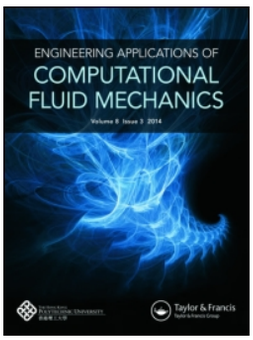

Engineering Applications of Computational Fluid Mechanics

ISSN: 1994-2060 (Print) 1997-003X (Online) Journal homepage: http://www.tandfonline.com/loi/tcfm20

# Effect of river flow on the quality of estuarine and coastal waters using machine learning models

Mohamad Javad Alizadeh, Mohamad Reza Kavianpour, Malihe Danesh, Jason Adolf, Shahabbodin Shamshirband & Kwok-Wing Chau

**To cite this article:** Mohamad Javad Alizadeh, Mohamad Reza Kavianpour, Malihe Danesh, Jason Adolf, Shahabbodin Shamshirband & Kwok-Wing Chau (2018) Effect of river flow on the quality of estuarine and coastal waters using machine learning models, Engineering Applications of Computational Fluid Mechanics, 12:1, 810-823, DOI: 10.1080/19942060.2018.1528480

**To link to this article:** https://doi.org/10.1080/19942060.2018.1528480



Published online: 10 Oct 2018.

ENGINEERING APPLICATIONS OF COMPUTATIONAL FLUID MECHANICS
2018, VOL. 12, NO. 1, 810–823
https://doi.org/10.1080/19942060.2018.1528480



# Effect of river flow on the quality of estuarine and coastal waters using machine learning models

Mohamad Javad Alizadeh[a], Mohamad Reza Kavianpour[a], Malihe Danesh[b], Jason Adolf[c], Shahabbodin Shamshirband [d,e] and Kwok-Wing Chau[f]

[a]Faculty of Civil Engineering, K. N. Toosi University of Technology, Tehran, Iran; [b]Faculty of Electrical and Computer Engineering, University of Science and Technology of Mazandaran, Behshahr, Iran; [c]Biology Department, Monmouth University, West Long Branch, NJ, USA; [d]Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam; [e]Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam; [f]Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hung Hom, Hong Kong

**ABSTRACT**

This study explores the river-flow-induced impacts on the performance of machine learning models applied for forecasting of water quality parameters in the coastal waters in Hilo Bay, Pacific Ocean. For this purpose, hourly recorded water quality parameters of salinity, temperature and turbidity as well as the flow data of the Wailuku River were used. Several machine learning models including artificial neural network, extreme learning machine and support vector regression have been employed to investigate the river-flow-induced impact on the water quality parameters from the current time up to 2 h ahead. Following the input structure of the machine learning models, two separate models based on including and excluding the river flow were developed for each variable to quantify the importance of the flow discharge on the accuracy of the forecasting models. The performance of different machine learning models was found to be close to each other and showing similar pattern considering accuracy and uncertainty of the forecasts. The results revealed that flow discharge influenced the water salinity and turbidity of the bay in which the models including the river flow as input variables had better performance compared with those excluding the flow time series. Among the water quality parameters investigated in this research, river flow made the most and least improvement on the efficiency of the models applied for forecasting of turbidity and water temperature, respectively. Overall, it was observed that water quality parameters can be properly forecasted up to several hours ahead providing a potentially valuable tool for environmental management and monitoring in coastal areas.

**ARTICLE HISTORY**
Received 21 May 2018
Accepted 22 September 2018

**KEYWORDS**
Water quality; river flow; machine learning; estuarine and coastal waters; salinity; turbidity

## 1. Introduction

Water quality parameters are important components to assess the health of the coastal environment and to guarantee suitable conditions for aquatic life. Estuarine and coastal waters are particularly susceptible to non-point/point source pollution conveyed by rivers and streams (Clark, 1995). These coastal areas are among the most important regions considering food supply and natural resources. Recently, anthropogenic pollution released in water bodies has been recognized as an important point of pollutants which necessitates serious attention to prevent drastic environmental problems. There are many estuaries that have been closed to commercial fishing due to pollution problems (Weiner and Matthews, 2003). Concerning an increasing demand on the use of estuarine waters, development of comprehensive water quality management programs is needed to evaluate conflicting uses of the estuary such as the discharge of wastewater, alterations of physiographic features, and alterations in the distribution and amount of freshwater inflow (Espey & Ward, 1972). Development of forecasting models of water quality parameters several hours ahead based on river flow can provide an early-stage alarm to prevent severe disaster in the coastal ecosystem by taking necessary actions in advance. Moreover, they can be employed as helpful tools for coastal monitoring purposes.

Water quality includes a wide variety of parameters that may be classified into three groups – biological, chemical and physical factors. In this study, three well-known physical properties of water quality including water temperature, salinity and turbidity are investigated. Sharp increase or decrease in these physical parameters can adversely affect water quality and microorganisms

**CONTACT** Shahabbodin Shamshirband shahaboddin.shamshirband@tdt.edu.vn

as well (e.g. high turbidity has side effects on flora and fauna). Water quality models can be constructed as physically based models or data-driven models. In the physical models, good knowledge of physics of the phenomena, relationships among different parameters (using mathematical descriptions) is mandatory. Chau and Jiang (2002) developed a three-dimensional numerical model of pollutant transport for the Pearl River Estuary. This model was applied for chemical oxygen demand (COD) distribution in the estuary and the results indicated the impact of pollutants especially during the wet season. Corbari, Lassini, and Mancini (2016) investigated intense short rainfall events on the water quality variables in the coastal waters. This study was carried out by means of remote sensing data including MODIS. The results showed that spatial and temporal water quality variation is dependent on the rainfall intensity and also on the distance from the shoreline. Data-driven models are becoming more popular due to their simplicity, ease of implementation and suitable performance. Dealing with water quality models, key elements of the water bodies can be designated as water quality indicators, or a combination of several water quality parameters can be formulated as water quality indices (WQIs). Several WQIs have been presented which consider different variables in the formulation. Gazzaz, Yusoff, Aris, Juahir, and Ramli (2012) used the WQI as a function of dissolved oxygen (DO), the concentration of suspended solids (SS), biochemical oxygen demand, COD, pH and ammonia nitrogen ($NH_3$-N). Sotomayor, Hampel, & Vázquez (2017) applied *k*-means classifying technique and a combined model of a KKN-genetic algorithm for water quality assessment in a river basin in southern Ecuador. In this study, a large number of water quality parameters have been taken under consideration. The results indicated that the efficiency of the employed techniques for water quality management in the river, especially when dealing with a complex dataset is required. Such models may require a large number of variables to estimate the WQIs. However, in some cases, these datasets are not available. Moreover, sometimes the models need to concentrate on some specific parameters. Therefore, development of water quality models based on individual important parameters rather than composite WQIs is a good alternative. Furthermore, some of the water quality parameters are mutually interrelated (e.g. water temperature with DO). Therefore, investigation of some particular parameters will suffice.

The data-driven (machine learning) models understand mathematical equations from analysis of concurrent input and output time series (Solomatine & Ostfeld, 2008). Artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS), support vector regression (SVR), extreme learning machine (ELM) and decision tree (D3) are some of the common machine learning techniques. The ANN and SVR are among the most widely used of machine learning methods. Recently, ELM has gained great popularity for time series and forecasting purposes in several fields. These models are being increasingly used for and forecasting purposes in hydrology, earth and environmental studies including water quality. Applications of different machine learning techniques such as genetic algorithm, artificial neural network and fuzzy inference system into water quality have been reviewed by K.-W. Chau (2006). ANN and ELM models have been employed for water quality forecasting in rivers and seas (Alizadeh & Kavianpour, 2015; Dogan, Sengorur, & Koklu, 2009; Nodoushan, 2018; Tomić, Antanasijević, Ristić, Perić-Grujić, & Pocajt, 2018; Wu, Wang, Chen, Cai, & Deng, 2018), for DO concentration modeling (Heddam & Kisi, 2017), for river discharge monitoring (Garel & D'Alimonte, 2017; Motahari & Mazandaranizadeh, 2017) and for analysis of chlorophyll dynamics (Tian, Liao, & Zhang, 2017). Fotovatikhah et al. (2018) provided a comprehensive survey on the computational intelligence applications in flood management systems. Yang et al. (2018) used SVR, ANN, principal component analysis methods for forecasting of water quality in Dianchi Lake. They investigated the spatial and temporal variations of lake water surface temperature and water quality. Barzegar, Moghaddam, Adamowski, and Ozga-Zielinski (2018) integrated different wavelet-ELM models through an ensemble process for multi-time-step-ahead forecasting of water quality. The results indicated the efficiency of the proposed technique. Other types of machine learning models such as ANFIS, genetic algorithm and ELM have been used in different fields of studies (Aghbashlo, Shamshirband, Tabatabaei, Yee, & Larimi, 2016; Chenar & Deng, 2018; Franco-Lopez, Ek, & Bauer, 2001; Jung, Popescu, Kelderman, Solomatine, & Price, 2010; Mohammadi, Shamshirband, Kamsin, Lai, & Mansor, 2016; Olyaie, Banejad, Chau, & Melesse, 2015; Wang, Xu, Chau, & Lei, 2014).

