Monthly Forecasting of Water Quality Parameters within Bayesian Networks: A Case Study of Honolulu, Pacific Ocean

Ehsan Jafari Nodoushan [a*]

[a] *Assistant Professor, Department of Civil Engineering, Bijar Branch, Islamic Azad University, Bijar, Iran.*

Abstract

This study investigates the efficiency of Bayesian network (BN) and also artificial neural network models for predicting water quality parameters in Honolulu, Pacific Ocean. Monthly forecasting of three important characteristics of water body including water temperature, salinity and dissolved oxygen have been taken under consideration. Two separate strategies were applied in which the first strategy was related to prediction of the water quality parameters based on previous time series of the same variable. In the second strategy, an attempt was made to forecast DO using different affecting parameters such as temperature, salinity, previous time series of DO, and amount of chlorophyll. The efficiency of the models were assessed by using error measures. Results revealed that the BN models are superior over the ANN models in case of temperature and DO forecasting. Also, it was found that the first strategy is more efficient than the second strategy for predicting DO concentration. The best BN models for temperature, salinity and DO were achieved when time series of the same parameter up to 3, 2, and 3 previous months applied as input variables respectively. Overall, it can be concluded that BN and ANN models can be successfully applied for water quality modelling and forecasting in coastal waters. Moreover, the current study demonstrated that the BN models have a great ability dealing with time series including incomplete or missing data.

Keywords: Bayesian Network, ANN, Salinity, Temperature, Dissolved Oxygen.

1. Introduction

Water quality in coastal and ocean waters is a key element for environmental management and safe aquatic life. A beforehand knowledge of water quality parameters can provide useful information for environmental management and planning. Water quality studies and monitoring programs are being increasingly used due to an increase in water pollution incident. In this regard, many researchers developed different models to predict different water quality parameters in rivers and oceans. Regarding water quality in coastal and estuarine waters, a wide variety of parameters can be influential which finding an exact relationship among them is not an easy task. Traditionally, the methods applied for water quality modelling and forecasting were based on linear relationships which mainly were not accurate enough due to ignoring nonlinear relationships among the variables [1].

Over the past several decades, many statistical analyses and artificial intelligence modelling methods have been successfully applied for water quality time series prediction. Jayawardena and Lai (1989) applied Auto-Regressive Integrated Moving Average (ARIMA) model to predict future water quality in Pearl River of China [2]. To predict water parameters such as BOD, electrical conductivity and chloride, Ahmad et al. (2001) developed the multiplicative ARIMA models for the river Ganges in India [3]. Chau (2006) reviewed the integration of AI techniques including knowledge-based system, genetic algorithm, artificial neural network, and fuzzy inference system into water quality modeling [4]. Said et al. (2011) used hydrographic data in order to investigate changes in Atlantic water characteristics

as a result of natural and anthropogenic activities [5]. Faruk (2010) proposed a hybrid ARIMA and neural network model gaining linear advantage of traditional ARIMA models and nonlinear feature of ANN models [6]. The proposed model was employed for predicting waters quality parameters of water temperature, boron, and dissolved oxygen. The results demonstrated superiority of the hybrid model over the ARIMA models and the ANN models. Gazzaz et al. (2012) applied fully-connected, three-layer perceptron neural network model for computing the water quality index [7]. The data obtained from Kinta River in Malaysia was used as case study. To compute the water quality index, data related to more than 20 water quality parameters recorded in the river were applied during the modelling process. Findings of this study revealed that the ANN can be successfully applied for predicting the water quality index. In other study for the Stillaguamish River in Washington, Arya and Zhang (2015) employed predictive models using order series method (OSM) for investigating the normality assumption, water quality parameters such as temperature and dissolved oxygen have been taken under consideration [8]. Gao et al. (2015) developed a Bayesian regularized back propagation ANN model for predicting monthly chlorophyll-a concentration in Meiliang Bay, Lake Taihu [9]. Alizadeh and Kavianpour (2015) developed wavelet-ANN models for daily forecasting of temperature, DO and turbidity in Hilo Bay, Pacific Ocean [1]. In this study, performance of ANN models and wavelet-ANN models for water quality forecasting have been investigated in which the results indicated outperformance of wavelet-ANN models compared with the ANN models. Heddam and Kisi (2017) investigated performance of extreme learning machine as a new approach for modelling dissolved oxygen concentration [10]. The results demonstrated high efficiency of the proposed approach for modelling in river ecosystem. Solgi et al. (2017) employed two types of predictive models called support vector regression (SVR) and adaptive neural fuzzy inference system (ANFIS) for predicting Biochemical Oxygen Demand (BOD) in Karun River, Iran. In order to improve performance of the models, they decomposed input time series by wavelet transform and selected efficient sub-signals using principle component analysis (PCA). The results of the models showed that the SVR model outperforms the ANFIS model. Also, it was found that combining the SVR with the wavelet transform (WSVR) was a good idea to improve the prediction of the BOD value in Karun River [11].

However, the previous models ignore uncertain entity embedded in sensors and instruments in measuring procedure. Therefore, they suffer lack of accurate predictions for water quality parameters and further investigations with considering uncertain inherent of the data including missed data, errors in measurements, or access to the measured data (Deng et al., 2015). In recent years, Bayesian Network (BN) models showed great ability in time series prediction including missing or incomplete data, uncertain, nonlinear and complex relationship among variables [12]. Alizadeh et al. (2017b) indicated that BN models outperforms the ANN models when applied for predicting of longitudinal dispersion coefficient [13]. Also, BN models have been applied to various problems, optimum management of groundwater contamination, nonlinear time series and spatio-temporal drought forecasting [14, 15 ,16].

The main objective of this study is to provide forecasting models for water quality parameters usingz Bayesian networks. An attempt is made to develop models with acceptable accuracy and also to handle uncertainty of the forecasts. In this study, monthly time series of temperature, DO, salinity and the amount of chlorophyll are considered as the water quality parameters. The study is carried out for a case study of Honolulu, Pacific Ocean. The results obtained through the BN models are compared with those of the ANN models. The rest of the paper is organized as follows. Section 2 describes ANN, BN theories, water quality parameters, data and the study area. Section 3 explains the modelling strategy and how ANN and BN models applied for this study. Results of the ANN and BN models are discussed in Section 4. Finally, Section 5 summarizes the major remarks of the study.

2. Material and Methods

2.1. Artificial Neural Networks (ANNs)

An ANN is a 'black box' approach which has great capacity in predictive modelling. It is a proper mathematical structure having an inter-connected assembly of simple processing elements or nodes. A feed forward neural network (FFNN) model, consisting of multiple layers of nodes in a directed graph, utilizes a supervised learning technique called back propagation for training the network [17]. Depending on the techniques to train a FFNN model, different algorithms of back propagation and metaheuristics can be employed [18]. The metaheuristic algorithms have more computational cost and complexity than the back propagation algorithms. Therefore, the Levenberg–Marquardt technique (a simplified version of Newton method) as a back propagation algorithm is applied in this study to train the ANN models. The simplicity, low computational cost and acceptable reliability of the algorithm made it as a popular training algorithm of ANN models. Basically, an ANN model can be mathematically formulated as:

$$y = \varphi \left(\sum_{i=1}^{n} w_i x_i + b \right) \tag{1}$$

In which w_i= the weight of the input node of x_i, b= the bias, φ = the activation function and n= number of data samples [19]. More information related to ANNs can be found in [20, 21].

