Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll *a* concentration in coastal waters

Shahaboddin Shamshirband, Ehsan Jafari Nodoushan, Jason E. Adolf, Azizah Abdul Manaf, Amir Mosavi & Kwok-wing Chau

Taylor & Francis
Taylor & Francis Group

Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll *a* concentration in coastal waters

Shahaboddin Shamshirband [a,b], Ehsan Jafari Nodoushan[c], Jason E. Adolf[d], Azizah Abdul Manaf[e], Amir Mosavi[f,g,h] and Kwok-wing Chau[i]

[a]Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam; [b]Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam; [c]Department of Civil Engineering, Bijar Branch, Islamic Azad University, Bijar, Iran; [d]Department of Biology, Monmouth University, West Long Branch, NJ, USA; [e]Faculty of Computing and Information Technology, University of Jeddah, Jeddah, Saudi Arabia; [f]Institute of Structural Mechanics, Bauhaus University Weimar, Weimar, Germany; [g]Institute of Automation, Obuda University, Budapest, Hungary; [h]Institute of Advanced Studies Koszeg, Koszeg, Hungary; [i]Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hong Kong, People's Republic of China

ABSTRACT

In this study, ensemble models using the Bates–Granger approach and least square method are developed to combine forecasts of multi-wavelet artificial neural network (ANN) models. Originally, this study is aimed to investigate the proposed models for forecasting of chlorophyll *a* concentration. However, the modeling procedure was repeated for water salinity forecasting to evaluate the generality of the approach. The ensemble models are employed for forecasting purposes in Hilo Bay, Hawaii. Moreover, the efficacy of the forecasting models for up to three days in advance is investigated. To predict chlorophyll *a* and salinity with different lead, the previous daily time series up to three lags are decomposed via different wavelet functions to be applied as input parameters of the models. Further, outputs of the different wavelet-ANN models are combined using the least square boosting ensemble and Bates–Granger techniques to achieve more accurate and more reliable forecasts. To examine the efficiency and reliability of the proposed models for different lead times, uncertainty analysis is conducted for the best single wavelet-ANN and ensemble models as well. The results indicate that accurate forecasts of water temperature and salinity up to three days ahead can be achieved using the ensemble models. Increasing the time horizon, the reliability and accuracy of the models decrease. Ensemble models are found to be superior to the best single models for both forecasting variables and for all the three lead times. The results of this study are promising with respect to multi-step forecasting of water quality parameters such as chlorophyll *a* and salinity, important indicators of ecosystem status in coastal and ocean regions.

1. Introduction

The development of suitable predictive models for chlorophyll concentration and salinity as water quality indicators in aquatic environments is a subject of ongoing concern due to its importance in aquatic life, environmental and ecological management (Liangliang & Daoliang, 2015). Chlorophyll concentration indicates the amount of phytoplankton, which plays an important role for fish and other marine life. It takes solar energy and provides a food supply for many different marine birds and animals. Chlorophyll *a* is a form of chlorophyll that is commonly used for water quality and ecological modeling. In addition to chlorophyll, water salinity and temperature, dissolved oxygen and turbidity are among the key water quality parameters. In estuaries and bays, water quality is a critical issue due to the high potential for anthropogenic pollution in runoff. Therefore, long-term and reliable forecasting models for these variables can be considered a key element in the sustainable use and management of marine food resources. Moreover, forecasts can serve as an early warning for environmental activities or events that may significantly degrade water quality.

Different variables can be named as water quality variables such as dissolved oxygen (DO), chlorophyll concentration, temperature, surface water temperature, salinity, turbidity, etc. Depending on the purpose of studies, these parameters can be modeled and simulated as individual water quality parameters or as a combined index of water quality (Mohammed & Abdulrazzaq, 2018) as well. Water quality modeling and forecasting are complex procedures due to nonlinear relationships among variables. In oceans and estuaries where tidal and wind waves

impact on surface waters, the complexity of the problem is further exacerbated. The water quality models can be categorized as conceptual, physical or empirical models (Tsakiris & Alexakis, 2012). In conceptually/physically based models, it is necessary to introduce mathematical relationships among variables. This approach requires an accurate knowledge of a large number of variables and their complex relationships. The traditional empirical models may suffer from insufficient accuracy for modeling and forecasting purposes. Over the two past decades, artificial intelligence (AI) based models have attracted the attention of many researchers due to their strong capability to recognize complex and nonlinear relationships among variables.

They have the advantage of employing only a few input variables to model or predict the target variable. Moreover, in artificial neural network (ANN) models, the mathematical relationship between input and target parameters is not defined by the user. A review of the application of different intelligence-based techniques such as ANN to model water quality can be found in Chau (2006). Ye and Cai (2009) used recurrent ANN to predict chlorophyll *a* concentration for seven days ahead in Three-Gorges Reservoir. They applied different types of datasets including hydrological, limnological and meteorological data to feed the model inputs. The results showed that the forecasting models are sensitive to different environmental input variables. AI-based models, especially ANN techniques, were frequently applied for forecasting purposes with streamflow and environmental processes thanks to easy implementation, low computational cost and suitable performance (Fotovatikhah et al., 2018; Motahari & Mazandaranizadeh, 2017; Olyaie, Banejad, Chau, & Melesse, 2015; Wang, Xu, Chau, & Lei, 2014; Zamanisabzi, King, Dilekli, Shoghli, & Abudu, 2018). They have been developed to predict a water quality index (Gazzaz, Yusoff, Aris, Juahir, & Ramli, 2012), monthly chemical oxygen demand concentration (Khalil, Awadallah, Karaman, & El-Sayed, 2012), daily water temperature, salinity and dissolved oxygen (Alizadeh & Kavianpour, 2015), etc. Nodoushan (2018) presented successful applications of ANN and Bayesian networks (BN) to forecast chlorophyll concentration on a monthly scale. The outputs showed that the BN models outperform ANN models. Regarding time series forecasting, ANN models have a flexible structure that can be combined with a wavelet transform. This feature increased their popularity because the wavelet-ANN model can predict the target variable with higher accuracy (Alizadeh, Kavianpour, Kisi, & Nourani, 2017; Heddam & Kisi, 2017). Generally, the water quality variables can be modeled individually or as water quality indices (WQIs). However, applying the water quality indices in comparison with the individual water quality data requires a wider range of datasets and parameters to be available. Also, for some special purposes, some parameters may be more valuable. Consequently, models with higher accuracy and reliability for this parameters should be developed.

