012) applied a fuzzy integrated assessment method to evaluate the ecological quality status of coastal waters in East China Sea. Palani et al. (2011) applied an artificial neural network (ANN) model to estimate atmospheric deposition concentration of total nitrogen (TN) and organic nitrogen (ON) concentrations in coastal aquatic ecosystems. The selected model input variables were nitrogen species from atmospheric deposition, total suspended particulates, pollutant standards index and meteorological parameters. They compared the results of the ANN model with the linear regression model. The results showed that the ANN model is relatively more accurate in its predictions. A number of studies already have been carried out regarding the application of the ANN models for the prediction of water quality variables in rivers and seas (Dogan et al., 2009; Faruk, 2010; Heydari et al., 2013; Musavi and Golabi, 2008; Palani et al., 2009; Singh et al., 2009). However, further explorations are needed in order to increase the accuracy of previous models and also to predict water quality parameters in oceans and bays.

The main objective of this study is to examine the ability and applicability of ANN and WNN models to predict daily values of salinity, temperature and DO based on the existing measured data of chlorophyll, DO, salinity, turbidity and water temperature. Moreover, separate ANN and WNN models are developed to predict hourly values of DO. Many different combinations of water quality variables with different time lags have been used in the input structure of the ANN and WNN models. In the WNN models, different mother wavelets as well as different wavelet decomposition levels are examined to get the best model performances. To evaluate the model performances, correlation coefficient (R) and root mean square error (RMSE) are computed for each model.

2. Materials and methods

2.1. Artificial neural network (ANN)

A usual ANN consists of three layers, input, output and hidden layer. Different back propagation algorithms can be employed to train an ANN model. In this study, the Levenberg–Marquardt back propagation algorithm, which is a simplified version of the Newton method, has been applied for training of the ANN. This algorithm is a second-order nonlinear optimization technique that is usually faster and more reliable than any other back propagation technique (Hagan and Menhaj, 1994; Ham and Kostanic, 2001). The training process can be viewed as finding a set of weights that minimize the error (e_p) for all samples in the training set (T). The performance function is a sum of squares of the errors as follows (Ham and Kostanic, 2001):

$$E(w) = \frac{1}{2}\sum_{p=1}^{p}(d_p - y_p)^2 = \frac{1}{2}\sum_{p=1}^{p}(e_p)^2, p = mT \tag{1}$$

where T is the total number of training samples, m is the number of output layer neurons, w represents the vector containing all the weights in the network, y_p is the actual network output, and d_p is the desired output, When training with the Levenberg–Marquardt algorithm, the changing of weights Δw can be computed as follows (Ham and Kostanic, 2001):

$$\Delta w_k = -[J_k^T J_k + \mu_k I]^{-1} J_k^T e_k \tag{2}$$

Then, the update of the weights can be adjusted as follows:

$$w_{k+1} = w_k + \Delta w_k \tag{3}$$

where J is the Jacobian matrix, I is the identify matrix, e is the network error, μ is the Marquardt parameter which is to be updated using the decay rate β depending on the outcome. In particular, μ is multiplied by the decay rate $\beta(0 < \beta < 1)$whenever $E(w)$ decreases, while μ is divided by β whenever $E(w)$ increases in a new step (Ham and Kostanic, 2001).

The optimum number of neurons in the hidden layer and training iteration number are important procedures in ANN models. There is no specific algorithm to determine the optimum number neurons in the hidden layer and it should be found based on a trial-and-error procedure.

2.2. Discrete wavelet transform (DWT)

Wavelet transforms are divided to continuous wavelet transform (CWT) and discrete wavelet transform (DWT). For practical applications in hydrology, researchers have access to a discrete time signal rather than to a continuous time signal (Alizdeh et al., 2015; Rajaee et al., 2010). By using DWT, the original time series will be decomposed into various sub-signals of details (d_i) and approximations (a_i) at different resolution levels. The approximations are the high-scale, low frequency components of the signal and the details are the low-scale, high frequency components (Zhang et al., 2014).

The current study will not delve into the theory behind wavelet transform (WT) and only the main concepts of the discrete wavelet transform (DWT) will be briefly presented. A mathematical overview of WT and a review of applications have been presented by (Labat et al., 2000). The WT performs the decomposition of a signal into a group of functions (Cohen and Kovacevic, 1996):

$$\psi_{j,k}(x) = 2^{j/2}\psi_{j,k}(2^j x - k) \tag{4}$$

where $\psi_{j,k}(x)$ is produced from a mother wavelet $\psi(x)$ which is dilated by j and translated by k. The mother wavelet has to satisfy the condition.

$$\int \psi(x)dx = 0 \tag{5}$$

The discrete wavelet function of a signal $f(x)$ can be calculated as follows:

$$c_{j,k} = \int_{-\infty}^{\infty} f(x)\psi_{j,k}^{*}(x)dx \tag{6}$$

$$f(x) = \sum_{j,k} c_{j,k}\psi_{j,k}(x) \tag{7}$$

where $c_{j,k}$ is the approximate coefficient of a signal. The mother wavelet is formulated from the scaling function $\varphi(x)$ as follows:

$$\varphi(x) = \sqrt{2}\sum h_0(n)\varphi(2x - n) \tag{8}$$

$$\psi(x) = \sqrt{2}\sum h_1(n)\varphi(2x - n) \tag{9}$$

where $h_1(n) = (-1)nh_0(1-n)$. Different sets of coefficients $h_0(n)$ can be found corresponding to wavelet bases with various characteristics. In the DWT, coefficients $h_0(n)$ play a critical role (Gupta and Gupta, 2007). The number of wavelet decomposition levels and wavelet mother wavelet applied in DWT affect WNN model performances. Therefore, to achieve the best model performance, an appropriate number of mother wavelet as well as an optimum number of wavelet decomposition levels should be taken into account.