2.2. Bayesian Networks (BNs)

A Bayesian network is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). Formally, Bayesian networks are DAGs whose nodes represent random variables in the Bayesian sense: they may be observable quantities, latent variables, unknown parameters or hypotheses. Edges represent conditional dependencies; nodes that are not connected represent variables that are conditionally independent of each other. Each node is associated with a probability function that takes, as input, a particular set of values for the node's parent variables, and gives (as output) the probability (or probability distribution, if applicable) of the variable represented by the node. A BN is composed of: (a) a set of variables and a set of directed links between the variables; (b) a set of mutually exclusive and exhaustive states for each variable; and (c) an assigned conditional probability for each variable with 'parents', which will be defined shortly. The relations between the variables in a BN are expressed in terms of family relationships, where a variable A1 is said to be the parent of B (B the child of A1) if the link goes from A1 to B. The dependencies are quantified by conditional probabilities for each node given its parents in the network. These dependencies are quantified through a set of CPTs (Conditional Table of Probabilities); each variable is assigned a CPT of the variable given its parents. In the case of a variable with no parents, the conditional probability structure reduces to the unconditional probability (UP) of that variable [22].

The Ups of the basic input parameters, often, are not known a priori, consequently, equal weights (1/n, where n is number of category considered for each basic input) can be assigned using the principle of insufficient reasoning [23]. In a BBN analysis, for n number of mutually exclusive parameters X_i (i = 1, 2, . . ., n), and a given observed data Y, the updated probability is computed by:

$$p(X_i|Y) = \frac{p(Y|X_i)*p(X_i)}{\sum_j p(Y|X_j)p(X_j)} \qquad (2)$$

In which $p(X_i|Y)$ is the posterior occurrence probability of X_i given the condition that Y occurs, $p(X_i)$ is the prior occurrence probability of X_i, $p(X_j)$ stands for the marginal (total) occurrence probability of Y and is effectively constant since the obtained data is in hand, and $p(Y|X_j)$ refers to the conditional occurrence probability of Y given that X_j occurs too (it is often viewed in this sense as the likelihood distribution)[24].

2.3. Study Area and Data

Generally, water quality consists of a wide variety of variables that covers physical, chemical and biological parameters. Some of these variables are mutually interrelated. In this study, four water quality variables including water temperature (T), DO, chlorophyll (Ch), and salinity (S) are applied. These are monthly average values of data recorded at 12 o'clock. These data were obtained from water quality buoy of WQB-KN at the Kilo Nalu Nearshore Reef Observatory, near Kakaako Waterfront Park in Honolulu (21.2887N and 157.865W). The data are available following the web server of (http://oos.soest.hawaii.edu/dchart/). 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. 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). Figure 1 illustrates the location of water quality buoy in Oahu, Honolulu County, and Pacific Ocean.
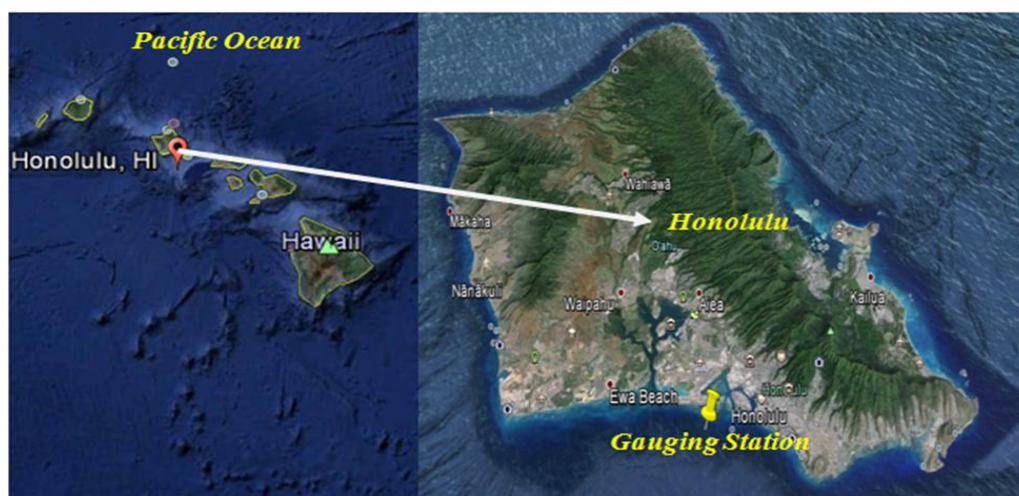

Figure 1. Location of the water quality buoy in Honolulu

The data consist of monthly values of the above-mentioned parameters from Aug. 2008 to Sep. 2014. These values were measured in the depth of 1 meter below the mean surface level (z = -1 m). The number of water samples selected for the predictive models are 74. In the data sets, the data records of water quality parameters for about 12 months is not available. Monthly time series of the water quality parameters are depicted in Figure. 2. Also, missing data can be observed in Figure 2.
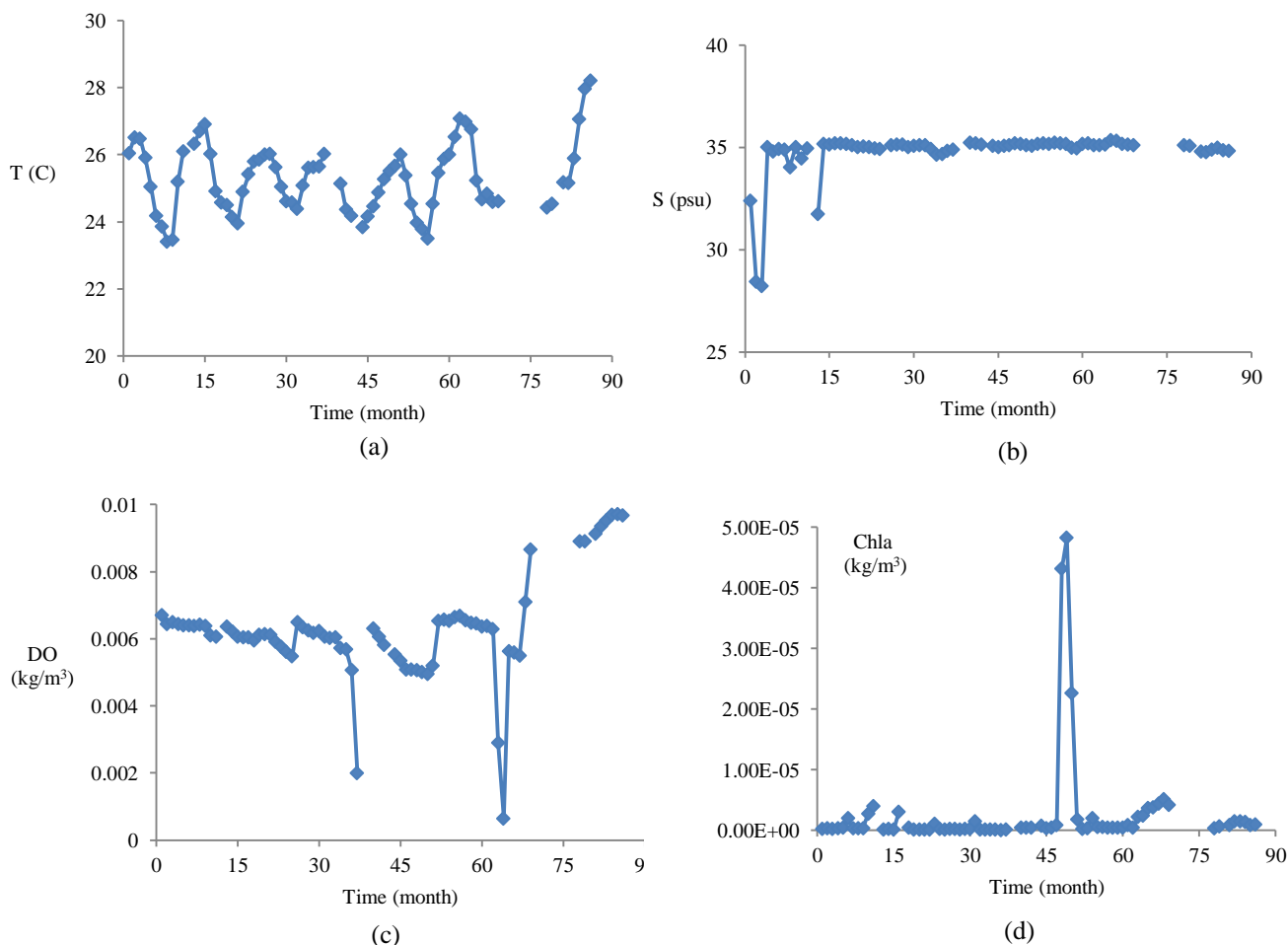