To date, no research study has been reported to explore river-flow-induced impacts on the water quality parameters in coastal and estuarine waters. Lack of enough data records of the river flow and water quality elements in coastal waters is a drawback for such studies. An interesting point related to the river flow impacts on the water quality in coastal waters is to consider the travel time of flow (or pollutant) from the river flow gauge to coastal waters and also its lability in the natural environment. In this regard, multi-time-step-ahead forecasting models can be of great importance. However, uncertainty about future events can reduce the efficiency of the long-term forecasting models.

Therefore, development of forecasting models with reliable outputs necessitates examining different data mining techniques.

The main objective of this study is to explore the effect of river flow on the performance of machine learning models for forecasting of water quality parameters in coastal and estuarine waters in Hilo Bay, Pacific Ocean. In this regard, surface water temperature, salinity and turbidity as water quality indicators are forecasted up to 2 h ahead. ANN, ELM, and SVR are employed to explore impacts of the river flow on the water quality parameters. The width of uncertainty band with 95% confidence level as well as root mean square error (RMSE) and coefficient of determination ($R^2$) and mean error is used to evaluate the performance of the models. The performance of different models is compared in terms of river flow impacts, time horizon, the forecasting technique and its time and space complexity procedure. A brief description of the methodology, datasets and study area and model development procedures are given in the next section. The results of the models are discussed in Section 3. Conclusions are presented in Section 4.

## 2. Materials and methods

### 2.1. Study area and data analysis

To explore the river flow impacts on the water quality parameters, the hourly flow data related to the Wailuku River entering Hilo Bay, Pacific Ocean were considered for the procedure. Moreover, the datasets of hourly water quality parameters including water temperature, salinity and turbidity in Hilo Bay were employed. These data were recorded by the Hilo Bay water quality buoy (HBB) moored within 1 m of the surface in one location within a small tropical estuary. The hourly data used in this study are average data of 15-minute records of the buoy. Quality of all data has been controlled and their accuracy has been validated. They have been provided by PacIOOS (Pacific Islands Ocean Observing System). The water quality measured by HBB is expected to be affected by freshwater inputs from the Wailuku River into Hilo Bay. The river is 45.1 km long and is the longest river in Hawaii. The coordinates for the buoy station (HBB) are 19.7430 N 155.0814 W, and the river station is USGS 16704000 Wailuku River at Piihonua, HI at 19.71214 N and 155.15080 W. The data cover hourly records from 2014 to 2016. Figure 1 illustrates the study area and Figure 2 shows the HBB components.

The datasets used in this study were recorded from January 2012 to December 2016 with 15-minute interval. Each hourly data are the average of four values. All the data have been normalized in a range of [0,1], as follows.

$$X'_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where $X'_i$ represents the normalized data and $X_i$, $X_{max}, X_{min}$ denote the observed data, maximum and minimum of the measured data, respectively. The statistical analysis including minimum 'Min', maximum 'Max', average 'Mean', standard deviation 'Sd' and skewness 'Skew' are given in Table 1 to give more details about the applied data.

The statistics show that the data have a wide range of variation. Especially, the flow data have high values of standard deviation that implies a high deviation of data from the average value.

### 2.2. Machine learning techniques

#### 2.2.1. Artificial neural networks

Feed forward neural networks (here called ANN) are a common type of artificial neural networks which applied in this study. An ordinary ANN model consists of input, hidden and output layers in which each layer has its nodes/neurons. Each layer is connected to the following layer via nodes. The nodes in the input layer which representing the input variables are transformed to hidden layer with weighted connections. The computations and processes are carried out in the hidden layer and then the nodes of the hidden layer are connecting to the output layer. Based on the relative importance of each input variable, appropriate weights between the connections of the nodes in the layer with those of the following layer are assigned.

Usually, the appropriate weights are determined through an iterative backpropagation algorithm in the training stage. Given $N$ distinct samples $(x_i, y_i)$, a single hidden layer ANN model with a linear activation function of output nodes in the general form can be expressed as (Huang, Zhu, & Siew, 2006):

$$o_j = \sum_{i=1}^{n} \beta_i g(w_i x_j + b_i), \quad j = 1, \ldots, N \tag{2}$$

where $x_j$ is the input value to node $j$, $o_j$ is the output at node $j$, $g$ is the hidden layer activation function (in this study means log-sigmoid) for the hidden layer, $b_i$ is the hidden layer bias, and $n$ is the number of nodes in the hidden layer. $w_i$ and $\beta_i$ are the weight between the input nodes and the $i$th hidden node, and the weight between the $i$th hidden node and the output nodes, respectively.