Ensemble models gain advantages of multiple individual models, resulting in superior performance over the best single forecast models (DeChant & Moradkhani, 2011; Najafi & Moradkhani, 2016; Seo, Liu, Moradkhani, & Weerts, 2014; Shi et al., 2015). Moreover, the forecasting uncertainty can be decreased by including forecasts of several individual models with different models than using a single model. Barzegar, Adamowski, and Moghaddam (2016) combined forecasts of multi-wavelet-ANN models to establish an ensemble model. They employed the model for monthly forecasting of electrical conductivity (EC) in Aji-Chay River, Iran. The performance evaluation of the models indicated the superiority of the ensemble model to the individual models.

To date, there is no reported study that employs ensemble models for simulation and prediction of water quality in estuaries and coastal waters. The forecasting of water quality parameters for more than one time step can be of great interest for coastal ecosystem and environmental management. Therefore, the development of reliable models with acceptable accuracy for multi-step ahead projection of the variables in estuaries and seas is an interesting subject of study.

This study is mainly organized to take advantage of ensemble models to achieve reliable multi-day ahead forecasts of chlorophyll *a* concentration and salinity in Hilo Bay, on the Island of Hawaii in the Pacific Ocean. In this regard, forecasts of different individual wavelet-ANN models were combined by means of bagging and boosting ensemble techniques. Daily values of chlorophyll *a* and salinity with up to three lags were applied to predict the same variable up to three days in advance. Performance and reliability of the models were evaluated using error measures and uncertainty analysis. In section 2, basic concepts of ANN and wavelet transform specifications of the study area and the data and a short description of ensemble models are presented. Section 3 provides details of the proposed models. The results are discussed in section 4. Concluding remarks are presented in the last section.

2. Materials and methods

2.1. Artificial neural networks

An ANN in its usual form consists of three different layouts that are linked to each other sequentially, e.g. the

input is linked to the hidden layer and the hidden one is connected to the output layer. The layers are made of neurons that connect each layer to the following layer. Input and output neurons represent the input and target variables, respectively. During the training process, the back propagation in a repetitive process is applied to assign appropriate weights to each input variable or node in such a way the more influential input variables get higher weights or stronger connection with the following layer's nodes. In brief, the mathematical expression for an ANN is given as Equation (1) (Jeong & Kim, 2005):

$$O_k = g_2 \left[\sum_{j=1}^{M} w_{kj} g_1 \left(\sum_{i=1}^{N} w_{ji} x_i + w_{jo} \right) + w_{ko} \right] \quad (1)$$

where x_i, O_k are the input and output values in node i and node k, respectively; g_1 (log-sigmoid) and g_2 (linear) are the hidden and output layers' activation functions, respectively; and N and M are the input and hidden neuron numbers. The weights between connections are defined by w_{ji} (the input node i and the hidden node j) and w_{kj} (the hidden node j and the output node k). Also, biases for neurons are represented with w_{jo} (the jth node in the hidden layer) and w_{ko} (the kth node in the output layer).

2.2. Discrete wavelet transform (DWT)

The wavelet is a useful tool to extract frequency and time information from a signal. In general, two types of wavelets including discrete and continuous transforms are recognized, in which each can employ different wavelet functions. Each wavelet function has its specific characteristics. Therefore, employing different wavelet functions for time series decomposition can lead to sub-signals with different specifications. Translation and dilation are two components of any wavelet function. According to Partal (2009), for a given signal of $x(t)$, the continuous wavelet transform (CWT) with translation (a) and dilation (b) can be expressed as:

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \psi^* \left(\frac{t - b}{a} \right) x(t) \mathrm{d}t \quad (2)$$

where t represents the time, ψ and $*$ denote the mother wavelet (wavelet function) and conjugate complex function, respectively. The discrete wavelet transform (DWT) has advantages over the CWT due to its involving less computational effort and complexity. Therefore, it has been selected for time series decomposition in this study. Following Grossmann and Morlet (1984) the DWT is

defined in Equation (3):

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a_0^m}} \psi \left(\frac{t - nb_0 a_0^m)}{a_0^m} \right) \quad (3)$$

where integer values of m and n are applied as the dilation and transformation parameters (Alizadeh et al., 2017); $a_0 > 1$ and $b_0 > 0$ are called the dilation step and location parameter, respectively. When dealing with DWT, the wavelet function and also decomposition level should be chosen appropriately in order to achieve desired outputs.

2.3. Case study and data analysis

In this study, historical daily average water temperature and salinity data recorded at Hilo Bay, Hawaii were used for forecasting. For this purpose, the water quality buoy (WQB-04) operating in Hilo Bay was selected as the case study. WQB-04 (located at 19.7430N, 155.0814W) is among several buoys installed as part of the Pacific Islands Ocean Observing System. WQB-04 measures water quality data at 15 min intervals in the upmost meter of the water column ($z = -0.75$ m). Daily data of chlorophyll a concentration and salinity are obtained by averaging 96 values per day. For the modeling process, daily average data from 1 January 2012 to 31 December 2016 were applied. The quality of all the data in this study has been controlled and low-quality data or days with missed data records were omitted from the modeling procedure. In this regard, the chlorophyll a data have fewer numbers than the salinity data because there are more missing data for chlorophyll. The chlorophyll a concentration is measured by means of fluorescence (light re-emitted by chlorophyll molecules). Chlorophyll dissipates the absorbed light by driving photochemical energy conversion (photosynthesis) and emitting fluorescence radiation. The EC characteristic is applied as an indirect tool to measure salinity, since saltwater has higher values of EC than freshwater with no dissolved salt. The data are divided into subsets for calibration and verification purposes. Figure 1 illustrates position of the buoy in Hilo Bay and the Pacific Ocean.

Data normalization to a range of [0, 1] was carried out to transpose the input variables into the same data range as the activation functions. Table 1 gives the basic statistical characteristics of the daily data including average 'Mean', minimum 'Min', maximum 'Max', standard deviation S_d and skewness coefficient C_{sx}.