2.3. Water quality parameters

Regarding the water quality parameters in rivers, there are many important variables which are usually interrelated. Some of the most important parameters in determination of water quality consists of; temperature, turbidity, EC, salinity, pH, the concentrations of suspended solids (SS), dissolved solids (DS), total solids (TS), ammonia nitrogen (NH3-N), dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), sodium (Na), potassium (K), calcium (Ca), magnesium (Mg), nitrate nitrogen (NO3-N), chloride (Cl), phosphate phosphorous (PO4-P). In addition to the mentioned parameters, knowing the amount of chlorophyll can be helpful in predicting some of the parameters, especially DO.

Prediction of water quality in bays, where fresh river water meets saline water of the sea and the ocean is more complicated. Oceans are more complex due to the presence of parameters such as wind-induced waves, tidal waves, and some other complex phenomena which happen over there. Therefore, in ocean water quality modelling, significant wave height, tidal waves characteristics, pressure fluctuations can enhance the predictive model accuracy. In this study, an attempt was made to deal with a limited number of parameters. These parameters include salinity, turbidity, DO, temperature and the amount of chlorophyll. Clearly, dealing with fewer variables has the advantage of simplicity and also being time and cost-effective. Sampling for a large number of parameters is time consuming and expensive. Moreover, applying these data for the model will increase the complexity of the model.

2.4. Gauging station and data analysis

Data used in this study includes daily and hourly values of water temperature, DO, chlorophyll, salinity, turbidity. The data are related to gauging station of WQB-04 which is located in Hilo Bay on the east side of the Big Island (−155.09 to −155.08 longitude and 19.73–19.74 latitude). These data consist of hourly and daily values of the above-mentioned parameters from Oct. 2010 to Oct. 2014. These values were measured in the depth of 1 meter below the mean surface level ($z = -1$ m). The water quality buoys (WQB) are part of the Pacific Islands Ocean Observing System (PacIOOS) and are designed to measure a variety of ocean parameters at fixed points. Continuous sampling of this outflow area provides a record of baseline conditions of the chemical and biological environment for comparison when there are pollution events such as storm runoff or a sewage spill. PacIOOS is one of the eleven regional observing programs in the U.S. which is supporting the emergence of the U.S. Integrated Ocean Observing System (**IOOS**®) under the National Oceanographic Partnership Program (**NOPP**). Data have been downloaded from the web server of (http://oos.soest.hawaii.edu/dchart/). Fig. 1 illustrates the gauging station located in the Hilo Bay.

It should be mentioned that in daily predictive models, the value of water temperature, turbidity, salinity, DO and the amount of chlorophyll measured at 12 o'clock were considered as daily values of the parameters. The number of water samples selected for the daily and hourly model developments are 650 and 540 respectively. These samples were divided into training (70%), validation (15%), and testing sets (15%). The required basic statistical measures (average 'Mean', standard deviation 'Sd', skewness coefficient 'Csx', minimum 'Min' and maximum 'Max' of the daily datasets are given in Table 1.

To bring all the data in the range [0, 1], data normalization has been carried out based on the following equation:

$$X'_i = \frac{X_{max} - X_i}{X_{max} - X_{min}} \tag{10}$$

where X'_i is the normalized value, X_i is the original value and X_{max} and X_{min} are the minimum and maximum of variables respectively.

2.5. Model evaluation

In this study, the performance of the models is evaluated by the indexes of the correlation coefficient (R) and root mean squared error (RMSE). In brief, the models' predictions are optimum if R and RMSE are found to be close to 1 and 0, respectively. Correlation coefficient (R) measures the strength and direction of the linear relationship between two variables that is defined as the (sample) covariance of the variables divided by the product of their (sample) standard deviations. RMSE indicates the discrepancy between the observed and forecasted values. A perfect fit between observed and forecasted values would have an RMSE of 0.

These indexes are defined as follows:

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_{i(measured)} - y_{i(predicted)})^2}{\sum_{i=1}^{n}(y_{i(measured)} - y_{i(mean)})^2}} \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_{i(measured)} - y_{i(predicted)})^2}{n}} \tag{12}$$

where n is the number of data, and y denotes the output variable.

2.6. Model development

Models were generally developed in daily and hourly time scales. For both daily and hourly predictions of water quality parameters, ANN and WNN models have been applied. In the WNN models, the decomposed time series obtained by the DWT are used in the input structure of ANN models while in the ANN models, the original time series (not decomposed) are used as inputs. For all the WNN models applied in this study, it has been found that 3 is the most efficient wavelet decomposition level. The examination of different types of wavelets including DMeyer (dmey), Daubechies (db2) and Coiflets (coif4) for the WNN models revealed that the 'coif4' is the most efficient type of wavelet for this study. The abovementioned mother wavelets are discrete wavelets in which each one has a specific property and function.