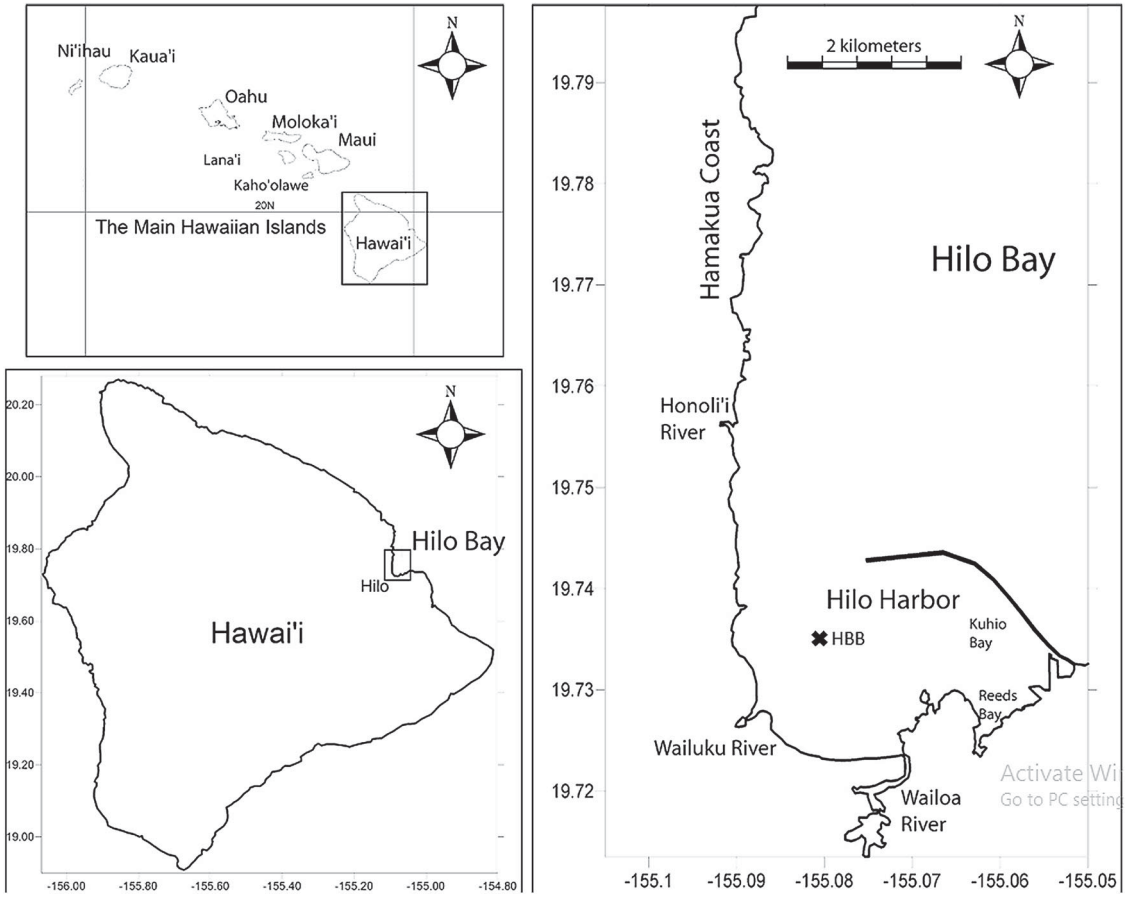


**Figure 1.** Study area in Hilo Bay, Hawai'i Island.

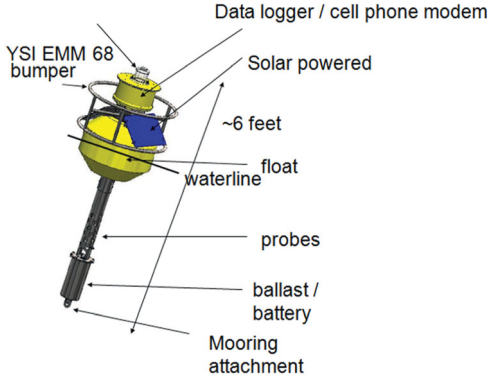


**Figure 2.** The Hilo Bay water quality Buoy (HBB) components.

**Table 1.** Data statistical analysis.

| Variable | Min | Max | Mean | SD | Skew |
|---|---|---|---|---|---|
| Salinity (ppt) | 5.127 | 35.695 | 28.25 | 4.54 | –1.28 |
| Turbidity (NTU) | 0 | 88.375 | 2.286 | 3.34 | 5.733 |
| Temperature (°C) | 18.082 | 29.5 | 24.996 | 1.403 | –0.360 |
| River flow ($m^3/s$) | 0.133 | 494.25 | 6.024 | 16.87 | 10.67 |

### 2.2.2. ELM

ELM proposed by Huang, Zhu, and Siew (2004) has become popular due to its faster implementation and better generalization compared to traditional ANNs. Unlike the gradient-based learning algorithm, ELM does not suffer from the stacking in local minima and over-fitting problems. Dealing with ELM, the input weights and hidden layer biases are chosen randomly while in ANN models, it is a time-dependent procedure due to determination through an iterative process. Also, in ELM the method, unlike the ANN, there is no need to tune all the parameters but determining the output weights analytically while choosing the input weights and hidden layer biases randomly. Assuming that there exist $w_i, \beta_i, b_i$ in which the target variable of $N$ sample $(Y_j)$ can be estimated with zero error (i.e. $Y_j = o_j$). Therefore Equation (3) can be rewritten in a compact form as:

$$H\beta = Y \tag{3}$$

where

$$H = \begin{bmatrix} g(w_1.x_1 + b_1) & \dots & g(w_n.x_1 + b_n) \\ \vdots & \dots & \vdots \\ g(w_1.x_N + b_1) & \dots & g(w_n.x_N + b_n) \end{bmatrix}_{N*n} \tag{4}$$

$$\beta = \begin{bmatrix} \beta_1^{\mathrm{T}} \\ \vdots \\ \beta_n^{\mathrm{T}} \end{bmatrix}_{n*m}, \quad Y = \begin{bmatrix} y_1^{\mathrm{T}} \\ \vdots \\ y_N^{\mathrm{T}} \end{bmatrix}_{N*m} \tag{5}$$

where $H$ is the hidden layer output matrix.

Working with ELM, the matrix $H$ can remain fixed once arbitrary values have been assigned to these parameters at the beginning of learning. Therefore, the model can be trained by finding a least-squares solution $\hat{\beta}$ for Equation (6).

$$\begin{aligned} &H(w_1, .., w_n, b_1, .., b_n)\hat{\beta} - Y \\ &= \min_{\beta} H(w_1, .., w_n, b_1, .., b_n)\beta - Y \end{aligned} \tag{6}$$

If $n = N$, the $H$ will be a square and invertible matrix and the training samples can be easily approximated with zero error. However, in real applications, it is usual to consider $n \ll N$, therefore, having a non-square matrix $H$, the exact solution may not exist. Therefore, a least square technique is used to obtain the solution.

$$\hat{\beta} = H^{\dagger} Y \tag{7}$$

where $H^{\dagger}$ is the Moore–Penrose generalized inverse of matrix $H$. A complementary introduction into ELM algorithm can be found in Huang et al. (2004) and Taormina and Chau (2015).

#### 2.2.3. SVR

Generally, support vector machines (SVMs) are a common type of data mining technique for classifying and regression purposes. They employ a hyperplane to separate data points of two categories. SVR is a type of support vector machine dealing with regression problems. Given a training dataset of $\{(x_i, y_i)\}_{i=1}^{n}$ where $n$ is the sample size and $x$ and $y$ represent input and output data, the method is applied to map the input space into an n-dimensional feature space using a non-linear function ($\varphi(x)$). Basically, the SVR function can be expressed as (Liu, Zhou, Chen, & Guo, 2014):

$$f(x) = (w.\varphi(x)) + b \tag{8}$$

where $w$ denotes the weight vector $w = \{w_1, \ldots, w_n\}$, and $b$ represents the bias. In SVR, the coefficients including the weight vector and the bias are estimated by defining a cost function. In the conventional regression models, the coefficients were obtained by minimizing square error while in SVR, they are determined using a new loss function known as the $\epsilon$-insensitive loss function (Liu et al., 2014).

$$L_{\varepsilon}(f(x), y) = \begin{cases} |f(x) - y| - \varepsilon \; for \; |f(x) - y| \geq \varepsilon \\ 0 \; \text{Otherwise} \end{cases} \tag{9}$$

where $L_{\varepsilon}$ is the loss function, $y$ is the target value, and $\varepsilon$ is the region of $\varepsilon$ insensitivity (defined by the user).

In SVR, the weight vector is derived using the regularized risk function as follows:

$$R_{reg} = C\frac{1}{n}\sum_{i=1}^{n} L_{\varepsilon}(f(x_i), y_i) + \frac{1}{2}w^2 \tag{10}$$

where $\frac{1}{2}w^2$ and $C$ are called regularization term and constant, respectively. The constant can be introduced by the user. Equation (10) can be rewritten as an optimization problem with the following cost function and constraints (Vapnik, 2013):

$$\text{minimize } \frac{1}{2}w^2 + C\sum_{i=1}^{n}(\xi_i, \xi_i^*) \tag{11}$$

$$\text{subject to } \begin{cases} y_i - (w.\varphi(x_i) + b) \leq \varepsilon + \xi_i \\ (w.\varphi(x_i) + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0, i = 1, \ldots, n \end{cases} \tag{12}$$

where $\xi_i$ and $\xi_i^*$ are the positive slack variables to measure the training samples' deviation outside the $\varepsilon$-insensitivity zone. Finally, the general form of the SVR regression function is formulated as (Vapnik, 2013):

$$f(x) = \sum_{i=1}^{n}(a_i - a_i^*)K(x, x_i) + b \tag{13}$$

where $a_i, a_i^* \geq 0$ are the Lagrangian multipliers that satisfy the equality $a_i a_i^* = 0$; and $K(x, x_i)$ is the kernel function(Liu et al., 2014). Different types of kernel function such as linear, Gaussian, polynomial, etc., can be employed in which selection of an appropriate type of the function is a mandatory step toward achieving suitable performance of SVR. Further details of SVM and SVR can be found in Vapnik (2013).