2.4. Ensemble modeling

Ensemble modeling is a suitable technique to decrease the bias and variance of the predictions. In the ensemble
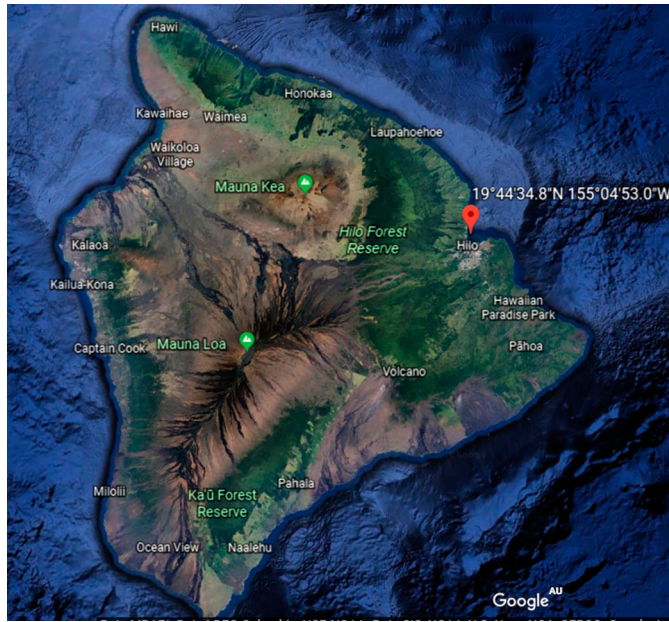
a)

b)

Figure 1. (a) The study area in Hawaii and (b) the water quality buoy (WQB).

Table 1. Statistical analysis of the whole dataset.

Variable	Mean	Min	Max	S_d	C_{sx}
Chlorophyll a (mg/m^3)	3.33	0.015	12.41	2.302	0.70
S (psu)	28.28	10.12	34.75	3.98	−1.26

model, forecasts of different individual models are combined to achieve the desired output. Regarding the variance and bias of predictions, bootstrap (bagging) or boosting ensemble methods can be employed. The bootstrap approach is more suitable for high-variance predictions whereas the boosting method is more efficient for high-bias predictions. In this study, two simple techniques including Bates–Granger and least square methods are applied to minimize the error of the forecasting models. To do this, forecasts of different independent models are combined to establish the ensemble model. The procedure employs an empirical formula/algorithm to assign appropriate weights to the independent models based on their variance of error/goodness of fit. In this research, a linear least square technique as well as the empirical technique proposed by Bates and Granger (1969) is employed to find optimum weights for independent models. Using the least square method and through a linear solution, the weights (w_i) are assigned in a way to achieve the least error (E). The general form of the least square method is given in Equation (4).

$$\sum_{i=1}^{k} w_i y_i - Y = E \tag{4}$$

where y_i = predictions, Y = measured values, and k = number of individual models. More details about ensemble methods and algorithms can be found in Zhou (2012).

In the Bates–Granger technique, weights of the individual models are found regarding the variance of the error of the models in the calibration period as per Equation (5):

$$w_i = \frac{1/\sigma_i^2}{1/\sigma_1^2 + .. + 1/\sigma_k^2} \tag{5}$$

where σ_i^2 represents the variance of error for model i over the calibration step. The main assumptions with this method are presuming a constant variance of residuals over time and unbiased forecasts for the models.

3. The proposed model and modeling procedure

To obtain acceptable forecasts of water temperature and salinity on a daily time scale, observed time series of the target variable up to three lags were decomposed using a discrete wavelet transform. The correct structure of the input variables indicating the number of lags was found using autocorrelation analysis. Several wavelet functions – Haar 'haar', discrete Meyer 'dmey', Coiflets 'coif1,coif5', Daubechies 'db2, db4, db6', Symlets 'sym4, sym10, sym20' – were employed for decomposition purposes. For the decomposition level of i, the DWT decomposes each data set to i subsets of details ($d_1, .., d_i$) and approximations ($a_1, \ldots, a_i$). Equation (6) was applied to

obtain a suitable decomposition level (Aussem, Campbell, & Murtagh, 1998):

$$L = \text{int}[\log(n)] \tag{6}$$

where L represents the wavelet decomposition level and n stands for number of data sample. Therefore, for this study, the decomposition level of 3 was recognized for data decomposition. Using the three levels of decomposition, six subsets of the data for each time series are obtained in which d_1, d_2, d_3 (high frequency component) and a_3 (high scale component) were incorporated as the input datasets in the forecasting models. Therefore, for each variable forecasting (water temperature or salinity) and also for each lead time $(t + 1/t + 2/t + 3)$, 10 different individual models based on the wavelet function were developed.

In the ensemble approach, each wavelet-ANN model is weighted based on variance of error (in the case of the Bates–Granger approach) and goodness of fit with the real data (for the least square technique). Therefore, models that have higher accuracy in forecasting gain higher values of weights. In cases of Bates–Granger approach, the weights are assigned using Equation (5). In the least square boosting ensemble model, the optimum weights (w_i) for the k individual models are achieved by solving a matrix of coefficients (Equation (7)). After finding the weights of the individual models in the ensemble forecast, the error of the model can be computed using Equation (4). It is noteworthy that some of the individual models with relatively low accuracy may be excluded in the ensemble model.

$$\begin{bmatrix} y_{11} & \cdots & y_{1k} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{nk} \end{bmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_k \end{pmatrix} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \tag{7}$$

where n, y and Y denote the number of data sample, forecasts of the individual models and the real data, respectively. Generally, the least-square-based ensemble model can be illustrated as shown in Figure 2.

To assess the performance of the individual and ensemble models, two statistical error measures of coefficient of determination (R^2) and root mean square error (RMSE) are used. These error measures can be defined as follows.

$$R^2 = \frac{\left(\sum_{i=1}^{n}(Y_i - \bar{Y})(y_i - \bar{y})\right)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - y_i)^2}{n}} \tag{9}$$

where Y_i, $\bar{Y}$, y_i and n denote observed water quality variable, mean observed value, predicted target variable and number of data, respectively.