(a)

(b)

(c)

(d)

Figure 2. Monthly time series of a) temperature, b) salinity, c) DO and d) chlorophyll

The statistical analysis of the data have been carried out using basic statistical measures including average ('Mean'), standard deviation ('Sd'), skewness coefficient ('Csx'), minimum ('Min') and maximum ('Max'). These indices for the monthly datasets are presented in Table 1.

Table 1. Statistical analysis of the data

Parameter	Min	Max	Mean	Sd	Csx
Temperature [C]	23.4	28.2	25.3	1.04	0.435
Salinity [psu]	28.2	35.4	34.8	1.22	-4.588
DO [kg/m^3]	6.6e-4	9.7e-3	6.3e-3	1.5e-3	-0.218
Chlorophyll [kg/m^3]	6.9e-8	4.8e-5	2.5e-6	7.8e-6	5.041

3. Modelling Strategy

Generally, the models applied for water quality prediction can be grouped in two types. First type employs different influential variables for the target prediction. In other words, values of several variables in time t are used to predict the target variable in time t. As they do not consider the correlation of the previous time steps, they cannot be suitable for future time series forecasting. The other type applies the previous time series of the same variable for the same variable prediction in the upcoming time [25-30]. In this study, two strategies are applied which in the first strategy ANN and BN models are employed to predict monthly values of temperature, salinity and DO using the preceding time series of the same parameter. In the other strategy, preceding time series of DO as well as time series of temperature, salinity and chlorophyll in the present and previous months were applied as input variables in order to predict DO in the current time step. The data were divided into training (80%) and testing sets (20%). In the ANN

models, the number of neurons in the hidden layer was 8. To achieve the most accurate model for each parameter, different models applying different time lags of the variable were examined. In the first strategy, preceding time series of temperature, salinity and DO up to 3 months ago were applied as input variables. Structure of the ANN and BN models developed for DO(t) forecasting with three input variables of DO(t-1), DO(t-2) and DO(t-3) is shown in Figure. 3. DO(t-1) ...DO(t-3) represents values of DO up to three time lags. For two other parameters including temperature and salinity a similar structure have been taken under consideration.
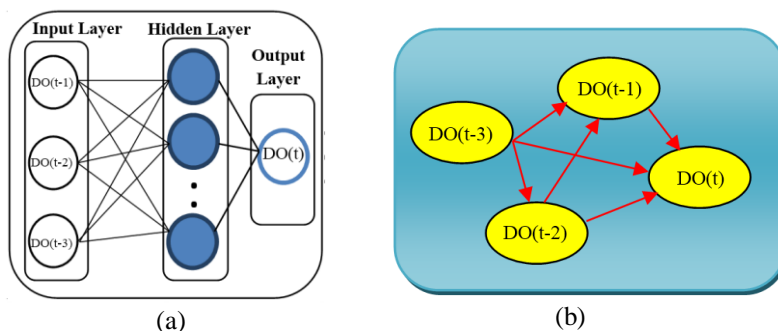

Figure 3. A schematic of the structure of the a) ANN model, b) BN model

In the BN model developed in this study, the state of the parameter DO in time t depends on the states of three other parameters: DO(t-1), DO(t-2) and DO(t-3). Therefore, DO(t) is considered conditionally dependent on the states of DO in preceding time steps. In Figure 2 the dependence between the variables is indicated by a simple graph, where the parameters are represented by nodes and their mutual dependences by arrows. Note that the directions of arrows indicate the cause-effect processes and their feasibility.

Another strategy in the model development was to consider other parameters such as temperature, salinity and chlorophyll as well as preceding time series of DO to predict DO values in time t. In this regard, different combinations of the variables have been examined to achieve the most accurate ANN and BN models. Fig. 4 gives an example of the structure of ANN and BN models for this purpose. It should be pointed out that the ANN and BN models in Figure.4 have different input variables in their structures. Referring to the graph in Fig. 4b, it is clear that, changes in T(t), Ch(t), T(t-1) and/or DO(t-1) values can all affect DO(t), a reverse process is not feasible. Moreover, in Figure. 4 there is no dependency between S(t) and DO(t).
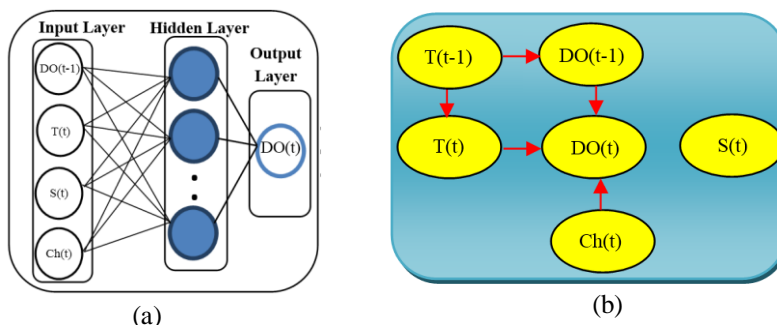

Figure 4. An example of the structure of a) ANN and b) BN models for the second strategy

The performance of the developed models was evaluated by using error measures such as the correlation coefficient (CC), the root mean square error (RMSE) and mean absolute error (MAE). In brief, the models' predictions are optimum if CC, RMSE and MAE are found to be close to 1, 0, and 0, respectively. These parameters are defined as:

$$CC = \frac{\sum_{i=1}^{n}(y_{i(observed)} - \bar{y}_{(observed)})(y_{i(predicted)} - \bar{y}_{(predicted)})}{\sqrt{\sum_{i=1}^{n}(y_{i(observed)} - \bar{y}_{(observed)})^2 \sum_{i=1}^{n}(y_{i(predicted)} - \bar{y}_{(predicted)})^2}} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{i(observed)} - y_{i(predicted)})^2}{n}} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_{i(observed)} - y_{i(predicted)}|}{n} \tag{5}$$

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

It should be mentioned that the ANN models were developed creating a program code via MATLAB software and the BN models were implemented in the Hugin Lite 8.3 software.

4. Results and Discussion

Results related to temperature and salinity forecasting models which developed based on the first strategy (using the preceding time series of the same variable) are presented in subsection 4.1. Results of DO forecasting models which developed using two distinguished strategies are presented in subsection 4.2.

