Applied Ocean Research 53 (2015) 208–217

Contents lists available at ScienceDirect

Applied Ocean Research

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

# Forecasting of chlorophyll-a concentrations in South San Francisco Bay using five different models

Taher Rajaee, Amir Boroumand*

*Department of Civil Engineering, University of Qom, Qom, Iran*

A R T I C L E I N F O

*Article history:*
Received 8 January 2015
Received in revised form 31 August 2015
Accepted 1 September 2015
Available online 3 October 2015

*Keywords:*
ANN
Wavelet transforms
GA
SVR
Multiple linear regressions
Eutrophication
Chlorophyll-a
San Francisco Bay

A B S T R A C T

Accurate and reliable eutrophication level forecasting models are necessary for characterizing complicated water quality processes in bays. In this study, the ability of coupled discrete wavelet transform (DWT) with artificial neural network (ANN) and multi linear regression (MLR) (WANN and WMLR), ANN, MLR and genetic algorithm-support vector regression (GA-SVR) models for chlorophyll-a level forecasting applications were considered. The data used to develop and validate the models were monthly chlorophyll-a (Chl-a) data recorded from January 1994 to December 2013 were obtained from the NO.36 station located in the South San Francisco bay, USA. In the proposed WANN and WMLR models, the observed time series of Chl-a were decomposed to sub time series at different scales by DWT. Afterwards, the sub time series were used as input data to the ANN and MLR systems to predict the 1 month ahead Chl-a. Also the genetic algorithm was linked to SVR models to search for the optimal SVR parameters. The relative performance of the proposed models was compared together and the results showed that the WANN models were found to provide more accurate monthly Chl-a forecasts compared to the other models. The determination coefficient was 0.87, −0.04, 0.31, −2.36 and 0.24 for the WANN, WMLR, ANN, MLR and GA-SVR models, respectively. In addition, the WANN model predicted extreme Chl-a values precisely. The results indicate that the WANN models are a promising new method for eutrophication level forecasting in bays such as those found in South San Francisco Bay.



## 1. Introduction

Eutrophication has been one of the major water quality problems in estuaries and coastal waters in many countries. In recent decades, human activities have considerably increased the of nutrients delivery to many estuarine and coastal areas. Eutrophic conditions, which include low dissolved oxygen concentrations, declining sea grasses and harmful algal blooms, may impact the uses of estuarine and coastal resources by reducing the success of commercial and sport fisheries, fouling swimming beaches, and causing other problems due to the decay of excess amounts of algae [1,2]. Chlorophyll-a concentrations may be used to determine a bay's trophic status. Chlorophyll is the green pigment in plants' leaves that allows them to create energy light through photosynthesis. By measuring chlorophyll, the amount of photosynthesizing plants is indirectly measured. In a bay water sample, these plants would be algae or phytoplankton. Chlorophyll is a measure of all green pigments whether they are alive or dead. Chlorophyll-a is a measure of the portion of the pigment that is still alive. Sunlight, temperature, nutrients, and wind all affect both algae numbers and Chlorophyll-a concentration. During the spring when water begins to warm, the days are sunnier, and nutrients are still plentiful, the first outbreak or "bloom" of algae may occur. As the days become increasingly warmer and sunnier, algae will continue to grow more; however, they may soon outgrow the available supply of nutrients. Consequently, the total amount of algae growth may be limited. As summer turns to fall and temperature and sunlight decrease, algae concentrations will decrease as well. Same pattern of variation can be seen at San-Francisco Bay.

In recent years, eutrophic condition has been monitored at the bays, in terms of both temporal and spatial variation. At San Francisco Bay, they run a long-term program to monitor chlorophyll-a. The measurements are done monthly at fixed stations in the bay.

Understanding and modeling the level of eutrophication (Chlorophyll-a) can be helpful to estuaries ecosystem management. In this regard, water quality and environmental models have been used widely to assist water resources' managers developing control strategies for estuarine water quality management.

Nutrients' load of phosphate and nitrate plays a key role in outbreak and growth of algal blooms. In other words, phosphate and nitrate are the indices that control this process. Management of bays upstream catchment land use, implementation of total maximum daily load plan and or nutrients' load allocation plans (increasing or decreasing the consumption of the fertilizers and

* Corresponding author. Tel.: +98 9125853026.
*E-mail address:* a.broumand@stu.qom.ac.ir (A. Boroumand).

0141-1187/$ – see front matter 
http://dx.doi.org/10.1016/j.apor.2015.09.001

chemical pesticides in upstream farms) may control eutrophication at the bays [3].

So catchments nutrient management is an important factor to decrease eutrophication process. The management incompetency occurs when there is no link between practical catchment models such as SWAT and water body simulation models including intelligent models. As a result linking between these models can be assisted for bay water quality managers.

Traditional models used for water bodies' qualitative modeling are complex models. Putting together the assumption and limitations of the models with data uncertainty leads to a very complex systems that make the modeling outcomes uncertain. On the other hand application of these models require sufficient expertise and experience and users' skills to calibrate, validate and verify the models affects the results directly. Hence, the limitation of these models on accurate forecasting the level of the algal blooms, and also the nonlinear relationship between the water quality and environmental indices and that of the level of Chlorophyll-a necessitate a new method using intelligent models. Artificial neural networks (ANNs) are the most used intelligent ones. Neural networks offer a number of advantages, including requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms. Disadvantages include its "black box" nature, greater computational burden, proneness to over fitting, and the empirical nature of model development [4].