For long-term forecasting, not knowing the exact values of input variables for predicting the target factor is a big problem. In this regard, the reliability of the long-term forecasting models due to using only past events of the input variables is expected to decrease remarkably.
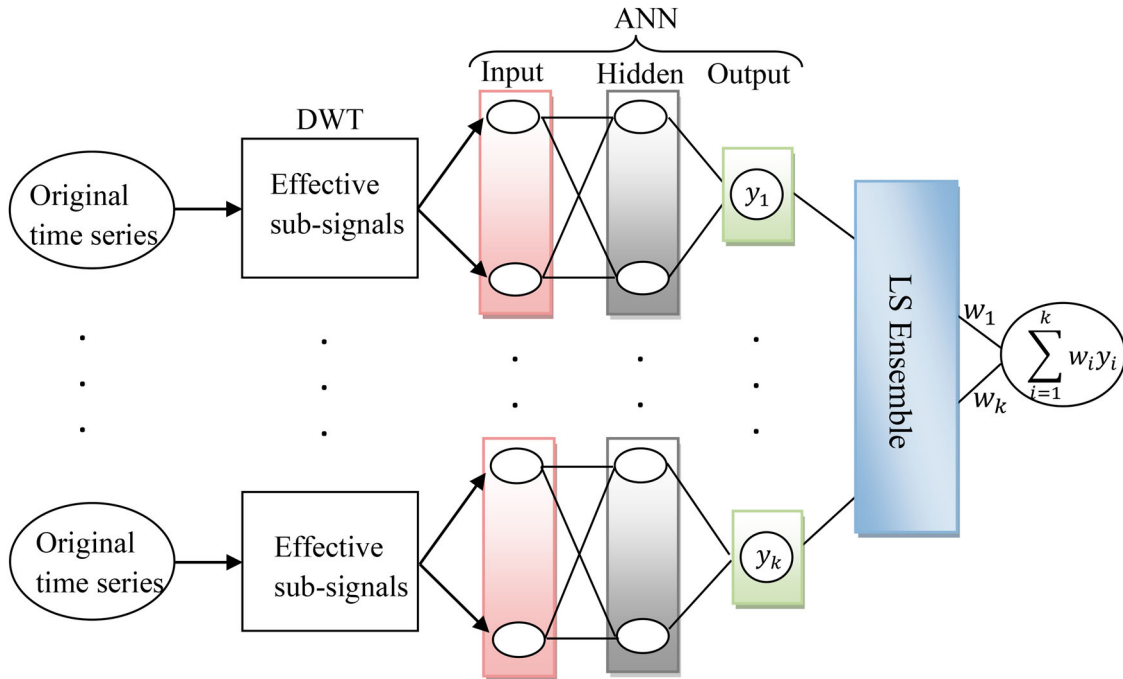

Figure 2. Schematic layout of the least-square-based ensemble wavelet-ANN model.

Therefore, uncertainty analysis conducted along with statistical error measures can be deemed a better indicator of the performance of the models. The accuracy expresses how precise the experiment is while the uncertainty is the component of a reported value which characterizes the range of values within which the real value is asserted to lie. Different sources of uncertainty can be recognized when dealing with data-driven and forecasting models, such as uncertainty in data sampling, tuning the correct network structure, considering important input variables, etc. However, in this study only a primary uncertainty analysis is carried out to find the range that the predictions are expected to lie in. For the uncertainty analysis, the mean ($\bar{e}$) and standard deviation of the error (S_e) are employed for the best individual model and for the ensemble model. The indices can be computed as per Equations (10 and 11):

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{10}$$

$$S_e = \sqrt{\sum_{i=1}^{n}(e_i - \bar{e})^2 / n - 1} \tag{11}$$

where $e_i = y_i - Y_i$, n = number of data sample. Using these indices, the width of uncertainty band for the 95% confidence level is written as $\pm 1.96 S_e$. It should be noticed that the results of the trial period only are reported in this study. The data for water temperature and salinity during trial period included the daily values from August 2015 to March 2016.

4. Results and discussion

4.1. Chlorophyll a concentration

To predict the chlorophyll a concentration up to three days in advance, historical time series of the same variable with three lags have been employed as input variables. Using different types of wavelet functions, 10 separate wavelet-ANN models were developed. Further, ensemble models including the Bates–Granger approach and least square technique were constructed to enhance the efficiency of the single models. The results for the trial period are presented in Table 2.

According to Table 2, the wavelet-ANN models have great capabilities to predict chlorophyll a up to three days in advance. Generally, the models provide acceptable predictions with high value of R^2 and low value of RMSE. The models with 'dmey' and 'sym20' wavelet functions are among the most efficient types of individual models. Shorter-period predictions ($t+1$) perform better than longer-period predictions ($t+2$, $t+3$).

Table 2. Results of the forecasting models for chlorophyll a concentration.

Model	$t+1$		$t+2$		$t+3$	
	R^2	RMSE (mg/m^3)	R^2	RMSE (mg/m^3)	R^2	RMSE (mg/m^3)
Haar	0.79	0.73	0.59	1.02	0.53	1.09
Dmey	0.93	0.40	0.85	0.61	0.75	0.78
coif1	0.87	0.56	0.71	0.84	0.70	0.85
coif5	0.93	0.40	0.82	0.65	0.73	0.84
db2	0.87	0.56	0.71	0.84	0.57	1.04
db4	0.87	0.56	0.79	0.72	0.60	1.01
db6	0.91	0.47	0.79	0.72	0.70	0.88
sym4	0.91	0.47	0.78	0.76	0.71	0.85
sym10	0.92	0.43	0.77	0.74	0.62	1.01
sym20	0.93	0.40	0.81	0.67	0.79	0.72
BG ensemble	0.96	0.31	0.85	0.59	0.81	0.69
LS ensemble	0.96	0.31	0.85	0.61	0.76	0.75

To establish the ensemble models, forecasts of the eight wavelet-ANN models (haar and coif1 were excluded in the ensemble models due to their relatively low performance) were combined to provide a better forecast of chlorophyll a concentration. As observed from Table 2, the Bates–Granger approach (BG ensemble) outperforms all the individual models (even the best single wavelet-ANN model) and the least square based ensemble model (LS ensemble). However, the results of the two ensemble methods are very close to each other. The forecasts of the BG ensemble models for all the three lead times have R^2 higher than 0.8, which implies good agreement among observed and forecasted values of the target variable. Moreover, the ensemble models have lower values of RMSE than the individual models. Therefore, the BG ensemble model is recognized as the most accurate model for chlorophyll a concentration forecasting up to three days in advance. As the models only use the chlorophyll historical data as input variables, the results are promising for the application of such methods to water quality and ecological modeling, especially in cases with limited data. Figures 3–5 depict scatter plot and forecasts of the BG ensemble model versus the observed values for one to three lead times.