2.6.1. Models for daily prediction

The daily ANN and WNN models were aimed to predict daily values of DO, water temperature and salinity. In this regard, different models based on different combinations of water quality variables and with different time lags were considered as input variables for the ANN and WNN models. Details regarding the input and output variables for both ANN and WNN models are given in Table 2. It should be noticed that the terms '$t, \ldots, t-4$' for daily and hourly models represent the amount of the parameters at the present time up to 4 days and hours back in time respectively. In Table 2, terms Ch, S, and Tur represents chlorophyll, salinity and turbidity respectively.

2.6.2. Models for hourly prediction

Hourly ANN and WNN models were developed to predict hourly values of DO in Hilo Bay. Similar to the daily models, different combinations of water quality variables with different time lags were considered as input variables to predict DO at time t (output). For each model, the optimum number of training iteration and neurons in the hidden layer of the ANN model was achieved by trial and error. The input and output variables applied for hourly

Fig. 1. Gauging station in Hilo Bay.

Table 1
Data statistical analysis.

Parameters	Mean	Sd	Csx	Min	Max
DO (kg/m^3)	7.21e−3	8.54e−4	−2.455	6.5e−4	9.53e−3
Salinity (psu)	29.75	4.16	−1.49	11.82	37
Temperature (C)	24.92	1.37	−0.707	19.42	28.3
Turbidity (ntu)	1.44	1.54	2.364	0	9.8
Chlorophyll (kg/m^3)	7.38e−6	9.18e−6	2.448	3e−8	5.93e−5

models have the same components as daily models No. 1 to 10. A schematic layout for a WNN model with decomposition level of 3 and for 1 input variable is depicted in Fig. 2. In the Fig. 2, a_i and d_i represent sub-signals related to the details and approximations of the original input respectively.

3. Results and discussion

This study employed two soft computing techniques including the ANN and WNN models to predict a variety of ocean water quality parameters in Hilo Bay, Pacific Ocean. To examine the ability and applicability of both ANN and WNN models, daily predictive models for DO, water temperature and salinity as well as hourly predictive models for DO concentration have been taken under consideration. Based on the input structure of the ANN and WNN models (Table 2), three distinguished patterns can be recognized. First, models which predict the values of the parameter at time t using the values of the other parameters in the input structure at time t. Second, Applying the values of the same parameter with different time lags (up to for previous days or hours) to predict it at time t. Third, having a conjunction model by using both preceding time series of the parameter and the values of other parameters at time t. Results related to the daily and hourly models are presented in the following subsections.

3.1. Models for daily prediction

Results related to the best models of each pattern for the prediction of daily values of DO, water temperature and salinity are presented in Tables 3 and 4. Table 3 presents the values of R and RMSE for the ANN models. In a similar way, Table 4 presents the values of R and RMSE for the best WNN models. The input and output variables related to each model No. can be observed in Table 2.

According to Table 3, it can be found that an accurate prediction of the water quality parameters is not achieved when the target values of previous time steps are excluded in the input variables. Low values of R for models No. 1, 11, and 21 demonstrate this claim. This conclusion can be drawn for all of the three considered parameters (DO, water temperature and salinity). Using only the same parameter with different time lags as input variables to

Table 2
Characteristics of the developed models.

Model No.	Input variables	Output	ANN/WNN
1	Ch(t), S(t), T(t), Tur(t)	DO(t)	ANN
2	DO($t-1$)	DO(t)	ANN
3	DO($t-1$), DO($t-2$)	DO(t)	ANN
4	DO($t-1$), DO($t-2$), DO($t-3$)	DO(t)	ANN
5	DO($t-1$), DO($t-2$), DO($t-3$), DO($t-4$)	DO(t)	ANN
6	Ch(t), S(t), T(t), Tur(t), DO($t-1$)	DO(t)	ANN
7	Ch(t), S(t), T(t), Tur(t), DO($t-1$)), DO($t-2$), DO($t-3$)	DO(t)	ANN
8	Ch(t), S(t), T(t), Tur(t)	DO(t)	WNN
9	DO($t-1$), DO($t-2$), DO(t-3)	DO(t)	WNN
10	Ch(t), S(t), T(t), Tur(t), DO($t-1$)), DO($t-2$), DO($t-3$)	DO(t)	WNN
11	Ch(t), S(t), DO(t), Tur(t)	T(t)	ANN
12	T($t-1$)	T(t)	ANN
13	T($t-1$), T($t-2$)	T(t)	ANN
14	T($t-1$), T($t-2$), T($t-3$)	T(t)	ANN
15	T($t-1$), T($t-2$), T($t-3$), T($t-4$)	T(t)	ANN
16	Ch(t), S(t), DO(t), Tur(t), T($t-1$)	T(t)	ANN
17	Ch(t), S(t), DO(t), Tur(t), T($t-1$), T($t-2$), T($t-3$)	T(t)	ANN
18	Ch(t), S(t), DO(t), Tur(t)	T(t)	WNN
19	T($t-1$), T($t-2$), T($t-3$)	T(t)	WNN
20	Ch(t), S(t), DO(t), Tur(t), T($t-1$), T($t-2$), T($t-3$)	T(t)	WNN
21	Ch(t), T(t), DO(t), Tur(t)	S(t)	ANN
22	S($t-1$)	S(t)	ANN
23	S($t-1$), S($t-2$)	S(t)	ANN
24	S($t-1$), S($t-2$), S($t-3$)	S(t)	ANN
25	S($t-1$), S($t-2$), S($t-3$), S($t-4$)	S(t)	ANN
26	Ch(t), T (t), DO(t), Tur(t), S($t-1$)	S(t)	ANN
27	Ch(t), T (t), DO(t), Tur(t), S($t-1$), S($t-2$), S($t-3$)	S(t)	ANN
28	Ch(t), T (t), DO(t), Tur(t)	S(t)	WNN
29	S($t-1$), S($t-2$), S($t-3$)	S(t)	WNN
30	Ch(t), T (t), DO(t), Tur(t), S(t-1), S($t-2$), S($t-3$)	S(t)	WNN