### 2.3. Procedures

To select and develop any data mining techniques for forecasting purposes, input selection, accuracy of the models, physical meaning of the relationship, size of the data in training stage and its homogeneity with the testing dataset and complexity of the models' structure have to be taken under consideration to guarantee efficiency and reliability of the models (Alizadeh, 2017). Different machine learning techniques have their own advantages

and disadvantage because of employing different formulations. Therefore, they are different in terms of structure, training procedures, learning time, etc. For instance, the dataset in traditional ANN models is usually divided into three subsets of training, validation and testing in which validation set is applied to control the model overfitting. On the other hand, the SVR and ELM need two datasets (no overfitting problem). In terms of learning time, the ELM and ANN are faster than the SVR technique. Moreover, they have different input structure but in this study, all the three techniques have been fed by the same input variables. In this study, an attempt made to the model and forecast water temperature, salinity and turbidity up to several hours ahead in Hilo Bay. To select the right input variables for each model, the correlation between each predictor with different lags and the target variable has been determined (Table 2).

As seen in Table 2, each water quality parameter is correlated mainly to its previous values. Moreover, an acceptable correlation between the river flow and water quality parameters especially salinity and turbidity are observed. A negative correlation between salinity and the river flow indicates freshwater intrusion to coastal and estuarine saline waters. The high correlation between the river flow and water turbidity in the estuary shows that the flow increases water turbidity due to making turbulence and spreading it. The river flow with 1 lag has the highest correlation with the water quality parameters in which implies the travel time (1 h) of flow from gauging station in the river to the Buoy in the Bay. Figure 3 illustrates a variation of water quality parameters and the flow time series for 18-day period in August 2014 in which the peak flow has happened in this time period. It is observed from the figure that turbidity and salinity fluctuations are in accordance with the flow discharge. Moreover, the peak flow has affected the water temperature. However, for the other flow values, the temperature is only slightly influenced by the river discharge.

This study investigates the efficiency of machine learning methods of ANN, ELM, and SVR to simulate and forecast the water temperature, salinity and turbidity in Hilo Bay for the current time ($t$) up to 2 h in advance ($t+2$). In each model, the same variable as the output variable with different lags (up to 3) is used as input variables. Moreover, the river flow data are included in the input structure of the developed models to forecast the target variable up to 2 h in advance. To provide more comparisons, separate models excluding the river flow data as an input variable are considered. In the model development, there are some parameters which need to

**Table 2.** Correlation analysis for water quality parameters.

| | Tem | | | Sal | | | Tur | | | River flow (Q) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | $t$ | $t-1$ | $t-2$ | $t$ | $t-1$ | $t-2$ | $t$ | $t-1$ | $t-2$ | $t$ | $t-1$ | $t-2$ |
| Tem ($t$) | 1 | 0.98 | 0.96 | 0.24 | 0.22 | 0.20 | −0.19 | −0.19 | −0.19 | −0.29 | −0.30 | −0.31 |
| Sal ($t$) | 0.24 | 0.22 | 0.21 | 1 | 0.96 | 0.93 | −0.54 | −0.54 | −0.54 | −0.58 | −0.59 | −0.60 |
| Tur ($t$) | −0.19 | −0.18 | −0.17 | −0.54 | −0.52 | −0.51 | 1 | 0.94 | 0.88 | 0.66 | 0.67 | 0.65 |

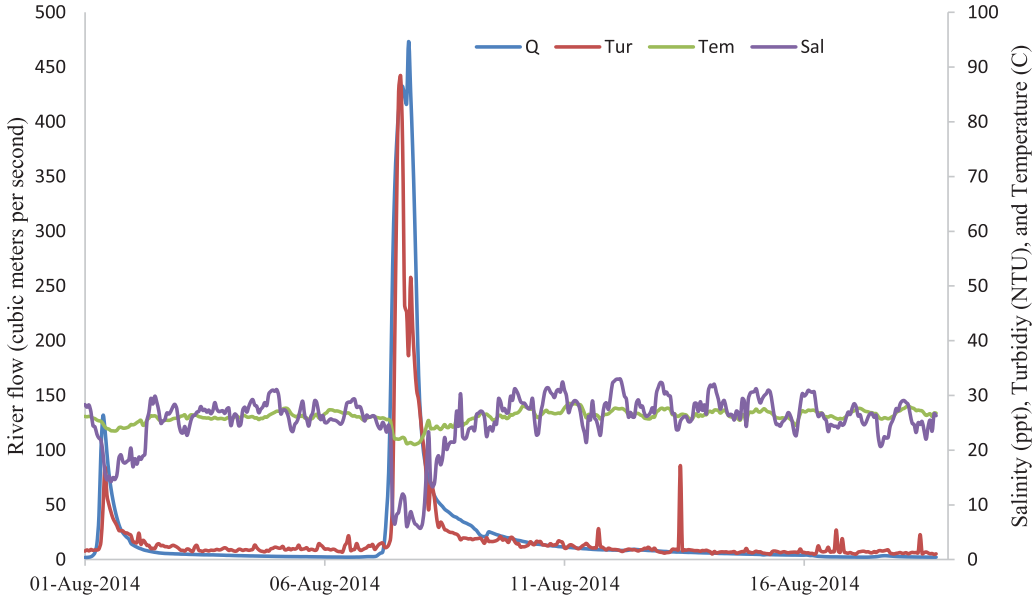


**Figure 3.** Time series of water quality parameters and flow discharge in August 2014.

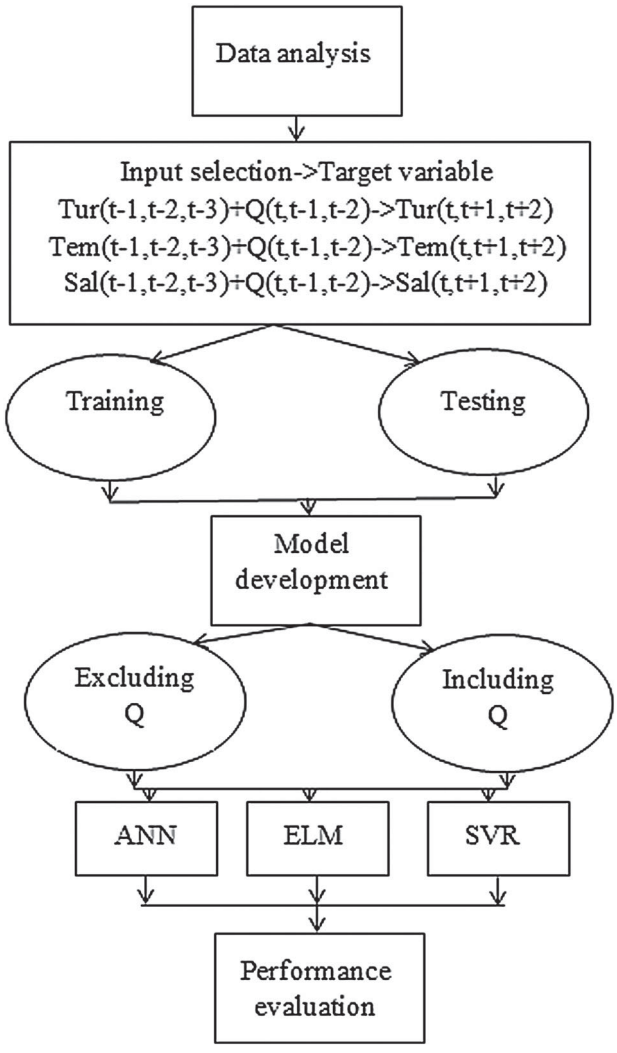


**Figure 4.** Schematic layout of the research study.

be tuned accordingly to reach the desired performance. In the ANN, the number of neurons in the hidden layer was 10, with the Levenberg–Marquardt algorithm. The activation function of ELM was set as 'sigmoid'. The linear kernel function for SVR is used. Other important characteristics of the models were set as default. The main steps of the study can be schematically illustrated as Figure 4.

The performance of the models is measured using three indices of the coefficient of determination ($R^2$), *RMSE* and width of uncertainty band ($\pm 1.96 S_e$). In Equation (26), $e_i = O_i - y_i$ and $\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i$ represent the prediction error and mean error, respectively.

$$R^2 = \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(O_i - \bar{O})\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (14)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - O_i)^2}{n}} \quad (15)$$

$$S_e = \sqrt{\sum_{i=1}^{n}(e_i - \bar{e})^2 / n - 1} \quad (16)$$

where $n$ is the number of data.