From Figures 3–5, it can be derived that acceptable forecasts of chlorophyll a can be achieved using the BG ensemble model. The proposed model provides accurate forecasts for the first lead time. The forecasts approximately overlap the real values. For two other lead times, the forecasts still represent high accuracy. However, the models for the second and third lead times underestimate the extreme values of chlorophyll a. The figures demonstrate that a roughly sound prediction for chlorophyll concentration can be achieved even three days in advance, which is valuable for water quality monitoring and assessment. Moreover, the proposed model can be employed as a useful tool to estimate the missing data for chlorophyll.
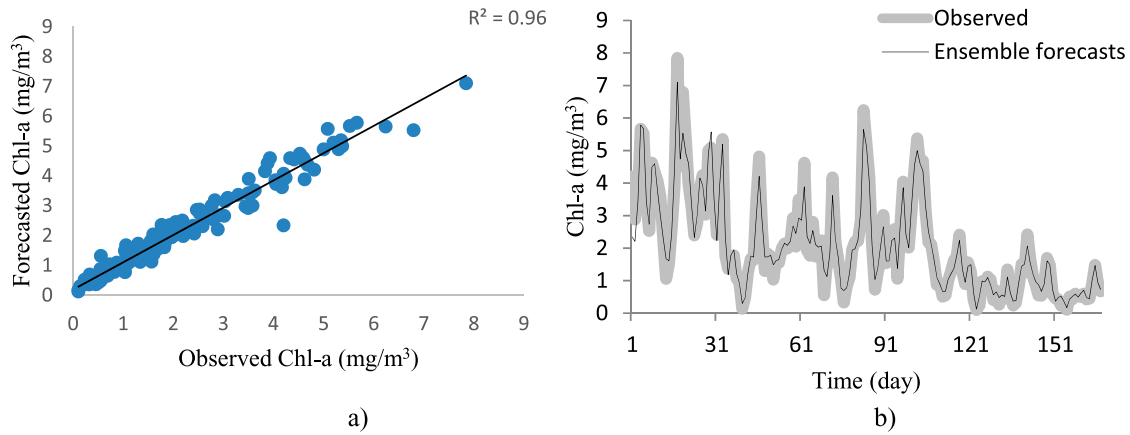
Figure 3. Ensemble-forecasted temperature against observed values: (a) scatter plot; (b) time series for $t + 1$.
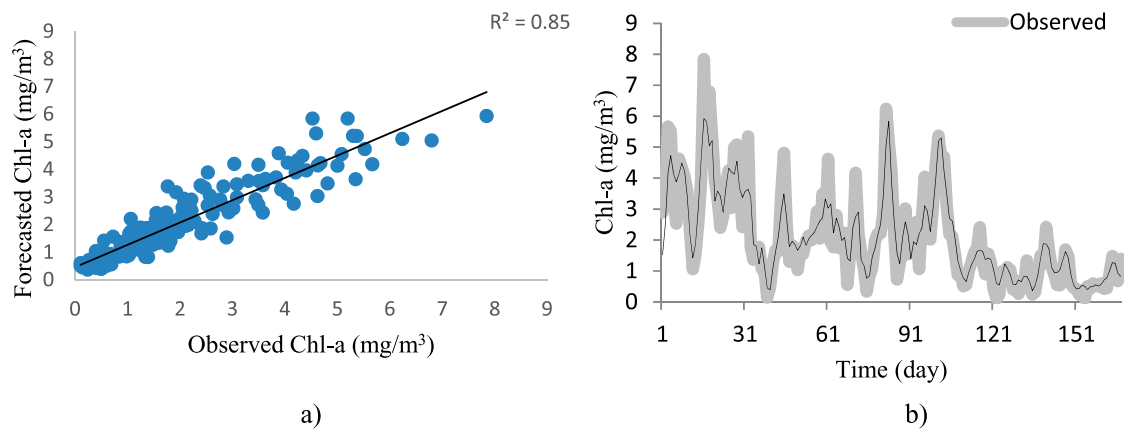

Figure 4. Ensemble-forecasted temperature against observed values: (a) scatter plot; (b) time series for $t + 2$.
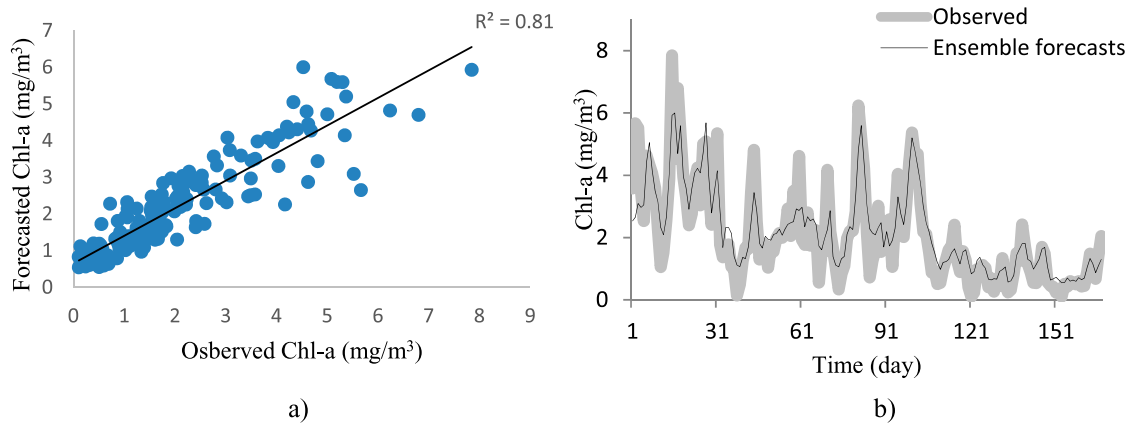

Figure 5. Ensemble-forecasted temperature against observed values: (a) scatter plot; (b) time series for $t + 3$.

4.2. Water salinity

To evaluate the validation and generalization of the results for chlorophyll *a*, the whole process has repeated for salinity. The water salinity up to three days in advance was forecasted by incorporating the historical time series of the salinity with three time lags. The results of the individual models as well as the ensemble models for the test data set are presented in Table 3.