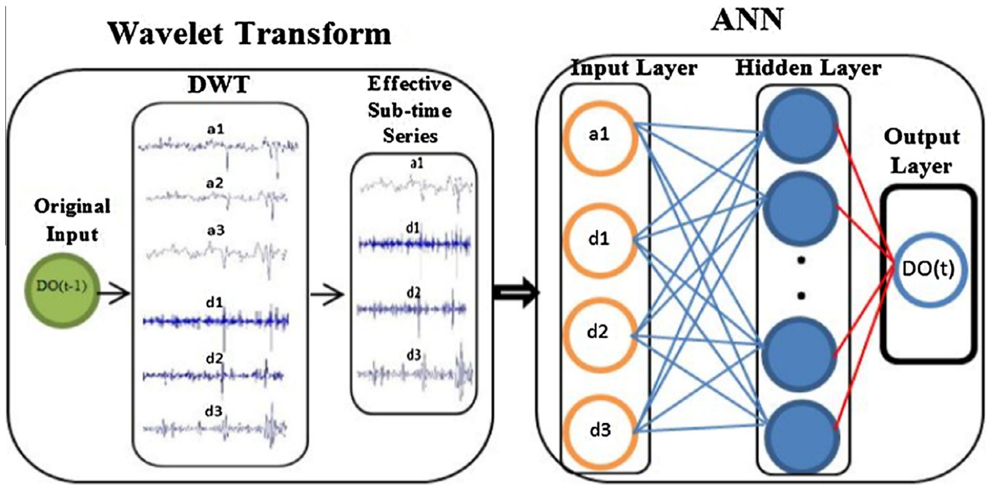

Fig. 2. Schematic of the WNN model for model No. 2.

Table 3
Results related to R and RMSE for daily ANN models.

Model No.	Training		Validation		Testing	
	R	RMSE	R	RMSE	R	RMSE
1	0.52	6.04e−4	0.54	7.42e−4	0.22	1.3e−3
2	0.64	5.57e−4	0.72	5.88e−4	0.71	1e−3
3	0.73	4.86e−4	0.66	6.26e−4	0.32	1.5e−3
4	0.65	5.34e−4	0.71	5.82e−4	0.72	9.31e−4
5	0.63	5.463e−4	0.72	5.682e−4	0.70	9.61e−4
6	0.76	4.54e−4	0.77	5.31e−4	0.74	8.93e−4
7	0.7	5.34e−4	0.8	5.56e−4	0.63	1.1e−3
11	0.6	0.12	0.62	0.08	0.41	1.24
12	0.84	0.08	0.67	0.07	0.67	0.74
13	0.85	0.07	0.68	0.07	0.69	0.75
14	0.85	0.07	0.68	0.07	0.7	0.70
15	0.85	0.07	0.68	0.07	0.69	0.73
16	0.82	0.08	0.75	0.06	0.70	0.85
17	0.88	0.07	0.78	0.06	0.76	0.73
21	0.28	4.1	0.33	4.48	−0.13	3.45
22	0.66	2.98	0.67	3.45	0.64	3.12
23	0.31	3.81	0.39	4.1	0.21	3.95
24	0.66	2.98	0.67	3.68	0.65	3.10
25	0.66	2.97	0.66	3.72	0.65	3.11
26	0.75	2.6	0.78	3.25	0.78	2.71
27	0.75	2.58	0.78	3.2	0.80	2.68

predict the parameter at time t has better performances rather than the models using other parameters at time t to predict the parameter which is excluded in the input structure of the model. Comparison of the model No. 1 with 6, 11 with 16 and 21 with 26 reveals that including the target variable with different time lags in the input structure improves the model performance significantly for the prediction of all the three considered parameters. For example, by adding DO values at time $t-1$ to the input variables of the model No. 1, the R value of testing period increases from 0.22 to 0.74 (more than 200%) and the RMSE decreases from 0.0013 to 0.0008 (model No. 6). This is also the case for the prediction of daily values of water temperature and salinity. Anyhow, it can be observed that the ANN model does not have enough accuracy for the prediction of daily values of the water quality parameters for the study. Moreover, the difference between R values for training and testing are significant in many cases, which is considered to be a weakness for the model. The same analysis can be drawn for RMSE values. Therefore, some more advanced models such as WNN models should be investigated.

Table 4
Results related to R and RMSE for daily WNN models.