4.1. Temperature and Salinity Forecasting Models

In this part, an attempt was made to develop monthly forecasting models for temperature and salinity using preceding time series of the target variable in the input structure of the ANN and BN models. For predicting temperature, three different ANN and BN models were developed in which the first model only had $T(t-1)$ as input variable. The second and third models applied up to two time lags ($T(t-1)$ and $T(t-2)$) and three time lags ($T(t-1)$, $T(t-2)$ and $T(t-3)$) respectively to predict $T(t)$. A similar procedure has been carried out for salinity forecasting. Table 2 presents the results related to CC, RMSE and MAE of different ANN and BN models developed for temperature and salinity predictions.

Table 2. Results of the ANN and BN models during testing period for temperature and salinity forecasting

Variable	Time lag (Input)			Output	ANN model			BN model		
	(t-1)	(t-2)	(t-3)		CC	RMSE	MAE	CC	RMSE	MAE
	✓			T(t)	0.13	1.24	1.099	0.833	0.675	0.592
Temperature	✓	✓		T(t)	0.732	0.83	0.746	0.92	0.477	0.372
	✓	✓	✓	T(t)	0.904	0.532	0.431	0.921	0.472	0.36
	✓			S(t)	0.884	1.004	0.469	0.586	1.94	1.256
Salinity	✓	✓		S(t)	0.76	1.134	0.648	0.564	2.007	1.49
	✓	✓	✓	S(t)	0.643	1.87	1.366	0.492	2.06	1.683

It can be derived from Table 2 that both of the soft computing techniques (ANN and BN) applied in this study provide a reasonable prediction of water temperature and salinity. High values of CC and low values of RMSE and MAE for both temperature and salinity forecasting demonstrate the efficiency of the proposed models. For temperature forecasting the best model performance was achieved when time series of temperature up to three precedent months were applied as input variables. The correlation coefficient of the ANN and BN models for this case is higher than 0.9 which indicate the great ability of the developed models. In this case, the RMSE and MAE are the least. For salinity forecasting the model including only $S(t-1)$ in its input structure shows the best performance. This model has the highest values of correlation coefficient and the lowest values of RMSE and MAE. A comparison between the accuracy of the ANN and BN models reveals that for the temperature forecasting the BN model outperforms the ANN model whereas for the salinity forecasting the ANN model is superior over the BN model. The best ANN and BN models for temperature forecasting have a CC of 0.9 and 0.92 respectively. Moreover, their RMSE and MAE are equal to 0.53, 0.47, 0.43 and 0.36 respectively. For the salinity predictions the values of CC, RMSE and MAE for the ANN model are 0.88, 1 and 0.46 and for the BN model are 0.58, 1.94 and 1.25. Scatterplots of the best ANN and BN models for the temperature and salinity forecasting illustrated in Figures 5 and 6 respectively. It is noteworthy that all the results presented in this study obtained for testing set.
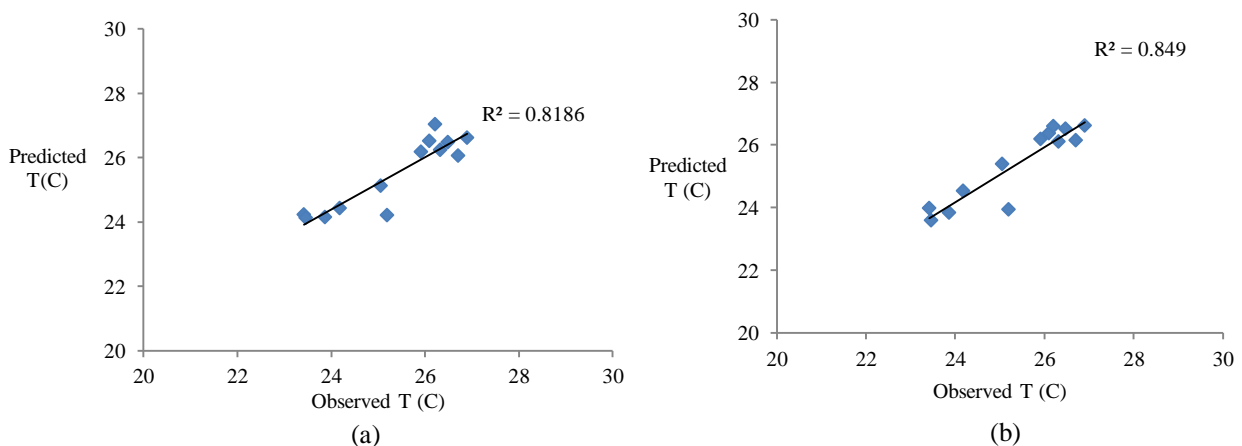

(a)   (b)

Figure 5. Scatterplots of the best a) ANN and b) BN models for the temperature during testing period

(a)   (b)
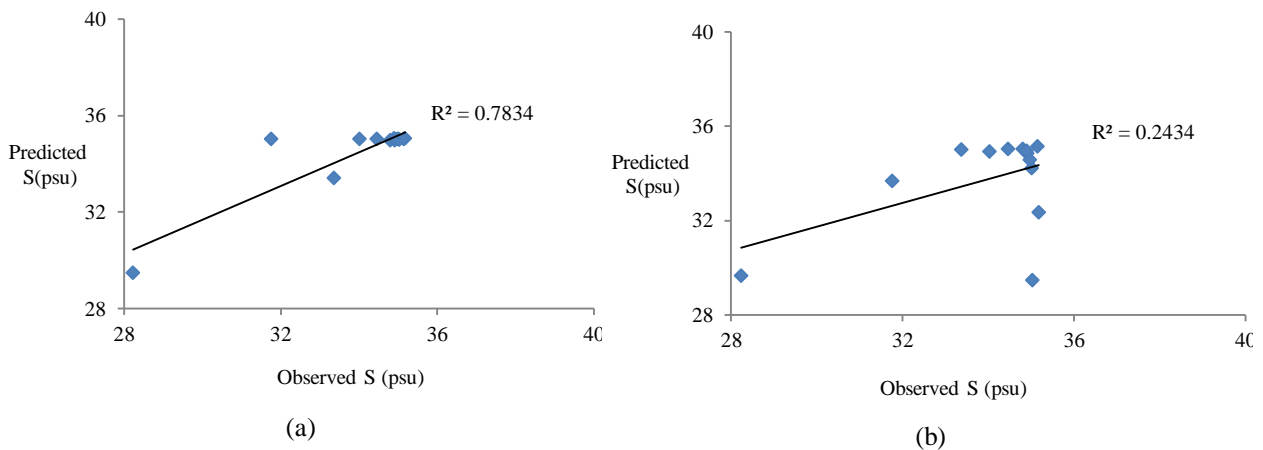
Figure. 6. Scatterplots of the best a) ANN and b) BN models for the salinity during testing period