As can be seen in Table 3, the individual models have good performance for first lead time forecasting. However, the models' efficiency decreases rapidly as they are applied to longer-period forecasting. For the third lead

Table 3. Results of the forecasting models for salinity.

Model	$t+1$		$t+2$		$t+3$	
	R^2	RMSE (psu)	R^2	RMSE (psu)	R^2	RMSE (psu)
Haar	0.78	1.79	0.61	2.39	0.30	3.35
Dmey	0.96	0.77	0.86	1.44	0.76	1.93
coif1	0.87	1.38	0.76	1.87	0.59	2.51
coif5	0.95	0.90	0.84	1.51	0.71	2.08
db2	0.86	1.42	0.75	2.05	0.75	1.92
db4	0.92	1.12	0.79	1.77	0.73	2.03
db6	0.92	1.09	0.78	1.83	0.74	2.05
sym4	0.91	1.14	0.78	1.83	0.68	2.22
sym10	0.94	0.95	0.82	1.61	0.80	1.80
sym20	0.95	0.91	0.79	1.81	0.77	1.89
BG ensemble	0.97	0.62	0.90	1.22	0.89	1.36
LS ensemble	0.98	0.57	0.89	1.30	0.87	1.39

time, no individual model provides an R^2 higher than 0.8. Therefore, an ensemble model can be helpful to improve the accuracy of forecasts. Comparing the results of the ensemble models with those of the individual models demonstrates that the ensemble models are able to increase the accuracy of the models significantly. The LS and BG ensemble models outperform the individual

models consisting of the best single wavelet-ANN model for all lead times. The superiority of the ensemble models to the individual models is more apparent when the models are employed to longer-period forecasting. In other words, the difference between the performance of the ensemble model and the individual models for three days ahead is greater than the difference between the ensemble and individual models for the next day ($t+1$). The ensemble models improve the performance of the best individual models in terms of R^2 on average about 2%, 4% and 9% for one, two and three lead times, respectively. A similar conclusion relating to RMSE can be derived. For the first and second lead times, the wavelet-ANN model embedded with the 'dmey' function gives the highest accuracy while for the third lead time the model using 'sym10' provides the best predictions. Therefore, there is not always a single model that is the best one and therefore ensemble models consisting of several individual models are advantages over the single models. Figures 6–8 illustrate scatter plot and time series of the ensemble forecasts and real values for one to three lead times.
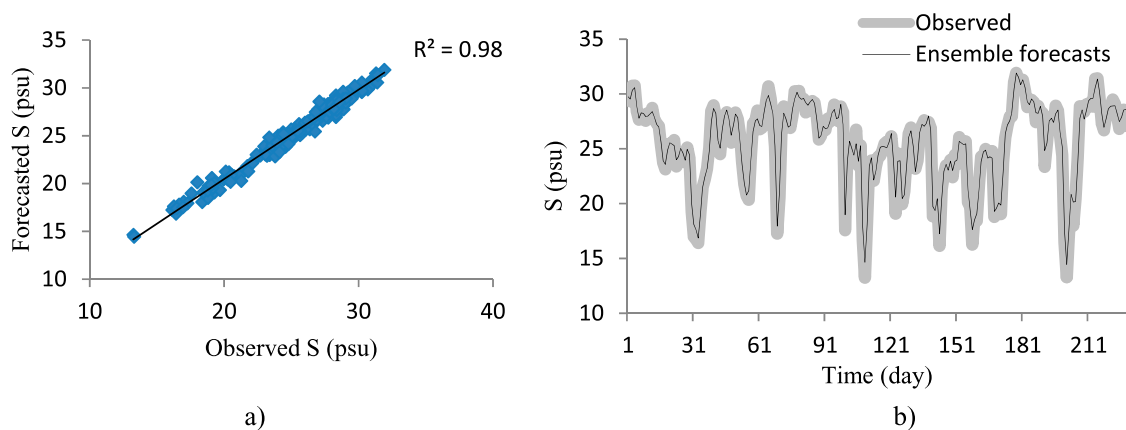
Figure 6. Ensemble-forecasted salinity against observed values: (a) scatter plot; (b) time series for $t+1$.
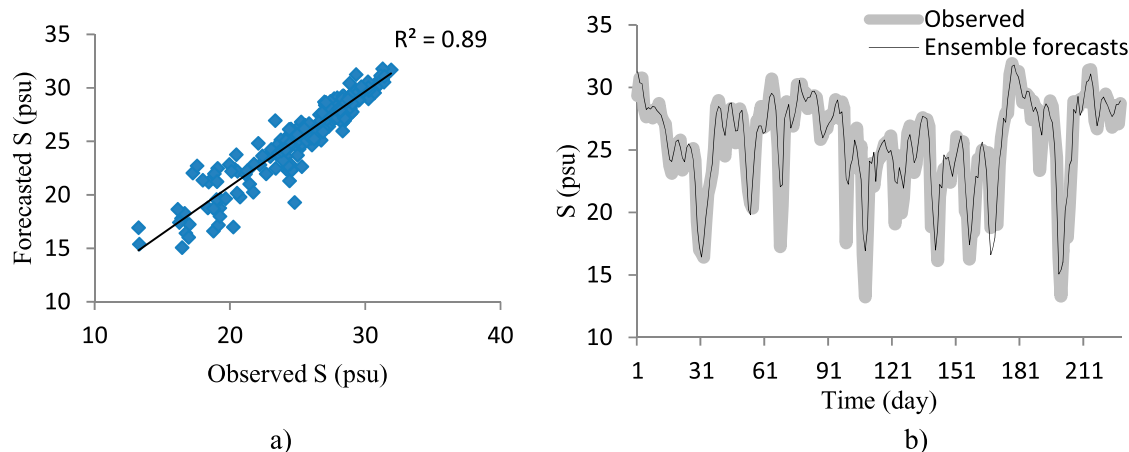
Figure 7. Ensemble-forecasted salinity against observed values: (a) scatter plot; (b) time series for $t+2$.
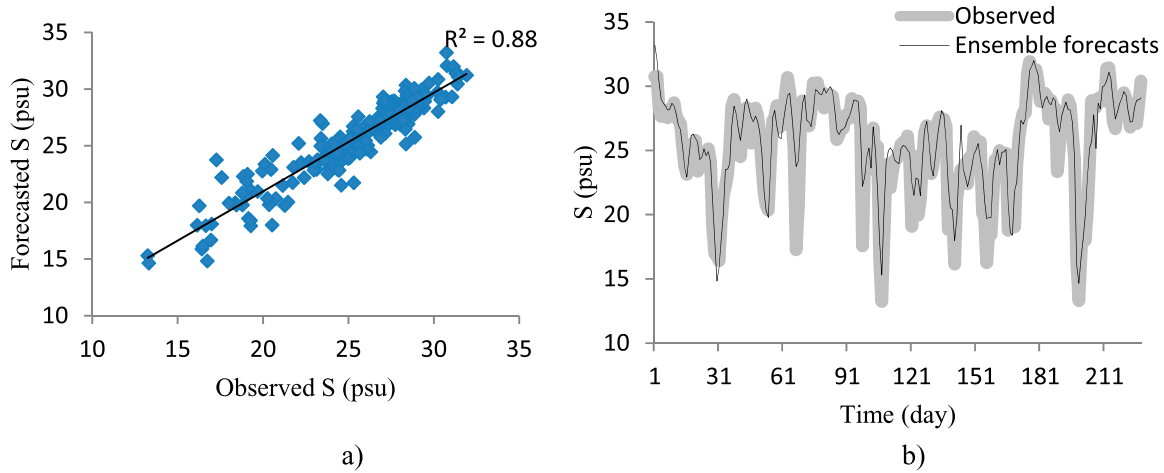
Figure 8. Ensemble-forecasted salinity against observed values: (a) scatter plot; (b) time series for $t+3$.