Model No.	Training		Validation		Testing	
	R	RMSE	R	RMSE	R	RMSE
8	0.36	6.78e−4	0.51	7.73e−4	0.27	1.3e−3
9	0.96	1.83e−4	0.96	2.09e−4	0.96	3.88e−4
10	0.91	2.86e−4	0.92	3.31e−4	0.84	6.22e−4
18	0.59	0.12	0.58	0.11	0.61	0.19
19	0.98	0.02	0.96	0.02	0.95	0.34
20	0.99	0.021	0.97	0.024	0.95	0.24
28	0.79	0.09	0.81	0.13	−0.18	3.66
29	0.96	0.04	0.96	0.05	0.95	1.09
30	0.96	0.04	0.96	0.05	0.92	1.23

To improve the efficiency of the existing ANN models, DWT has been used to decompose the original time series (input variables). Then, the effective sub-time series were imposed as new input variables for the ANN models. Results related to the best WNN models including all the three patterns for the input structure and for predicting of the three parameters of DO, water temperature and salinity are given in Table 4. Details related to the input structure and output variable presented in Table 2.

Regarding Table 4, it is implied that applying WNN models provides an accurate prediction of the water quality parameters for the area under consideration. The best WNN model for predicting each parameter can be achieved when the values of the target variable up to 3 days ahead ($t-1$, $t-2$, and $t-3$) are used in the input structure (all the three predicted variables had the same pattern in the input structure, i.e. second pattern). The best WNN model for the prediction of DO, water temperature and salinity has R value of 0.964, 0.912, and 0.967 respectively. Moreover, the WNN models which have the other parameters as well as the target variable with a time lag can provide an appropriate prediction of the target variable at the present day. On the other hand, the low values of R and great values of RMSE for models No. 8, 18, and 28 demonstrate that excluding the target variable in the input structure and using only the other parameters does not give an appropriate prediction for the water quality variables of this study.

Comparison of the results presented in Tables 3 and 4 indicates that the WNN models outperform the ANN models. Applying the WNN models instead of the ANN models during testing period make an improvement in the R values by 32%, 35%, and 47% for predicting of DO, water temperature and salinity respectively. For testing period, predicted values versus observed values of DO, water temperature and salinity are illustrated in Figs. 3–5 respectively.

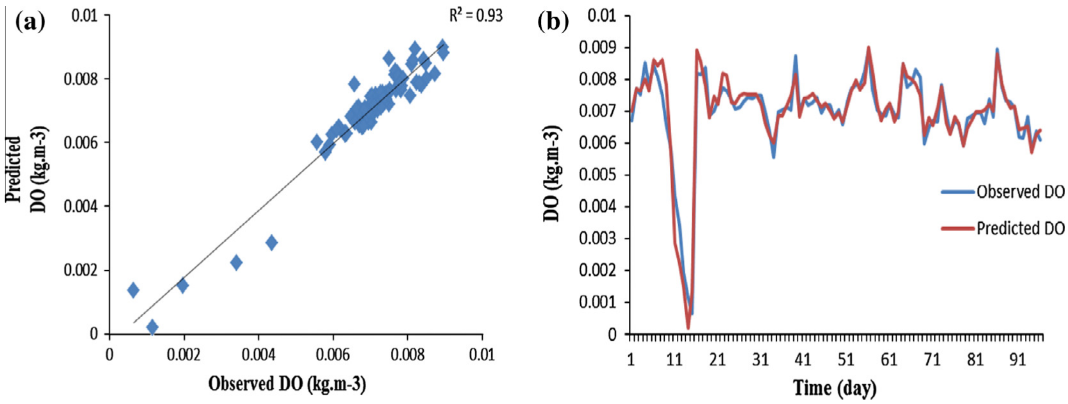

Fig. 3. (a) Scatter plot of the observed versus the predicted DO values and (b) Comparison of the predicted and the measured DO for the testing set of the WNN model (model No. 9).

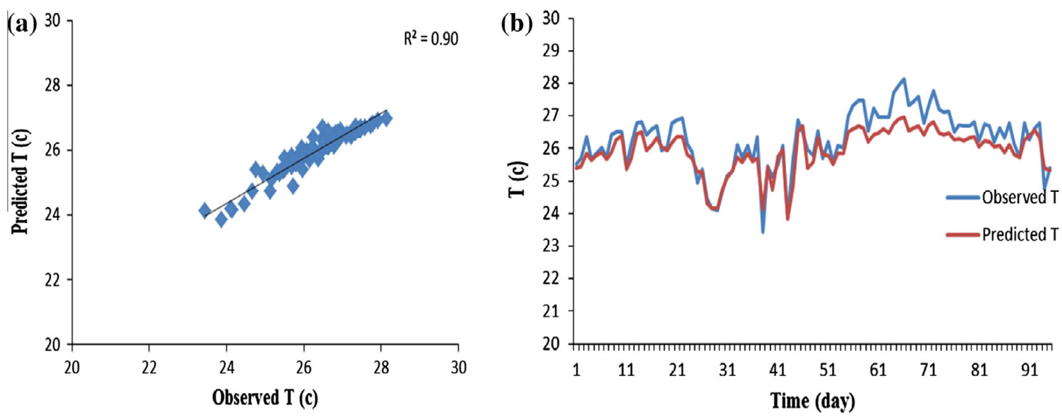

Fig. 4. (a) Scatter plot of the observed versus the predicted temperature values and (b) Comparison of the predicted and the measured temperature for the testing set of the WNN model (model No. 19).