To provide more comparison, time series of the observed and predicted values of temperature obtained by the best ANN and BN models depicted in Figure 7. Also, the time series of salinity for testing set obtained through the best ANN and BN models are shown in Figure 8.
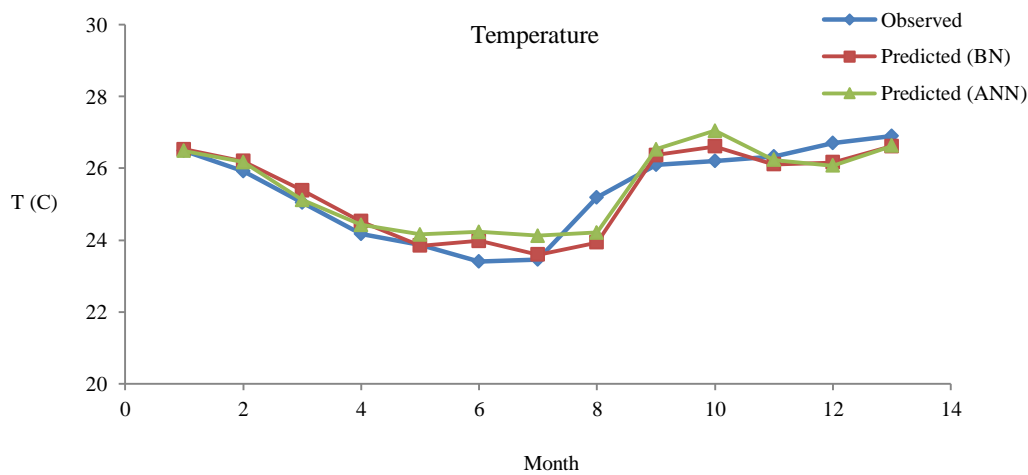

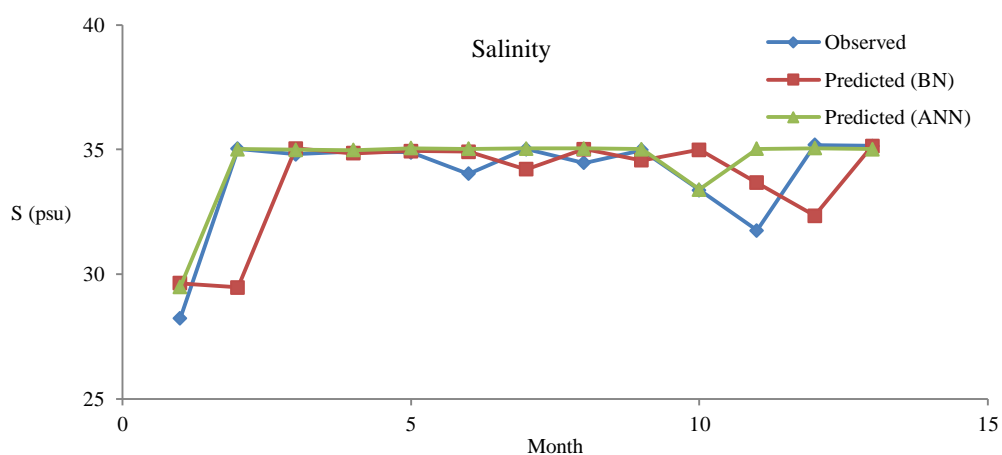**Figure. 7. Time series of the monthly temperature obtained during testing set**

Figure. 8. Time series of the monthly salinity obtained during testing set

It can be derived from Figures. 7 and 8 that an accurate prediction of monthly temperature and salinity can be achieved by using the soft computing techniques (ANN and BN models). For predicting temperature, a very high correlation between predicted and measured data can be observed when BN model is applied. For salinity prediction, the ANN model provide a high accurate prediction. Generally, the ANN model for both temperature and salinity forecasting a slightly overestimate the target values. Regarding BN model, no general conclusion can be extracted.

4.2. DO Forecasting Models

To predict monthly time series of DO, two kinds of strategy have been taken under consideration. In the first one, the procedure was similar to the previous subsection in which preceding time series of the DO were applied to predict DO in time step t. In the latter strategy, some other variables such as temperature, salinity and the amount of chlorophyll were added to the input structure. Table 3 presents results of the ANN and BN models developed for monthly forecasting of DO and during testing set. For all the developed models for DO forecasting, DO(t) was considered as target value (models' output).

Table 3. Results of the ANN and BN models for DO forecasting and for testing set

Model No.	Input variables	ANN model			BN model		
		CC	RMSE	MAE	CC	RMSE	MAE
1	DO(t-1)	0.544	1.82e-4	1.64e-4	0.568	1.29e-4	9.7e-5
2	DO(t-1), DO(t-2)	0.56	1.48e-4	1.34e-4	0.572	1.27e-4	9.6e-5
3	DO(t-1, DO(t-2), DO(t-3)	0.617	1.47e-4	1.34e-4	0.634	1.27e-4	9.6e-5
4	Ch(t), S(t), T(t)	0.142	2.2e-3	1.99e-3	0.319	3.88e-3	2.5e-3
5	Ch(t), S(t), T(t), DO(t-1)	0.317	8.01e-4	4.86e-4	0.418	2.43e-3	1.88e-3
6	Ch(t), T(t-1), T(t), DO(t-1)	0.077	4.18e-4	3.29e-4	0.6	1.9e-4	1.5e-4
7	DO(t-1), DO(t-2), T(t), T(t-1),Ch(t)	0.219	3.4e-4	3.03e-4	0.529	2.53e-4	2.05e-4

According to Table 3, models which developed following the first strategy provide more accurate prediction of DO than those models applying the second strategy. Moreover, the best BN model outperforms the best ANN models for both strategies. The highest correlation coefficient (0.634) was achieved for model No. 3 when three previous time steps of DO time series were applied as input parameters/parents in ANN/BN model. Overall, the best model for DO forecasting has RMSE and MAE of 1.27e-4 and 9.6e-5 respectively. It was found during modelling procedure that including chlorophyll in model development does not affect the model's performance significantly. Anyhow, including chlorophyll improves a little the model's efficiency. Also, the best BN model developed based on the second strategy does not have salinity as parent or effective parameter. In the second strategy, the best ANN and BN models have different input variables. In this regard, model No. 6 for the BN and model No. 7 for the ANN model gives the best performance. To provide more comparison, the prediction accuracy of different models illustrated in Figure 9.
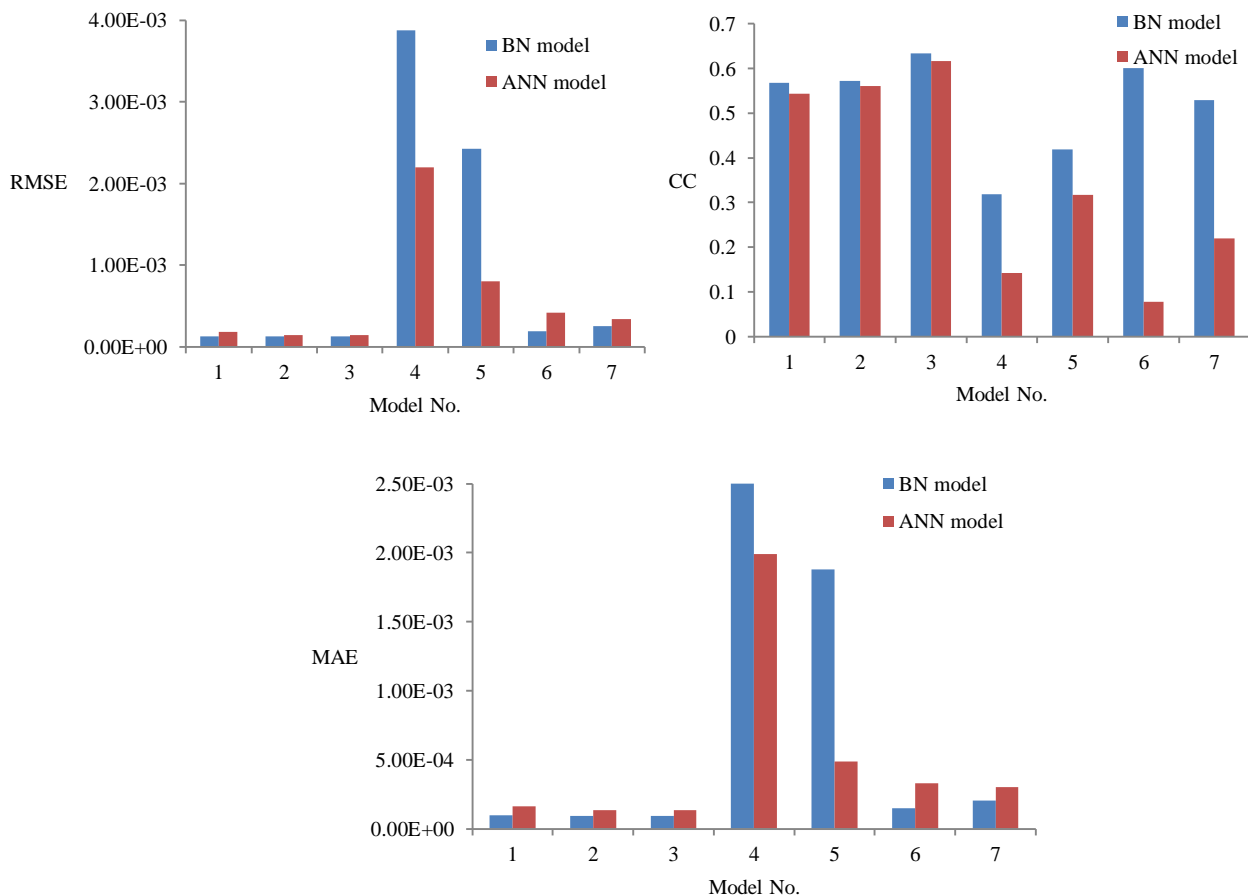