The last decade has seen a tremendous growth in interest in the application of ANNs to water resources and environmental problems. In recent years ANNs have found a number of applications in prediction of eutrophic conditions. Lee et al. [5] used ANN to modeling of coastal algal blooms of Hong Kong coastal waters. They showed that the algal concentration in the eutrophic sub-tropical coastal water is mainly dependent on the antecedent algal concentrations in the previous 1–2 weeks. Their study also showed that an ANN model with a small number of input variables is able to capture trends of algal dynamics, but data with a minimum sampling interval of 1 week is necessary. Kuo et al. [6] applied a back-propagation (BP) ANN to reservoir eutrophication prediction in central Taiwan. They showed that the ANN is able to predict water quality indicators with reasonable accuracy. Melesse et al. [7] used a multilayer perceptron-back propagation (MLP-BP) algorithm of ANN to predict the level of eutrophication (chlorophyll-a) from water quality parameters monitored at two Florida Bay water quality monitoring stations (FLAB03 and FLAB14). They studied seven input data scenarios, and compared the models performance. They showed that the prediction with antecedent chlorophyll-a alone gave a stable result with smaller error and higher performance attributed to easier and more efficient training. They also showed that the MLP-BP technique is applicable to the monitoring and prediction of algal blooms and will be crucial to coastal watershed management. Huo et al. [8] used ANN model to relate the key factors that influence a number of water quality indicators such as dissolved oxygen, total phosphorus, chlorophyll-a and secchi disk depth in Lake Fuxian, China. Their results indicated that the ANN model performs well in ten months prediction and the ANN is able to predict eutrophication indicators with reasonable accuracy. Saghiet al. [9] developed a feed forward ANN model to analysis trophic state index (TSI) in the Dez Dam reservoir; Iran. They showed that ANN is a suitable tool for quality modeling of dam reservoir and increment and decrement of nutrients in trend of eutrophication.

In recent years, wavelet analysis and ANNs was used for water resources and environmental engineering problems. Kisi [10] considered the accuracy of WANN and single ANN models in monthly stream flow prediction and resulted that the WANN performs much better than single ANN. Partal and Cigizoglu [11] applied WANN for prediction of suspended sediment load in rivers. Rajaee et al. [12,13] proposed a model by combining wavelet analysis and the neuro-fuzzy (NF) approach to predict daily suspended sediment load (S). In the developed WNF models, the daily observed time series of river discharge and S were decomposed into several sub time series at different scales using discrete wavelet transform. The results indicated that the WNF model more efficient than the NF and sediment rating curve models in the prediction of S. In another study, Rajaee et al. [14] applied a WANN model to Iowa River suspended sediment load prediction, USA. They showed that wavelet-transformed data improve the ability of predicting model by capturing useful information on various resolution levels. SatyajiRao et al. [15] proposed a WANN model for the prediction of daily runoff in the West flowing Rivers of India. The results of daily runoff time series modeling illustrated that the performances of WANN models are more effective than the ANN models.

A number of studies have applied wavelet analysis and ANNs for eutrophication modeling in lakes. Kim et al. [16] proposed a WANN model to chlorophyll-a concentration forecasting 1, 3, and 7 days ahead, in Korea. They showed that WANN models constitute a promising new method for short-term chlorophyll-a concentration forecasting in large lakes. Wang et al. [17] proposed a hybrid WANN method for chlorophyll-a simulation in the lake Baiyangdian, North China. They compared the performance of the proposed WANN model for monthly chlorophyll-a simulation in the lake ecosystem with a multiple stepwise linear regression (MSLR) model, an autoregressive integrated moving average (ARIMA) model and a regular ANN model. Their results showed that the WANN model was suitable for chlorophyll-a simulation providing a more accurate performance than the MSLR, ARIMA, and ANN models. Nourani et al. [18] reviewed applications of hybrid wavelet–artificial intelligence models in hydrology.

In recent years, support vector regression (SVR) was used for water resources and environmental engineering problems. Many applications SVR can be found in water resource and environmental engineering literature. Some of these studies in the field of water quality modeling are presented in Table 1 [19–23].

In this research, WANN, ANN, WMLR, MLR, and GA-SVR models were developed for one month ahead forecasting of eutrophication in San Francisco Bay gauging station in the USA, and were compared together. Based on the author's findings, this paper is the first application of the WANN, WMLR, and GA-SVR models for bay eutrophication prediction.

This paper is organized as follows: in Section 2, a brief description about wavelet, ANN, GA, SVR, MLR models and models performance criteria's are presented, respectively. Gauging station and data analysis are presented in Section 3. Section 4 represents the proposed models. The models application for chlorophyll-a prediction and results are summarized in Section 5. The concluding remarks will be the last section.

## 2. Methods

### 2.1. Wavelet analysis

A wavelet system is a set of building blocks to construct or represent a signal or function. It has become a popular analytical tool due to its ability to simultaneously elucidate both spectral and temporal information within the signal. This property overcomes the basic shortcoming of Fourier analysis, which is that the Fourier spectrum contains only globally averaged data. Wavelet transforms, which provide information in both the time and frequency domains of a signal, give considerable information about the physical structure of data. It provides a time-frequency representation of a signal in many different periods of the time domain [24]. The time-scale