## 3. Results and discussion

### 3.1. Turbidity

Table 2 shows that river flow has the highest correlation with turbidity compared to salinity and temperature. As observed, the turbidity in the current hour has the highest correlation with the river discharge in the previous hour which can be due to travel time between hydrometric station recording the flow data and the water quality buoy. Moreover, there is a remarkable dependency of turbidity with those of flow data in other time horizons including *t*, and *t–1*. Therefore, the machine learning models including the flow data were constructed using the flow data in *t*, *t–1*, and *t–2*. These models were employed to predict turbidity up to 2 h ahead. The results are presented in Table 3. The models excluding and including the river flow as input variables are denoted with '–Q' and '+Q' respectively.

Regarding Table 3, it can be obtained that the models including the river flow (regardless of the machine learning type) provide more accurate forecasts of turbidity compared with those of excluding the river flow in the input structure. Therefore, it can be derived that the river flow can affect the turbidity of such large water bodies remarkably. Effects of the river flow on the turbidity in $t+1$ and $t+2$ indicate that it takes some hours to recover disturbance and turbidity caused by the river flow in the previous hours. Following the performance of the models at different times, it can be seen that the models' efficiency decreases as the time horizon increases. However,

**Table 3.** Results of machine learning models for the turbidity in the test period.

| | | ANN | | ELM | | SVR | |
|---|---|---|---|---|---|---|---|
| Time | | −Q | +Q | −Q | −Q | −Q | +Q |
| *t* | *RMSE* | 1.04 | 0.98 | 1.517 | 1.394 | 1.49 | 1.48 |
| | $R^2$ | 0.877 | 0.892 | 0.871 | 0.892 | 0.876 | 0.877 |
| | $1.96S_e$ | 2.916 | 2.739 | 2.97 | 2.73 | 2.93 | 2.917 |
| *t+1* | *RMSE* | 1.518 | 1.261 | 2.033 | 1.81 | 2.022 | 1.919 |
| | $R^2$ | 0.78 | 0.83 | 0.772 | .818 | 0.778 | 0.8 |
| | $1.96S_e$ | 3.898 | 3.434 | 3.982 | 3.544 | 3.947 | 3.749 |
| *t+2* | *RMSE* | 2.329 | 2.0719 | 2.304 | 2.054 | 2.324 | 2.293 |
| | $R^2$ | 0.706 | 0.769 | 0.707 | 0.765 | 0.704 | 0.711 |
| | $1.96S_e$ | 4.555 | 4.057 | 4.512 | 4.02 | 4.544 | 4.485 |

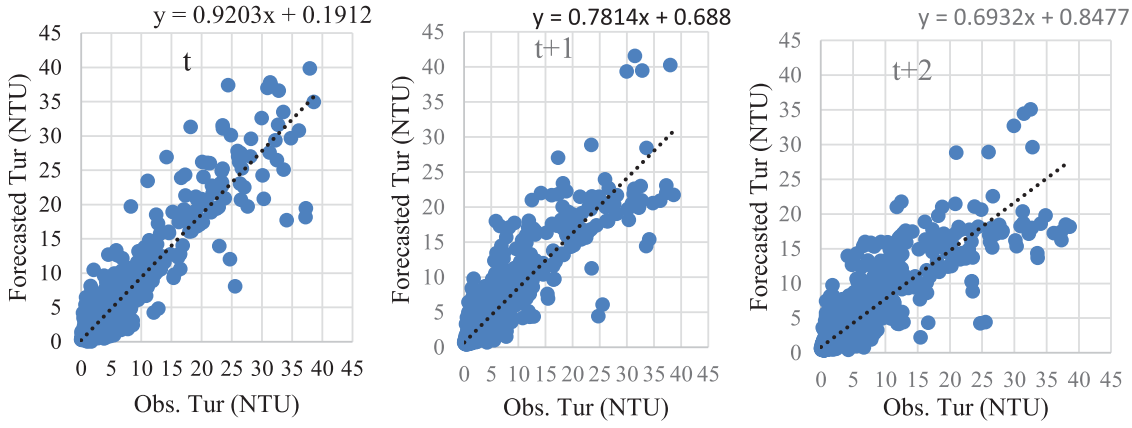


**Figure 5.** Scatterplots of turbidity for the ANN models.

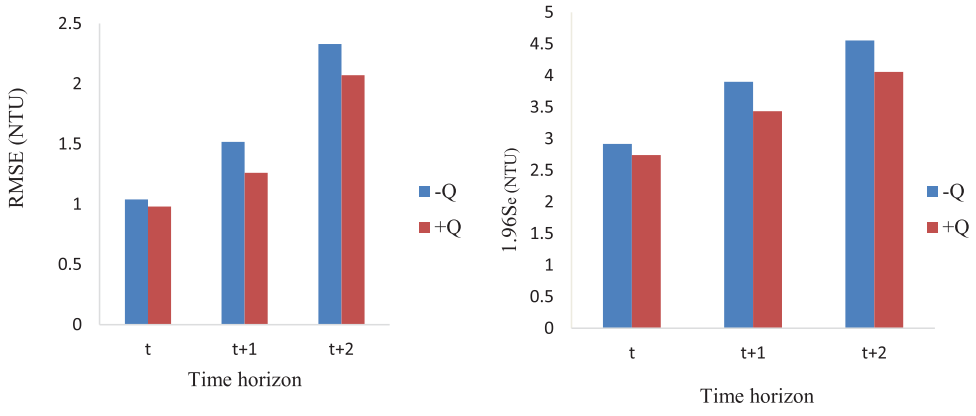


**Figure 6.** RMSE and uncertainty band of turbidity forecast during the test period.

the results for all three times investigated in this study are acceptable. Considering the performance of different machine learning models, it can be found that all three types of the models (ANN, ELM, and SVR) behave in a similar manner and the error measures for all of them are close to each other. Therefore, only forecasts of one model (the ANN) for different times are depicted against those of the observed data during the test period (Figure 5). As the models including the river flow have a higher coefficient of determination, here only their scatterplots are illustrated.

Generally, good agreement was obtained between observed and forecasted data. As the time horizon increases, the correlation between observed and forecasted data decreases. In Figure 5, for time $t + 1$ and $t + 2$, the data have more deviation from the linear trend line. In time $t$, the extreme values have been fairly forecasted while for times $t + 1$ and $t + 2$, the values have been generally underestimated. Comparison of the RMSE and $1.96S_e$ for different times and the ANN models with and without the river flow inputs is shown in Figure 6.

Figure 6 implies that the model for time t has the best accuracy and the uncertainty band for its output is narrower than the other times. As the time horizon increases, the uncertainty band is going to increase with a sharper rate. Therefore, the models applied for multi-hour-ahead forecasting may embed a high amount of uncertainty.

### 3.2. Water salinity

Among the variables considered in this study, it was found the water salinity in the Hilo Bay is mostly dependent on its values in the previous time steps and also on the river flow. Salinity is inversely correlated with those of the river flow with the highest correlation in *t–2*. Therefore, the water salinity in previous time steps, as well as the river flow (in the case), were applied to forecast the salinity in $t$, $t + 1$ and $t + 2$. Table 4 gives the results of the different models including and excluding the river flow up to 2 h ahead.

Results presented in Table 4 suggest that the models including the river flow slightly outperform the models without the flow inputs. Higher values of correlation, lower values of RMSE and also uncertainty band indicate the superiority of the models in which gain the flow data in their input structures. The forecasted time series of the ANN models during the test period are plotted versus the real values in Figure 7.