Figure. 9. Comparison of different ANN and BN models' efficiency for DO forecasting

Regarding Figure. 9 it can be concluded that models No. 4 and 5 are not efficient models for DO forecasting. Therefore, excluding previous time series of DO in the input structure affect the models' performances negatively. It can be derived that without previous time series of DO, an accurate prediction model for DO forecasting cannot be achieved. Overall, the BN models applied in this study are superior over the ANN models. Regardless models No. 4 and 5, all the other BN models outperform the ANN models. These BN models have higher values of CC and lower values of RMSE and MAE. Scatterplots of the best ANN and BN models obtained from the first and the second strategy are shown in Figures 10 and 11 respectively.
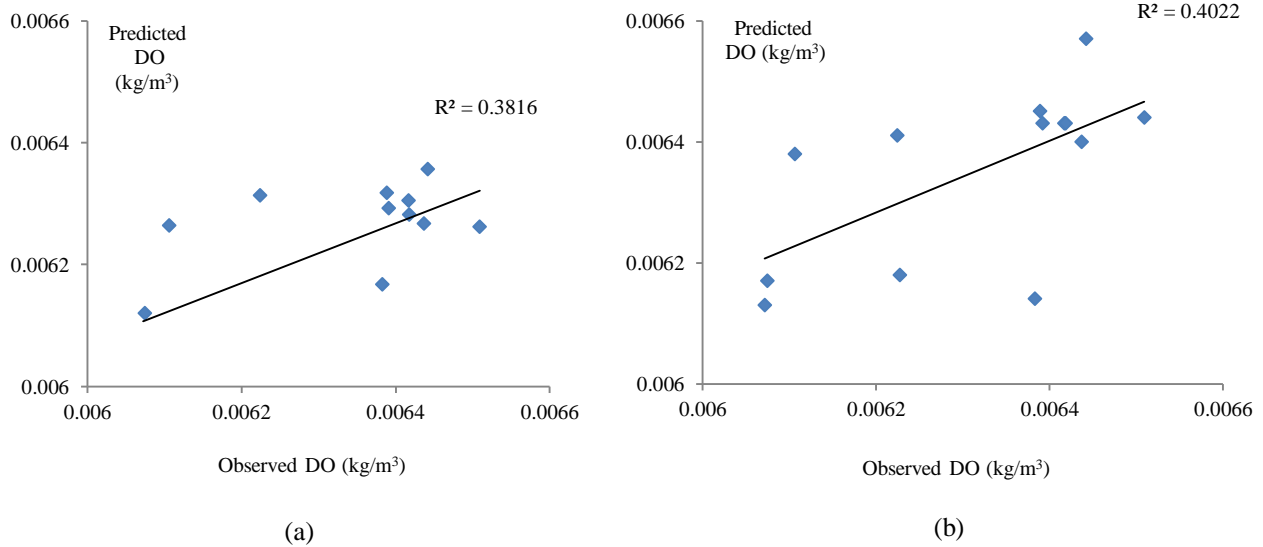

(a)   (b)

Figure 10. Scatterplots of the best a) ANN and b) BN models for DO forecasting based on the first strategy

(a)   (b)
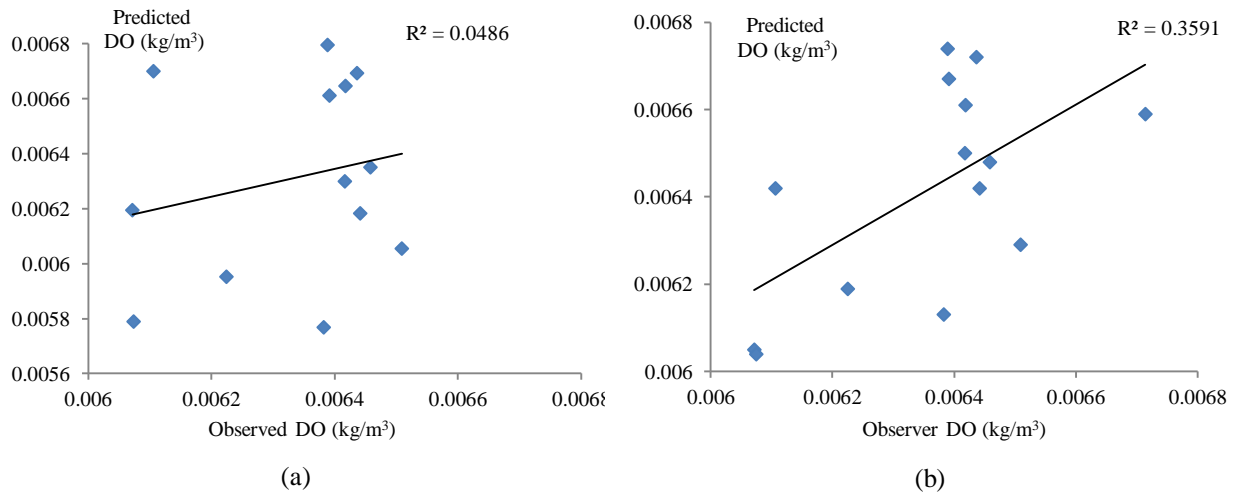
Figure. 11. Scatterplots of the best a) ANN and b) BN models for DO forecasting based on the second strategy