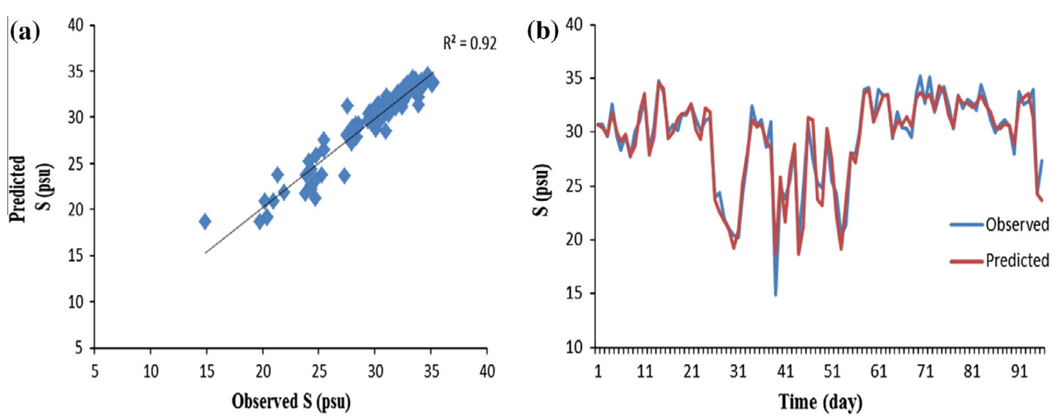

Fig. 5. (a) Scatter plot of the observed versus the predicted salinity values and (b) Comparison of the predicted and the measured salinity for the testing set of the WNN model (model No. 29).

These figures. demonstrate the efficiency of the wavelet-ANN models for the prediction of the ocean water quality parameters. According to Figs. 3a, 4a and 5a. a very high correlation between observed and predicted values of DO, Temperature and salinity is observed. High values of R^2 (greater than 0.9) indicate the high accuracy of the developed models. Fig. 3b. shows that the WNN model has a great ability to predict DO values. Moreover, the proposed model gives an accurate prediction for the extreme and minimum values of DO. For the water temperature prediction, the WNN model provides a bit smaller values in comparison with the actual values. Anyway, a relatively good correlation is observed between the measured and predicted values (Fig. 4b). Regarding Fig. 5b, it can be seen that predicted and observed values overlap each other very well. Overall, the proposed models provide a good prediction for all the three considered water quality parameters. Among three variables, the WNN models for DO and temperature have the most and least accuracy respectively.

Table 5
Results related to R and RMSE for hourly ANN models.

Model No.	Training		Validation		Testing	
	R	RMSE	R	RMSE	R	RMSE
1	0.96	1.58e−4	0.73	3.39e−4	0.87	3.38e−4
2	0.94	1.88e−4	0.94	1.81e−4	0.92	2.04e−4
3	0.94	1.87e−4	0.94	1.85e−4	0.92	2.05e−4
4	0.95	1.7e−4	0.94	1.81e−4	0.94	1.85e−3
5	0.95	1.69e−4	0.94	1.81e−4	0.93	1.87e−4
6	0.94	1.86e−4	0.94	1.88e−4	0.93	2.01e−4
7	0.92	1.82e−4	0.89	2.68e−4	0.94	2.02e−4

Table 6
Results related to R and RMSE during testing period and for hourly WNN models.

Model No.	Training		Validation		Testing	
	R	RMSE	R	RMSE	R	RMSE
8	0.94	1.7e−4	0.95	1.64e−4	0.92	2.08e−4
9	0.99	8.5e−5	0.99	8.61e−5	0.98	1.03e−4
10	0.99	1.35e−4	0.98	1.32e−4	0.97	1.53e−4

3.2. Models for hourly prediction

In this part, ANN and WNN models have been developed to predict the hourly values of DO concentration. The input structure of the ANN models applied for hourly prediction of DO can be divided in three categories; (1) using amount of chlorophyll, water temperature, turbidity and salinity in time t (2) using values of DO up to 4 days ahead ($t-1, \ldots, t-4$) and (3) applying a combination of the amount of chlorophyll, turbidity, salinity, water temperature in time t and DO values with 1 to 3 time lags. The results related the ANN models during testing period are given in Table 5.

As illustrated in Table 5, the ANN models can be successfully applied in order to predict hourly values of DO. Models No. 2 to 7 have great values of R close to 1 and small values of RMSE near to 0 that shows the models accuracy. Anyhow, model No. 1 does not present an efficient prediction of DO. The hourly ANN models present more accurate predictions than the daily ANN models. For the prediction of DO, the greatest value of R for the daily ANN model is 0.74, while it reaches about 0.94 for hourly model.

Results related to the hourly WNN models for DO prediction are shown in Table 6. It can be concluded that predictive WNN models for the prediction of hourly values of DO have the greatest accuracy. Similar to the previous subsection, the WNN models outperform the ANN models. Model No. 8 which only has the water temperature, salinity, turbidity and the amount of chlorophyll in time t gives an acceptable prediction of DO in time t. This finding is interesting and can be used to predict missing data of a variety of ocean parameters when the values of the target variable in the preceding time steps are not available.

Regarding Figs. 6 and 7, it can be observed that the WNN model has greater value of R^2 in comparison with the ANN model. Therefore, higher correlation between observed and predicted values of DO is obtained for the WNN model than for the ANN model. Also, part b) of these figures demonstrates that both ANN and WNN model provide an accurate prediction for hourly values of DO.