**Table 4.** Results of machine learning models for the salinity.

| | | ANN | | ELM | | SVR | |
|---|---|---|---|---|---|---|---|
| Time | | −Q | +Q | −Q | +Q | −Q | +Q |
| *t* | *RMSE* | 1.299 | 1.275 | 1.29 | 1.28 | 1.296 | 1.288 |
| | $R^2$ | 0.90 | 0.903 | 0.90 | 0.902 | 0.9 | 0.902 |
| | $1.96S_e$ | 2.539 | 2.493 | 2.528 | 2.497 | 2.535 | 2.515 |
| *t* + *1* | *RMSE* | 1.812 | 1.789 | 1.821 | 1.783 | 1.816 | 1.794 |
| | $R^2$ | 0.805 | 0.815 | 0.806 | 0.817 | 0.806 | 0.807 |
| | $1.96S_e$ | 3.549 | 3.463 | 3.537 | 3.432 | 3.54 | 3.48 |
| *t* + *2* | *RMSE* | 1.856 | 1.815 | 2.09 | 2.02 | 2.082 | 2.031 |
| | $R^2$ | 0.75 | 0.76 | 0.748 | 0.767 | 0.746 | 0.763 |
| | $1.96S_e$ | 4.053 | 3.934 | 4.046 | 3.87 | 4.063 | 3.96 |

According to Figure 5, the model provided a good forecast for all range of the salinity data. A high correlation can be found and all the points converge the trend line. For the other times, there are some points distracted from the line. However, the general performance of the models regardless of the time horizon is acceptable. Results related to the RMSE and uncertainty band of the forecasted target variable during the test period are presented in Figure 8.

Figure 8 reveals that the models with the flow have lower values of RMSE and uncertainty. The models' performance deteriorates as the time horizon grows. However, the forecasts for times *t* + *1* and *t* + *2* are still reliable. Generally, it can be found that the machine learning models such as ANN, ELM, or SVR have a suitable capability to forecast water salinity in coastal and estuarine waters several hours ahead.

### 3.3. Surface water temperature

According to correlation analysis, the surface temperature in the current hour (*t*) is mostly correlated to the temperature in the preceding time steps. However, its relation with the river flow is not very high. The main reason for this inverse correlation can be lower, colder temperature of the river compared with warmer the estuarine water. The low correlation may be due to the fact that the amount of river flow entering in the estuary is not remarkable in comparison to the estuarine water volume. As this study is concentrated on the river-induced impacts, two sets of models including and excluding the flow as input parameter have been constructed. Therefore, the input structure of the forecasting models for temperature consists of its values in preceding hours (*t–1, t–2, t–3*) and flow time series from the current time (*t*) up to 2 h ago (*t–2*) (in case of including model) are considered. The results are given in Table 5.

Regarding Table 5, the surface water temperature can be efficiently forecasted for several hours ahead. The

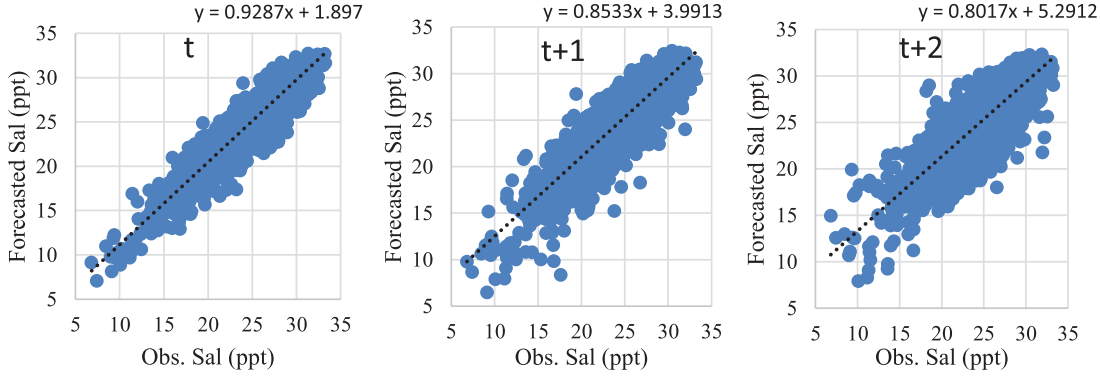


**Figure 7.** Scatterplots of salinity for the ANN models.

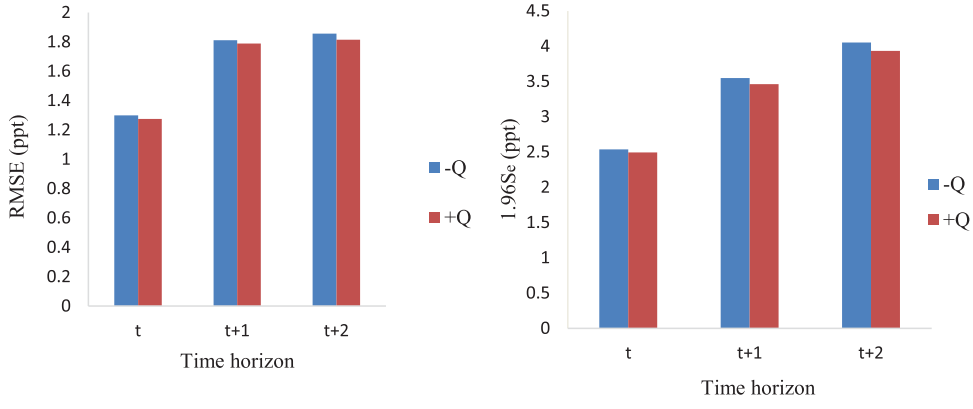


**Figure 8.** RMSE and uncertainty band of salinity forecast during test period.

**Table 5.** Results of machine learning models for the surface temperature.

| | | ANN | | ELM | | SVR | |
|---|---|---|---|---|---|---|---|
| Time | | −Q | +Q | −Q | +Q | −Q | +Q |
| *t* | *RMSE* | 0.289 | 0.286 | 0.287 | 0.289 | 0.288 | 0.287 |
| | $R^2$ | 0.95 | 0.951 | 0.95 | 0.95 | 0.95 | 0.95 |
| | $1.96S_e$ | 0.563 | 0.559 | 0.562 | 0.564 | 0.564 | 0.58 |
| *t* + 1 | *RMSE* | 0.356 | 0.35 | 0.401 | 0.399 | 0.347 | 0.352 |
| | $R^2$ | 0.903 | 0.905 | 0.903 | 0.905 | 0.928 | 0.927 |
| | $1.96S_e$ | 0.783 | 0.776 | 0.784 | 0.777 | 0.677 | 0.681 |
| *t* + 2 | *RMSE* | 0.47 | 0.46 | 0.468 | 0.464 | 0.47 | 0.466 |
| | $R^2$ | 0.868 | 0.873 | 0.867 | 0.873 | 0.87 | 0.878 |
| | $1.96S_e$ | 0.912 | 0.898 | 0.916 | 0.9 | 0.914 | 0.899 |

relatively high coefficient of determination, low values of RMSE and uncertainty confirm good agreement of the forecasted values against the real values. The results show that for time *t*, the uncertainty in temperature forecasting is about 0.5*C* in which it indicates the reliability and efficiency of the models. Moreover, the models can predict the temperature for 2 h ahead with adequate accuracy and reliability. The uncertainty band for time *t* + *2* is less than unity which confirms high performance of the applied models for temperature forecasting. The results are promising for water temperature monitoring and environmental management in coastal and estuarine waters. Figure 9 illustrates the scatterplots of the ANN models for temperature for the test dataset.

In Figure 9, there is a high similarity between forecasted and observed values of surface temperature for all three times under consideration. The forecasts are roughly close to 1:1 line which indicates the accuracy of the models. For more illustrations, Figure 10 presents the results for RMSE and width of uncertainty band of the forecasting models during the test period. In the figure, the RMSE and $1.96S_e$ have relatively low values which imply the accuracy and reliability of the developed models for temperature forecasting. Moreover, the models including the river flow have a bit better performance than those of excluding the flow as an input variable. However, the difference is negligible. The point is that it takes more time for heat transfer among the water layers in the ocean and coastal waters. Moreover, the amount of the water released by the river flow in compare with such a huge water body is not remarkable.