As it can be seen from Figures 11 and 12, the best correlation is derived for the first strategy and BN model. The highest R^2 which equals 0.402 is belong to the BN model developed based on the first strategy. Therefore, applying previous time series of DO with the BN model can provide an acceptable prediction of DO for one time step ahead (here one month ahead of time). Figure 12 illustrates the time series of observed DO and predicted values obtained by the best ANN and BN models.
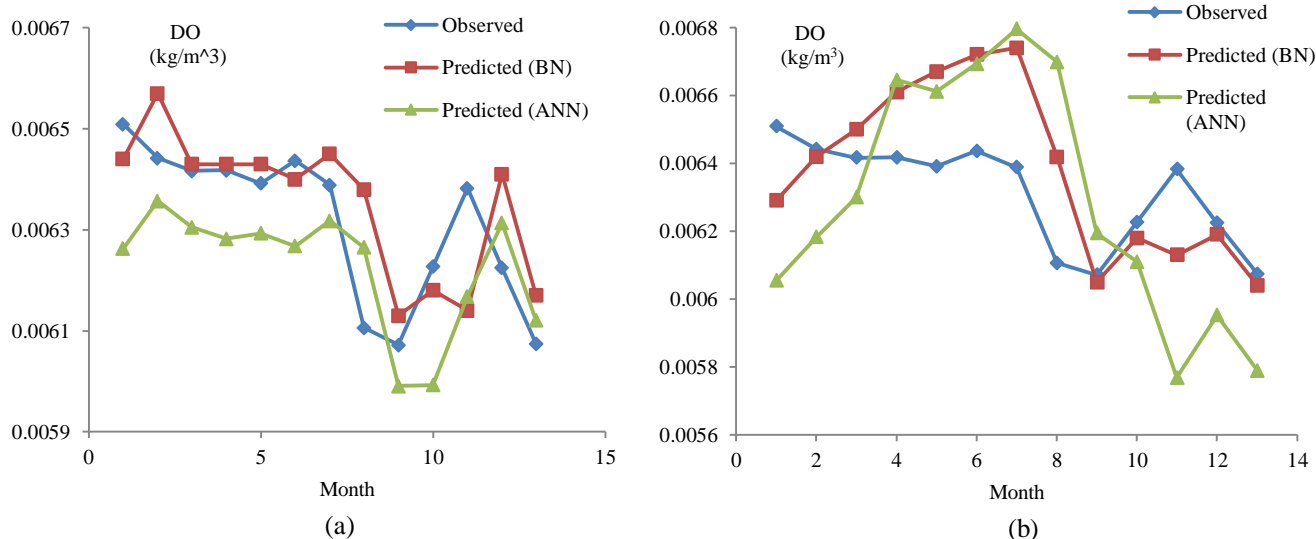
Figure 12. Time series of the monthly DO obtained through a) the first strategy and b) the second strategy during testing set

Figure. 12 demonstrates that applying the first strategy to develop forecasting DO model is superior over the second strategy. A relatively good agree between predicted and observed values of DO can be obtained for the ANN and BN models developed based on the first strategy. Overall, the BN models for both strategies overestimate the DO values while the ANN models underestimate for the first strategy and overestimate for the second one. The best BN model obtained from the first strategy provides an accurate prediction for average to high values of DO. For low values of DO, applying the BN model based on the second strategy can be helpful.

There was a number of incomplete time series or missed data (e.g. 6 months missing data of all the variables in the year 2014) that makes it more complex to give an acceptable prediction for water quality parameters. Anyhow, the proposed BN model showed a great ability dealing with time series including missing data. Therefore, BN models can be successfully applied for predicting time series with missing data, nonlinear problems and variables with interrelated and complex relationships.

5. Conclusion

In the current study, an attempt was made to performance of two types of data driven models including ANN and BN models for predicting water quality parameters. The monthly data such as salinity, temperature, DO and the amount of chlorophyll recorded in Honolulu, Pacific Ocean were applied for model development. The main purpose of this study was to develop models which can provide accurate prediction of the DO, salinity and temperature. Moreover, an attempt was made to investigate capability of the BN models in dealing with incomplete time series or missing data. Generally two types of forecasting models were developed in which one type of the models was constructed only using the previous time series of the target variable as input. The second type was implemented by combining other parameters in the present and previous time steps and also previous time series of the target variable as input parameters. Results of the different models were evaluated by using error measures.

Results of this study revealed that the proposed models can be successfully applied for water quality prediction. Overall, the BN models were found to be more efficient than the ANN models. Anyhow, for salinity prediction, the ANN model outperformed the BN model. The best models for temperature and salinity were obtained when the time series of three and two preceding months respectively applied as input parameters. The best BN models for DO forecasting had three previous time steps of DO in its input structure. It was found that the first strategy applied for DO prediction is more efficient than the second one which combines different other water quality parameters. For temperature and salinity forecasting, a very high correlation between observed and predicted values can be derived. The correlation coefficient of the best model for temperature and salinity has correlation coefficient higher than 0.9. Moreover, the RMSE and MAE for these models are low. For predicting average and high values of DO, the BN model provides an accurate prediction. On the other hand, the ANN model gives an acceptable prediction for low values of DO.

Finding of this study demonstrated that BN models are powerful tools dealing with incomplete time series or missing data. They have a great ability in finding the relationship between complex, nonlinear and interrelated parameters. Also, it was obtained during modelling procedure that BN models are strongly dependent on input variables and relationship defined among variables (structure learning). Generally, it can be concluded that an acceptable forecasting of water quality parameters in one month ahead of time can be achieved by using BN models and only previous time series of the same variable. This can be helpful in environmental management, ocean

monitoring programs and fishery sciences. It is suggested to investigate the effect of some other parameters such as significant wave height, wind speed and some other water quality parameters in order to increase the accuracy of the forecasting models of water quality parameters.

6. References

[1] Alizadeh, Mohamad Javad, and Mohamad Reza Kavianpour. "Development of Wavelet-ANN Models to Predict Water Quality Parameters in Hilo Bay, Pacific Ocean." Marine Pollution Bulletin 98, no. 1–2 (September 2015): 171–178. doi:10.1016/j.marpolbul.2015.06.052.

[2] Jayawardena, Amithirlgala W., and Feizhou Lai. "Time Series Analysis of Water Quality Data in Pearl River, China." Journal of Environmental Engineering 115, no. 3 (June 1989): 590–607. doi:10.1061/(asce)0733-9372(1989)115:3(590).

[3] Ahmad, Shamshad, Iqbal H Khan, and B.P Parida. "Performance of Stochastic Approaches for Forecasting River Water Quality." Water Research 35, no. 18 (December 2001): 4261–4266. doi:10.1016/s0043-1354(01)00167-1.

[4] Chau, Kwok-wing. "A Review on Integration of Artificial Intelligence into Water Quality Modelling." Marine Pollution Bulletin 52, no. 7 (July 2006): 726–733. doi:10.1016/j.marpolbul.2006.04.003.

[5] Said, Mohamed A., Makram A. Gerges, Ibrahim A. Maiyza, Maged A. Hussein, and Ahmed A. Radwan. "Changes in Atlantic Water Characteristics in the South-Eastern Mediterranean Sea as a Result of Natural and Anthropogenic Activities." Oceanologia 53, no. 1 (March 2011): 81–95. doi:10.5697/oc.53-1.081.

[6] Ömer Faruk, Durdu. "A Hybrid Neural Network and ARIMA Model for Water Quality Time Series Prediction." Engineering Applications of Artificial Intelligence 23, no. 4 (June 2010): 586–594. doi:10.1016/j.engappai.2009.09.015.

[7] Gazzaz, Nabeel M., Mohd Kamil Yusoff, Ahmad Zaharin Aris, Hafizan Juahir, and Mohammad Firuz Ramli. "Artificial Neural Network Modeling of the Water Quality Index for Kinta River (Malaysia) Using Water Quality Variables as Predictors." Marine Pollution Bulletin 64, no. 11 (November 2012): 2409–2420. doi:10.1016/j.marpolbul.2012.08.005.