Generally, hourly WNN models are recognized as the most efficient models for the prediction of the ocean water quality

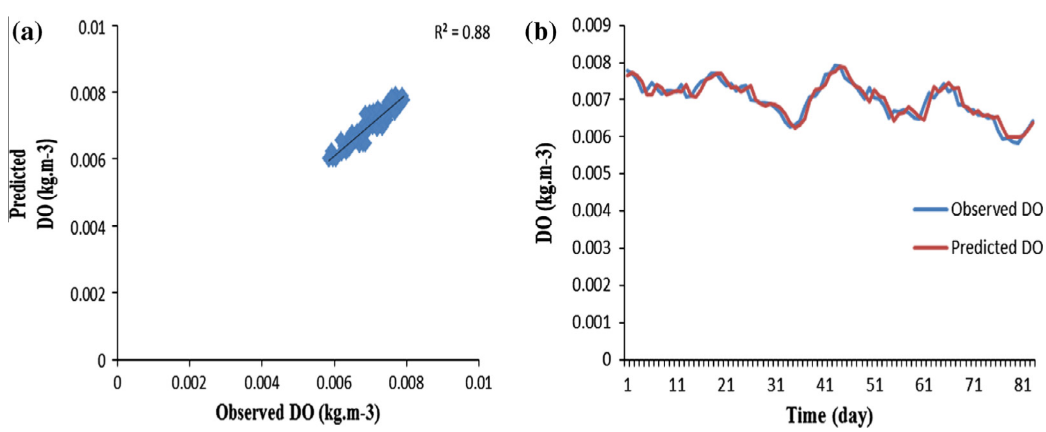

Fig. 6. (a) Scatter plot of the observed versus the predicted DO values and (b) Comparison of the predicted and the measured DO for the testing set of the ANN model (model No. 4).

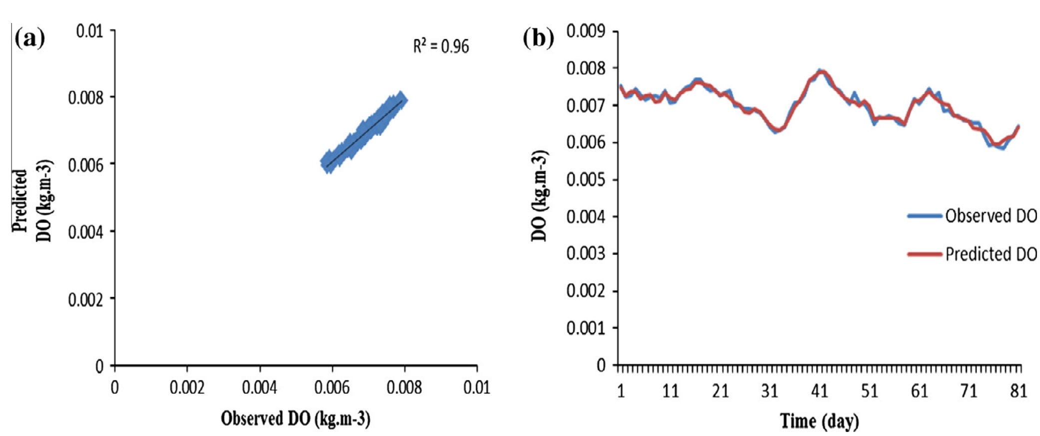

Fig. 7. (a) Scatter plot of observed versus predicted DO values and (b) Comparison of the predicted and measured DO for the testing set of WNN model (model No. 9).

parameters in the study. Moreover, the best model performance has been achieved for the daily and hourly WNN models when the values of the target values up to 3 preceding time steps were used as input to predict the same variable in time step t (model No. 9 for the prediction of the daily and hourly values of DO, models No. 19, and 29 for the prediction of daily values of water temperature and salinity).

As it is clear, tidal waves, wind-induced waves, pressure fluctuation and other ocean phenomena can affect water quality parameters in oceans and bays. Results of this study show that it is possible to have a sound prediction of the water quality parameters without including some important parameters such as significant wave height and pressure. However, it seems that including such parameters in the input structure of the ANN and WNN models can increase their performances.

4. Conclusions

The present study provides an application of ANN and WNN models for the prediction of a variety of ocean water quality parameters including daily values of DO, water temperature, salinity as well as hourly values of DO in Hilo Bay, Pacific Ocean. An attempt was made to get an appropriate prediction for the parameters based on different combinations of the variables in the input structure of ANN and WNN. In the model development, only a limited number of variables including DO, water temperature, salinity and turbidity were applied and many of the important parameters in the ocean water quality such as wave height, pressure, phosphate, nitrate and ammonia were excluded. In any case, a very good correlation was observed between the measured and predicted values when the WNN models were applied. Results of the study reveals that wavelet transform as a pre-processing technique can improve the performance of an existing ANN model remarkably.

For all the cases, the WNN models outperform the ANN models and hourly models are more accurate than daily models. Among daily and hourly ANN and WNN models applied in this study, hourly WNN models are the most accurate models in terms of correlation coefficient and RMSE. For the hourly WNN models, the correlation coefficient ranges from 0.92 to 0.99. The best model for daily prediction of all the three parameters has the same pattern regarding the input variables (models No. 9, 19, and 29 for the prediction of DO, water temperature and salinity respectively). The same pattern for hourly model can be recognized (using the target variable with 3 preceding time steps as input, $t-1$, $t-2$, $t-3$).