In general, this study shows that the machine learning models can be successfully applied to forecast the water quality parameters in coastal and estuarine water up to a few hours ahead. Comparing the results of different variables demonstrate that temperature can be forecasted with higher accuracy than the salinity and turbidity. The Hilo Bay water quality buoy is moored within

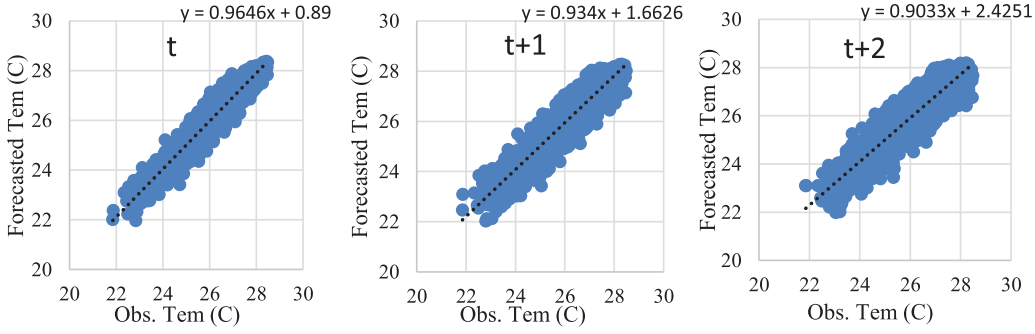


**Figure 9.** Scatterplots of surface temperature for the ANN models.

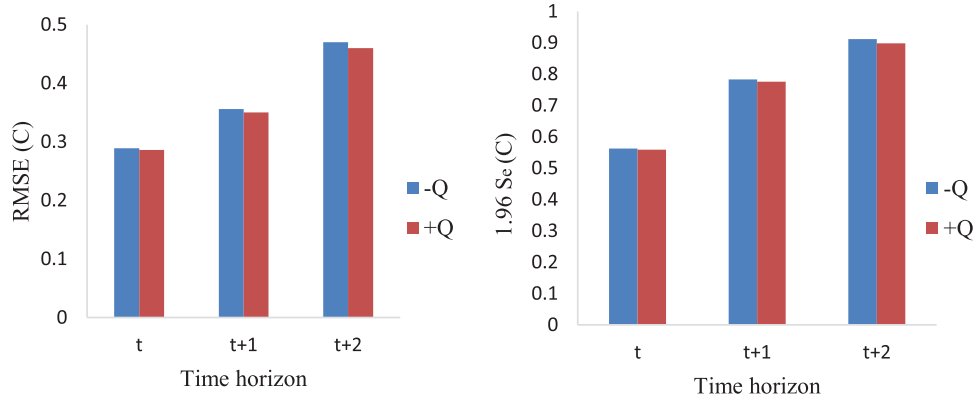


**Figure 10.** RMSE and uncertainty band of temperature forecast during test period.

1 m of the surface in one location within a small tropical estuary, and consequently the salinity measured by HBB is affected by freshwater inputs to Hilo Bay from the Wailuku River and submarine groundwater discharge (Mead & Wiegner, 2010; Paquay, Mackenzie, & Borges, 2007) as well as by tides (2-week Spring-Neap cycle). Storm events that result in elevated Wailuku River flow occur aperiodically and can depress salinity for several days (Mead & Wiegner, 2010); Paquay et al. (2007). However, the effect of Wailuku discharge and the Spring-Neap tide cycle on temperature is relatively smaller– while a strong storm can depress salinity 10–20 ppt, storm flows only affect temperature 2–4°C. On the other hand, turbidity is affected by the river flow with higher intensity, especially, in the case of storm flow, the sediments and other SSs increase the turbidity of the coastal waters. As the elevated temperature is an important factor in coral bleaching (Couch et al., 2017; Jokiel & Brown, 2004), accurate forecasts of local temperature in areas such as Hilo Bay where corals live is an important tool in understanding and managing coral bleaching. In general, the findings of this study are in a good accordance with the physics of the phenomena and characteristics of the water quality variables in the study area.

Wailuku River flow has a dominant influence on several biogeochemical processes and water quality parameters in Hilo Bay. Paquay et al. (2007) concluded that $pCO_2$ in Hilo Bay was largely driven by Wailuku R. flow conditions and its influence (along with the Wailuku R.) on the salinity gradient leading out of the bay. Mead and Wiegner (2010) showed that the surface water metabolic balance shifts from net autotrophy during low flow conditions to net heterotrophy during high flow conditions. Water quality parameters including nutrient concentrations and turbidity (Wiegner, Mead, & Molloy, 2013), C-source quality (Atwood, Wiegner, & MacKenzie, 2012), and microbial pollutants (Wiegner et al., 2017) have been shown to be related to storm vs. base flow river conditions in Hilo Bay. The models produced in the present study demonstrate the capability to model the influence of Wailuku river flow on these and other important parameters of Hilo Bay and perhaps other tropical estuarine systems.

Results of this study demonstrate that the river flow depleting in estuarine and coastal waters can affect water quality there. Moreover, it takes the coastal and estuaries some hours to recover themselves from the changes made by the flow. This study was carried out for depleting a freshwater and roughly clean water into the coastal waters. However, for rivers conveying industrial and domestic wastes, the problem is more serious, and the flow can degrade the water quality of that water body with frequently higher intensity. Especially, the problem would be worse when the river flow is released in smaller water bodies such as lakes and lagoons. Therefore, employing the proposed approach in this study can be helpful for the water quality monitoring and management. Moreover, a quantitative assessment of the river water quality and its impacts on the water quality it is depleted can be of great interest.

**Table 6.** Time and space complexity analysis of the procedure.

| | ANN | | ELM | | SVR | |
|---|---|---|---|---|---|---|
| | −Q | +Q | −Q | +Q | −Q | +Q |
| Time (s) | 14.01 | 14.20 | 6.83 | 7.00 | 21.89 | 22.39 |
| Space usage (%) | 39 | 40 | 38 | 38 | 55 | 56 |

### 3.4. Time and space complexity of the models

Dealing with big data and machine learning models, it is of great interest to know about time and space complexity of the models. In other words, runtime and required memory for implementation of the models can be considerable to select the suitable model. In this regard, this study provides a relative comparison for the three models. The results in terms of run time and memory space usage in percentile are given in Table 6. As the consuming time for models

As observed, the ELM models require less computational time and memory space. On the other hand, SVR models have more complexity in terms of time and space. The models including and excluding the river flow as input variables roughly show the same characteristics in terms of time-consuming and space usage. However, it should be noticed that these values are strongly dependent on the specifications of the processors, data size, a number of input elements, programming language or the software capability, transfer functions, data format, etc. Providing more sophisticated analysis for time and space complexity of the machine learning models can be the direction of future studies.

## 4. Conclusions

The current research was aimed to investigate the effect of the Wailuku River flow on the water quality of the Hilo Bay. Moreover, an attempt was made to achieve models with acceptable performance for water quality parameter forecasting in the upcoming hours. The correlation analysis between the river flow and the water quality indicators showed that turbidity in time t is mostly depended on the flow in time *t–1*. However, temperature and salinity showed a higher correlation with the flow with more lags

($t$–$2$). This issue reflects the fact that turbidity is immediately affected when the river flow enters the Bay. However, it takes some hours for salinity and temperature to be affected by the flow probably because of stratification and time-consuming heat transfer between layers. It is noteworthy that the highest correlation and lowest correlation between the river flow with the water quality parameters were related to turbidity ($R^2 = 0.67$) and temperature ($R^2 = -0.29$), respectively. The inverse correlation (negative values) of flow with salinity and temperature denotes release of fresh and colder water of the river flow into saline and warmer coastal waters.

Results of this study revealed that including the river flow data in the input structure of the models improve the accuracy and reliability of the forecasts. Its influence on the water quality variables varies from the remarkable effect on turbidity to negligible effects on the surface water temperature. For example, the ANN models for turbidity improves the model performance in average about 11%, 5.5%, and 10% in terms of *RMSE*, $R^2$, and width of the uncertainty band respectively. In a similar way, these results for temperature decreases to 1.6%, 0.3%, and 1.1%. The river flow and especially the storm flow can intensify turbidity by conveying sediments and making turbulence. Moreover, they can depress salinity depending on the flow discharge.