[8] Arya, Farid Khalil, and Lan Zhang. "Time Series Analysis of Water Quality Parameters at Stillaguamish River Using Order Series Method." Stochastic Environmental Research and Risk Assessment 29, no. 1 (June 11, 2014): 227–239. doi:10.1007/s00477-014-0907-2.

[9] Gao, Hailong, Xin Qian, Ruibin Zhang, Rui Ye, Zhe Liu, and Yu Qian. "Bayesian Regularized Back-Propagation Neural Network Model for Chlorophyll-a Prediction: A Case Study in Meiliang Bay, Lake Taihu." Environmental Engineering Science 32, no. 11 (November 2015): 938–947. doi:10.1089/ees.2015.0164.

[10] Heddam, Salim, and Ozgur Kisi. "Extreme Learning Machines: a New Approach for Modeling Dissolved Oxygen (DO) Concentration with and Without Water Quality Variables as Predictors." Environmental Science and Pollution Research 24, no. 20 (May 30, 2017): 16702–16724. doi:10.1007/s11356-017-9283-z.

[11] Solgi, Abazar, Amir Pourhaghi, Ramin Bahmani, and Heidar Zarei. "Improving SVR and ANFIS Performance Using Wavelet Transform and PCA Algorithm for Modeling and Predicting Biochemical Oxygen Demand (BOD)." Ecohydrology & Hydrobiology 17, no. 2 (April 2017): 164–175. doi:10.1016/j.ecohyd.2017.02.002.

[12] Deng, Weihui, Guoyin Wang, and Xuerui Zhang. "A Novel Hybrid Water Quality Time Series Prediction Method Based on Cloud Model and Fuzzy Forecasting." Chemometrics and Intelligent Laboratory Systems 149 (December 2015): 39–49. doi:10.1016/j.chemolab.2015.09.017.

[13] Alizadeh, Mohamad Javad, Hosein Shahheydari, Mohammad Reza Kavianpour, Hamid Shamloo, and Reza Barati. "Prediction of Longitudinal Dispersion Coefficient in Natural Rivers Using a Cluster-Based Bayesian Network." Environmental Earth Sciences 76, no. 2 (January 2017). doi:10.1007/s12665-016-6379-6.

[14] Farmani, Raziyeh, Hans Jørgen Henriksen, and Dragan Savic. "An Evolutionary Bayesian Belief Network Methodology for Optimum Management of Groundwater Contamination." Environmental Modelling & Software 24, no. 3 (March 2009): 303–310. doi:10.1016/j.envsoft.2008.08.005.

[15] Kocadağlı, Ozan, and Barış Aşıkgil. "Nonlinear Time Series Forecasting with Bayesian Neural Networks." Expert Systems with Applications 41, no. 15 (November 2014): 6596–6610. doi:10.1016/j.eswa.2014.04.035.

[16] Madadgar, Shahrbanou, and Hamid Moradkhani. "Spatio-Temporal Drought Forecasting Within Bayesian Networks." Journal of Hydrology 512 (May 2014): 134–146. doi:10.1016/j.jhydrol.2014.02.039.

[17] Goharnejad, Hamid, Abolfazl Shamsai, and Seyed Abbas Hosseini. "Vulnerability Assessment of Southern Coastal Areas of Iran to Sea Level Rise: Evaluation of Climate Change Impact." Oceanologia 55, no. 3 (August 2013): 611–637. doi:10.5697/oc.55-3.611.

[18] Bishop, Christopher. "Neural Networks." Handbook of Neural Computation (January 1996). doi:10.1201/9781420050646.ptb6.

[19] Hagan, M.T., and M.B. Menhaj. "Training Feedforward Networks with the Marquardt Algorithm." IEEE Transactions on Neural Networks 5, no. 6 (1994): 989–993. doi:10.1109/72.329697.

[20] Ham, F.M., Kostanic, I. "Principles of neurocomputing for science and engineering". McGraw-Hill Higher Education (2000).

[21] Kabir, Golam, Solomon Tesfamariam, Alex Francisque, and Rehan Sadiq. "Evaluating Risk of Water Mains Failure Using a Bayesian Belief Network Model." European Journal of Operational Research 240, no. 1 (January 2015): 220–234. doi:10.1016/j.ejor.2014.06.033.

[22] Ismail, Mohamed A., Rehan Sadiq, Hamid R. Soleymani, and Solomon Tesfamariam. "Developing a Road Performance Index Using a Bayesian Belief Network Model." Journal of the Franklin Institute 348, no. 9 (November 2011): 2539–2555. doi:10.1016/j.jfranklin.2011.07.015.

[23] Pearl, Judea. "BAYESIAN INFERENCE." Probabilistic Reasoning in Intelligent Systems (1988): 29–75. doi:10.1016/b978-0-08-051489-5.50008-4.

[24] Alizadeh, Mohamad Javad. "Discussion on 'Gene Expression Models for Prediction of Longitudinal Dispersion Coefficient in Streams' by Sattar, A.M.A., Gharabaghi, B., 2015. Journal of Hydrology 524, 587–596." Journal of Hydrology (January 2017). doi:10.1016/j.jhydrol.2017.01.007.

[25] Basant, Nikita, Shikha Gupta, Amrita Malik, and Kunwar P. Singh. "Linear and Nonlinear Modeling for Simultaneous Prediction of Dissolved Oxygen and Biochemical Oxygen Demand of the Surface Water — A Case Study." Chemometrics and Intelligent Laboratory Systems 104, no. 2 (December 2010): 172–180. doi:10.1016/j.chemolab.2010.08.005.

[26] Han, Hong-Gui, Qi-li Chen, and Jun-Fei Qiao. "An Efficient Self-Organizing RBF Neural Network for Water Quality Prediction." Neural Networks 24, no. 7 (September 2011): 717–725. doi:10.1016/j.neunet.2011.04.006.

[27] He, Li-Ming (Lee), and Zhen-Li He. "Water Quality Prediction of Marine Recreational Beaches Receiving Watershed Baseflow and Stormwater Runoff in Southern California, USA." Water Research 42, no. 10–11 (May 2008): 2563–2573. doi:10.1016/j.watres.2008.01.002.

[28] Maier, Holger R., Ashu Jain, Graeme C. Dandy, and K.P. Sudheer. "Methods Used for the Development of Neural Networks for the Prediction of Water Resource Variables in River Systems: Current Status and Future Directions." Environmental Modelling & Software 25, no. 8 (August 2010): 891–909. doi:10.1016/j.envsoft.2010.02.003.

[29] Heddam, Salim. "Comment on 'Performance of ANFIS Versus MLP-NN Dissolved Oxygen Prediction Models in Water Quality Monitoring A. Najah & A. El-Shafie & O. A. Karim & Amr H. El-Shafie. Environ Sci Pollut Res (2014) 21:1658-1670.'" Environmental Science and Pollution Research 22, no. 5 (November 14, 2014): 3983–3983. doi:10.1007/s11356-014-3813-8.

[30] Najah, A., A. El-Shafie, O. A. Karim, and O. Jaafar. "Integrated Versus Isolated Scenario for Prediction Dissolved Oxygen at Progression of Water Quality Monitoring Stations." Hydrology and Earth System Sciences 15, no. 8 (August 29, 2011): 2693–2708. doi:10.5194/hess-15-2693-2011.