In spite of the hourly WNN models, a suitable prediction of daily values of the target variable cannot be achieved when the preceding time steps of the target variable is excluded from the input structure (models No. 8, 18 and 28). It can be derived that many known and unknown parameters affecting the water quality parameters and to have an efficient models, some other parameters such as significant wave height and pressure should be included in the model development. Anyhow, using hourly WNN model, a sound prediction of DO at time t can be achieved when water temperature, salinity and the amount of chlorophyll at time t are considered as input variables. This result is interesting because by only using the hourly values of 3 parameters, the values of the other parameter in time t can be predicted with a relatively high accuracy ($R > 0.9$). It is helpful to estimate the value of some parameters without measurement and sampling and only based on some related measured parameters.

Findings of this study can be employed to give an accurate prediction for the missing data and for times in which continuous monitoring is difficult. The technique applied for this study (WNN) has the advantages of high accuracy, time and cost reduction in comparison to other computational techniques. The WNN models can be successfully applied in order to predict daily and hourly values of DO, water temperature, and salinity in bays and oceans. Moreover, by providing accurate prediction for some of time intervals, they help to reduce sampling numbers which can be expensive and time consuming. Results of this study are encouraging to develop some other WNN models in order to predict more water quality parameters. Also, this approach can be examined to predict a variety of ocean parameters in different bays and areas of seas and oceans.

References

Alizadeh, M.J., Kavianpour, M.R., Tahershamsi, A., Shahheydari, H., 2015. A wavelet-ANN approach to investigate the effect of seasonal decomposition of time series in daily river flow forecasting. In: 10th International Congress on Civil Engineering, Tabriz, Iran.

Alizdeh, M.J., Joneyd, P.M., Motahhari, M., Ejlali, F., Kiani, H., 2015. A wavelet-ANFIS model to estimate sedimentation in dam reservoir. Int. J. Comput. Appl., 114

Bahaa, M.K., Ayman, G.A., Hussein, K., Ashraf, E.-S., 2012. Application of artificial neural networks for the prediction of water quality variables in the nile delta. J. Water Resour. Protection, 2012.

Cohen, A., Kovacevic, J., 1996. Wavelets: The mathematical background, Proc. IEEE. Citeseer.

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

Faruk, D.Ö., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Eng. Appl. Artif. Intell. 23, 586–594.

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

Govindaraju, R.S., 2000. Artificial neural networks in hydrology. II: hydrologic applications. J. Hydrol. Eng. 5, 124–137.

Gupta, K.K., Gupta, R., 2007. Despeckle and geographical feature extraction in SAR images by wavelet transform. ISPRS J. Photogramm. Rem. Sens. 62, 473–484.

Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. Neural Networks, IEEE Trans. 5, 989–993.

Ham, F., Kostanic, I., 2001. Principles of neurocomputing for science and engineering.

Heydari, M., Olyaie, E., Mohebzadeh, H., Kisi, Ö., 2013. Development of a neural network technique for prediction of water quality parameters in the Delaware River, Pennsylvania. Middle East J, Sci. Res. 13, 1367–1376.

Kisi, O., 2008. River flow forecasting and estimation using different artificial neural network techniques.

Labat, D., Ababou, R., Mangin, A., 2000. Rainfall–runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses. J. Hydrol. 238, 149–178.

Musavi, S., Golabi, M., 2008. Application of artificial neural networks in the river water quality modeling: Karoon River, Iran. J. Appl. Sci., 2324–2328

Nourani, V., Alami, M.T., Aminfar, M.H., 2009. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Eng. Appl. Artif. Intell. 22, 466–472.

Nourani, V., Kisi, Ö., Komasi, M., 2011. Two hybrid Artificial Intelligence approaches for modeling rainfall–runoff process. J. Hydrol. 402, 41–59.

Okkan, U., 2012. Wavelet neural network model for reservoir inflow prediction. Scientia Iranica 19, 1445–1455.

Palani, S., Liong, S.-Y., Tkalich, P., Palanichamy, J., 2009. Development of a neural network model for dissolved oxygen in seawater. Indian J. Mar. Sci. 38, 151.

Palani, S., Tkalich, P., Balasubramanian, R., Palanichamy, J., 2011. ANN application for prediction of atmospheric nitrogen deposition to aquatic ecosystems. Mar. Pollut. Bull. 62, 1198–1206.

Rajaee, T., Nourani, V., Zounemat-Kermani, M., Kisi, O., 2010. River suspended sediment load prediction: application of ANN and wavelet conjunction model. J. Hydrol. Eng. 16, 613–627.

Ravansalar, M., Rajaee, T., Ergil, M., 2015. A hybrid model of Wavelet with Artificial Neural Network for monthly prediction of river water quality.

Singh, K.P., Basant, A., Malik, A., Jain, G., 2009. Artificial neural network modeling of the river water quality—a case study. Ecol. Model. 220, 888–895.

Tryland, I., Myrmel, M., Østensvik, Ø., Wennberg, A.C., Robertson, L.J., 2014. Impact of rainfall on the hygienic quality of blue mussels and water in urban areas in the Inner Oslofjord, Norway. Mar. Pollut. Bull. 85, 42–49.

Weiner, R., Matthews, R., 2003. Environmental engineering. Butterworth-Heinemann.

Wu, H., Chen, K., Chen, Z., Chen, Q., Qiu, Y., Wu, J., Zhang, J., 2012. Evaluation for the ecological quality status of coastal waters in East China Sea using fuzzy integrated assessment method. Mar. Pollut. Bull. 64, 546–555.

Zhang, F., Dai, H., Tang, D., 2014. A conjunction method of wavelet transform-particle swarm optimization-support vector machine for streamflow forecasting. J. Appl. Math., 2014