From Figure 6, it can be seen that the ensemble model provides acceptable forecasts of salinity for the whole range of the data in the first lead time. For the second and third lead times, the forecasts are generally in a good agreement with those of the observed data. However, the models overestimate low values of salinity. For medium and high values, the forecasts of the BG ensemble for one to three days ahead are relatively satisfactory. In general, acceptable forecasts of salinity up to three days ahead can be achieved when the ensemble model is employed. A comparison between the results of ensemble forecasting of salinity conducted by this study with those of the previous study in terms of R^2 indicates the superiority of the ensemble models. The current model gives an R^2 of 0.98 for the next day while it did not exceed 0.93 in the previous study using a single wavelet-ANN model (Alizadeh & Kavianpour, 2015).

4.3. Uncertainty analysis

Uncertainty analysis is a useful tool to investigate the reliability of forecasting models. Table 4 presents the results of the uncertainty analysis for the best single model and the ensemble model in terms of mean error $\bar{e}$ and the width of uncertainty band (for the 95% confidence level). Thus, the performance of the models can be evaluated from the points of view of accuracy and reliability. As the results of the BG ensemble and LS ensemble models are roughly similar to each other, the uncertainty analyses are conducted only for the best individual and BG ensemble models.

Generally, the ensemble models are superior to the best single model for each variable and lead time due to lower values of mean error and narrower width of uncertainty band. On the other hand, it is not always an easy task to find the best single model. Moreover, it requires

Table 4. Uncertainty analysis for the best single and ensemble models.

		Chlorophyll a		Salinity	
		Best single	Ensemble	Best single	Ensemble
$t+1$	$\bar{e}$	0.023	−0.004	0.090	0.028
	Width of uncertainty band	±0.742	±0.642	±1.473	±1.256
$t+2$	$\bar{e}$	0.020	0.016	0.057	0.034
	Width of uncertainty band	±1.305	±1.212	±2.623	±2.064
$t+3$	$\bar{e}$	0.026	0.058	0.185	0.035
	Width of uncertainty band	±1.506	±1.385	±2.937	±2.194

examination of the efficiency of many different models to obtain the most efficient one. Therefore, combining the forecasts of several individual models in ensemble models can be thought of as a useful way to enhance the accuracy of model forecasts. For any time series and case study, individual models may have good predictions for some points and other models may give better predictions for some other points. Therefore, ensemble models may outperform single models because they gain the advantages of several models.

Uncertainty analysis for different lead times reveals that with increasing lead time the uncertainty of the forecasts increases. For more lead times, the width of uncertainty band increases for both forecasting variables. The comparison of the results for chlorophyll a and salinity in Table 4 demonstrates that the forecasting models for salinity have a greater degree of uncertainty than the models for temperature. Negative and positive values for mean error in Table 4 represent the models' underestimation and overestimation, respectively. Low values of mean error may indicate a symmetry distribution of the

errors but should not be confused with very accurate predictions. For example, the mean error for the second lead time of chlorophyll *a* has a lower value than the mean error for the first lead time whereas the forecasts for the first lead time are more accurate. This is mainly because the errors with negative and positive signs neutralize each other when one computes mean error. Moreover, the width of the uncertainty band for the first lead time forecasts are significantly narrower than the third lead time forecasts of chlorophyll *a*. The models applied for salinity have higher values of uncertainty band for salinity than with chlorophyll *a* due to higher level of variation of salinity than of chlorophyll. The chlorophyll average value is 3.33 mg/m^3 where salinity has an average of 28.28 psu. However, more accurate forecasts are expected for salinity than for chlorophyll *a* because of higher R^2. The uncertainty band with 95% confidence level for chlorophyll *a* and salinity for one to three days ahead forecasting changes from 0.64 to 1.38 mg/m^3 and from 1.25 to 2.2 psu, respectively. These ranges with their average values imply an acceptable level of reliability and accuracy for the BG ensemble forecasts.

5. Conclusions

In this study, ensemble models based on least square and Bates–Granger techniques were constructed in order to achieve acceptable forecasts of the parameters in Hilo Bay, Hawaii. Multi-wavelet-ANN models consisting of several functions were employed to this end. The models were applied for chlorophyll *a* and salinity forecasting in coastal water up to three days in advance. To predict each variable, the historical data of the same variable in the past three days were incorporated as the model inputs.

Results of multi-wavelet-ANN models indicate that the wavelet functions of 'demy', 'sym20' and 'sym10' are among the most efficient types for this study. The best forecasting model is obtained for the 'dmey' wavelet-ANN, which gives an R^2 of 0.93 and 0.96 for chlorophyll *a* and salinity, respectively, in the next day ($t + 1$). The worst model relates to 'haar' wavelet-ANN with R^2 of 0.53 and 0.3 for chlorophyll *a* and salinity, respectively. With increasing forecasting time horizon, the accuracy and reliability of the forecasts decrease. A comparison between the models for the first and third lead times reveals that the R^2 of the best model varies from 0.97 to 0.86 for the temperature and from 0.96 to 0.80 for the salinity. The temporal variation of the models' performance (with lead time) declines more sharply for salinity than for temperature.