Comparing the performance of the ANN, ELM and SVR models demonstrated that all of them can provide reliable estimates of the water quality parameters. Moreover, their forecasts for different variables and times do not differ significantly, even though the SVR models require much longer execution time. The error measures (coefficient of determination, RMSE and width of the uncertainty band) for these models are roughly in the same range. The accuracy of the models deteriorate with an increase in forecasting time horizon. It happens with a higher rate of turbidity rather than salinity and temperature that implies turbidity may be prone to immediate change from different oceanic phenomena and storm flow compared with the other water quality variables. Evaluating performance of the models for different variables indicates that water temperature can be forecasted with higher accuracy than salinity and turbidity. The models provide forecasts of temperature in times $t$, $t+1$, and $t+2$ with $R^2 > 0.85$ and $0.286 < RMSE < 0.92$ in which they confirm high accuracy for the forecast. Interestingly, the width of uncertainty band with 95% confidence level does not exceed the unity (1°C) for 2 h ahead forecasting. The uncertainty band is narrower for shorter times ($t$, $t+1$). For salinity and turbidity, the uncertainty band is in a range of 2.5–4 (ppt or NTU).

Findings of this study are promising to develop models for water quality monitoring and beforehand forecasting. Conducting similar studies for rivers depleting industrial and domestic wastes in inland waters or coastal seas is essential. According to the results of this study, acceptable forecasts of the water quality indicators in a few hours ahead can be achieved in which it can be used as an early-stage alarms to prevent more damages to aquatic hydro-environment. In this study, only the river flow data were included in model development while knowing more details about the river flow such as its physical and chemical components can be helpful to improve the models' efficiency. For regions which are prone to continuous inputs of sewage and effluent, the inclusion of details of the river flow (e.g. contaminant measurements) in the forecasting models is essential to achieve models with higher accuracy and reliable estimates. Moreover, applying sophisticated models to analyze the computational time and space complexity of the machine learning models can be the direction of future studies.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

*Shahabbodin Shamshirband* http://orcid.org/0000-0002-6605-498X

## References

Aghbashlo, M., Shamshirband, S., Tabatabaei, M., Yee, L., & Larimi, Y. N. (2016). The use of ELM-WT (extreme learning machine with wavelet transform algorithm) to predict energetic performance of a DI diesel engine running on diesel/biodiesel blends containing polymer waste. *Energy*, *94*, 443–456.

Alizadeh, M. J. (2017). Discussion on "Gene expression models for prediction of longitudinal dispersion coefficient in streams" by Sattar, AMA, Gharabaghi, *B*., 2015. *Journal of Hydrology*, 524, 587–596.

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.

Atwood, T. B., Wiegner, T. N., & MacKenzie, R. A. (2012). Effects of hydrological forcing on the structure of a tropical estuarine food web. *Oikos*, *121*(2), 277-289.

Barzegar, R., Moghaddam, A. A., Adamowski, J., & Ozga-Zielinski, B. (2018). Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. *Stochastic Environmental Research and Risk Assessment*, *32*(3), 799–813.

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

Chau, K., & Jiang, Y. (2002). Three-dimensional pollutant transport model for the Pearl River Estuary. *Water Research*, *36*(8), 2029–2039.

Chenar, S. S., & Deng, Z. (2018). Development of genetic programming-based model for predicting oyster norovirus outbreak risks. *Water Research*, *128*, 20–37.

Clark, J. R. (1995). *Coastal zone management handbook*. Boca Raton, FL: CRC Press.

Corbari, C., Lassini, F., & Mancini, M. (2016). Effect of intense short rainfall events on coastal water quality parameters from remote sensing data. *Continental Shelf Research*, *123*, 18–28.

Couch, C. S., Burns, J. H., Liu, G., Steward, K., Gutlay, T. N., Kenyon, J., ... Kosaki, R. K. (2017). Mass coral bleaching due to unprecedented marine heatwave in Papahānaumokuākea Marine National Monument (Northwestern Hawaiian Islands). *PloS One*, *12*(9), e0185121.

Dogan, E., Sengorur, B., & Koklu, R. (2009). Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *Journal of Environmental Management*, *90*(2), 1229–1235.

Espey, W., & Ward, G. (1972). Estuarine water quality models. *Water Research*, *6*(10), 1117–1131.

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.

Franco-Lopez, H., Ek, A. R., & Bauer, M. E. (2001). Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sensing of Environment*, *77*(3), 251–274.

Garel, E., & D'Alimonte, D. (2017). Continuous river discharge monitoring with bottom-mounted current profilers at narrow tidal estuaries. *Continental Shelf Research*, *133*, 1–12.

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.

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.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2004). *Extreme learning machine: A new learning scheme of feedforward neural networks*. Neural networks, 2004. Proceedings. 2004 IEEE International Joint Conference on.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1), 489–501.

Jokiel, P. L., & Brown, E. K. (2004). Global warming, regional trends and inshore environmental conditions influence coral bleaching in Hawaii. *Global Change Biology*, *10*(10), 1627–1641.

Jung, N.-C., Popescu, I., Kelderman, P., Solomatine, D. P., & Price, R. K. (2010). Application of model trees and other machine learning techniques for algal growth prediction in Yongdam reservoir, Republic of Korea. *Journal of Hydroinformatics*, *12*(3), 262–274.

Liu, Z., Zhou, P., Chen, G., & Guo, L. (2014). Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting. *Journal of Hydrology*, *519*, 2822–2831.

Mead, L. H., & Wiegner, T. N. (2010). Surface water metabolism potential in a tropical estuary, Hilo Bay, Hawai'i, USA, during storm and non-storm conditions. *Estuaries and Coasts*, *33*(5), 1099–1112.

Mohammadi, K., Shamshirband, S., Kamsin, A., Lai, P., & Mansor, Z. (2016). Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure. *Renewable and Sustainable Energy Reviews*, *63*, 423–434.

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.

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.

Paquay, F. S., Mackenzie, F. T., & Borges, A. V. (2007). Carbon dioxide dynamics in rivers and coastal waters of the "big island" of Hawaii, USA, during baseline and heavy rain conditions. *Aquatic Geochemistry*, *13*(1), 1–18.

Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: Some past experiences and new approaches. *Journal of Hydroinformatics*, *10*(1), 3–22.

Sotomayor, G., Hampel, H., & Vázquez, R. F. (2017). Water quality assessment with emphasis in parameter optimisation using pattern recognition methods and genetic algorithm. *Water Research*, *130*, 353–362.

Taormina, R., & Chau, K.-W. (2015). Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and extreme learning machines. *Journal of Hydrology*, *529*, 1617–1632.

Tian, W., Liao, Z., & Zhang, J. (2017). An optimization of artificial neural network model for predicting chlorophyll dynamics. *Ecological Modelling*, *364*, 42–52.

Tomić, A. Š., Antanasijević, D., Ristić, M., Perić-Grujić, A., & Pocajt, V. (2018). A linear and non-linear polynomial neural network modeling of dissolved oxygen content in surface water: Inter-and extrapolation performance with inputs' significance analysis. *Science of The Total Environment*, *610*, 1038–1046.

Vapnik, V. (2013). *The nature of statistical learning theory*. Berlin: Springer Science & Business Media.

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.

Weiner, R. F., & Matthews, R. A. (2003). *Environmental engineering*. Oxford: Butterworth-Heinemann.

Wiegner, T., Edens, C., Abaya, L., Carlson, K., Lyon-Colbert, A., & Molloy, S. (2017). Spatial and temporal microbial pollution patterns in a tropical estuary during high and

low river flow conditions. *Marine Pollution Bulletin*, *114*(2), 952–961.

Wiegner, T. N., Mead, L. H., & Molloy, S. L. (2013). A comparison of water quality between low-and high-flow river conditions in a tropical estuary, Hilo Bay, Hawaii. *Estuaries and Coasts*, *36*(2), 319–333.

Wu, Z., Wang, X., Chen, Y., Cai, Y., & Deng, J. (2018). *Assessing river water quality using water quality index in lake taihu basin, China*. *Science of The Total Environment*, *612*, 914–922.

Yang, K., Yu, Z., Luo, Y., Yang, Y., Zhao, L., & Zhou, X. (2018). Spatial and temporal variations in the relationship between lake water surface temperatures and water quality-A case study of dianchi lake. *Science of The Total Environment*, *624*, 859–871.