Comparing the forecasts obtained from different models demonstrates the superiority of the ensemble models. Also, the outperformance of the ensemble models over the individual models is enhanced with increasing lead time. Regarding the results of the forecasting models for chlorophyll *a* in the third lead time, the ensemble model improves the performance of the best single model in terms of R^2 from 0.75 to 0.81 and in terms of RMSE from 1.80 to 1.36 mg/m^3. For salinity on the third day, the performance of the best single model is improved from 0.8 to 0.87 in terms of R^2 and it is improved from 1.89 to 1.39 psu in terms of RMSE. A similar conclusion for two other lead times can be derived. Moreover, it was found that the ensemble models outperform the individual models with respect to the uncertainty of the predictions. The ensemble models for both chlorophyll *a* and salinity have a narrower width of uncertainty band than the best individual models. According to this study, it can be concluded that the ensemble models developed for chlorophyll *a* and salinity provide acceptable forecasts up to three days in advance. The proposed model is recommended to be applied for forecasting of other water quality parameters. Hence, its application for different areas with different characteristics can serve as a useful tool for environmental monitoring purposes. The consideration of new techniques for ensemble purposes could be subject of future studies.

ORCID

Shahaboddin Shamshirband 🆔 http://orcid.org/0000-0002-6605-498X

References

Alizadeh, M. J., & Kavianpour, M. R. (2015). Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Marine Pollution Bulletin*, *98*(1), 171–178.

Alizadeh, M. J., Kavianpour, M. R., Kisi, O., & Nourani, V. (2017). A new approach for simulating and forecasting the rainfall-runoff process within the next two months. *Journal of Hydrology*, *548*, 588–597.

Aussem, A., Campbell, J., & Murtagh, F. (1998). Wavelet-based feature extraction and decomposition strategies for financial forecasting. *Journal of Computational Intelligence in Finance*, *6*(2), 5–12.

Barzegar, R., Adamowski, J., & Moghaddam, A. A. (2016). Application of wavelet-artificial intelligence hybrid models for water quality prediction: A case study in Aji-chay river, Iran. *Stochastic Environmental Research and Risk Assessment*, *30*(7), 1797–1819.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Or*, *20*, 451–468.

Chau, K.-w. (2006). A review on integration of artificial intelligence into water quality modelling. *Marine Pollution Bulletin*, *52*(7), 726–733.

DeChant, C. M., & Moradkhani, H. (2011). Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. *Hydrology and Earth System Sciences*, *15*(11), 3399–3410.

Fotovatikhah, F., Herrera, M., Shamshirband, S., Chau, K.-w., Faizollahzadeh Ardabili, S., & Piran, M. J. (2018). Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work. *Engineering Applications of Computational Fluid Mechanics*, *12*(1), 411–437.

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. *Marine Pollution Bulletin*, *64*(11), 2409–2420.

Grossmann, A., & Morlet, J. (1984). Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, *15*(4), 723–736.

Heddam, S., & Kisi, O. (2017). 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*(20), 16702–16724.

Jeong, D.-I., & Kim, Y.-O. (2005). Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. *Hydrological Processes*, *19*(19), 3819–3835.

Khalil, B. M., Awadallah, A. G., Karaman, H., & El-Sayed, A. (2012). Application of artificial neural networks for the prediction of water quality variables in the Nile delta. *Journal of Water Resource and Protection*, *4*(06), 388–394.

Liangliang, G., & Daoliang, L. (2015). A review of hydrological/water-quality models. *Frontiers of Agricultural Science and Engineering*, *1*(4), 267–276.

Mohammed, S. E., & Abdulrazzaq, K. A. (2018). Developing water quality Index to assess the quality of the drinking water. *Civil Engineering Journal*, *4*(10), 2345–2355.

Motahari, M., & Mazandaranizadeh, H. (2017). Development of a PSO-ANN model for rainfall-runoff response in basins.

Case Study: Karaj Basin. *Civil Engineering Journal*, *3*(1), 35–44.

Najafi, M. R., & Moradkhani, H. (2016). Ensemble combination of seasonal streamflow forecasts. *Journal of Hydrologic Engineering*, *21*(1), 04015043.

Nodoushan, E. J. (2018). Monthly forecasting of water quality parameters within Bayesian networks: A case study of Honolulu, Pacific Ocean. *Civil Engineering Journal*, *4*(1), 188–199.

Olyaie, E., Banejad, H., Chau, K.-W., & Melesse, A. M. (2015). A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: A case study in United States. *Environmental Monitoring and Assessment*, *187*(4), 189.

Partal, T. (2009). River flow forecasting using different artificial neural network algorithms and wavelet transform. *Canadian Journal of Civil Engineering*, *36*(1), 26–38.

Seo, D.-J., Liu, Y., Moradkhani, H., & Weerts, A. (2014). Ensemble prediction and data assimilation for operational hydrology.

Shi, H., Li, T., Liu, R., Chen, J., Li, J., Zhang, A., & Wang, G. (2015). A service-oriented architecture for ensemble flood forecast from numerical weather prediction. *Journal of Hydrology*, *527*, 933–942.

Tsakiris, G., & Alexakis, D. (2012). Water quality models: An overview. *Eur. Water*, *37*, 33–46.

Wang, W.-c., Xu, D.-m., Chau, K.-w., & Lei, G.-j. (2014). Assessment of river water quality based on theory of variable fuzzy sets and fuzzy binary comparison method. *Water Resources Management*, *28*(12), 4183–4200.

Ye, L., & Cai, Q. (2009). Forecasting daily chlorophyll a concentration during the spring phytoplankton bloom period in Xiangxi Bay of the three-gorges reservoir by means of a recurrent artificial neural network. *Journal of Freshwater Ecology*, *24*(4), 609–617.

Zamanisabzi, H., King, J. P., Dilekli, N., Shoghli, B., & Abudu, S. (2018). Developing an ANN based streamflow forecast model utilizing data-mining techniques to improve reservoir streamflow prediction accuracy: A case study. *Civil Engineering Journal*, *4*(5), 1135–1156.

Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms*. London: CRC press.