Pol. J. Environ. Stud. Vol. 23, No. 4 (2014), 1281-1287

Original Research

Applying Artificial Neural Networks for the Estimation of Chlorophyll-*a* Concentrations along the Istanbul Coast

Ruya Samli[1], Nuket Sivri[2], Selcuk Sevgen[1]*, Vildan Zulal Kiremitci[2]

[1]Computer Engineering Department,
[2]Environmental Engineering Department,
Engineering Faculty, Istanbul University, 34320, Avcilar, Istanbul, Turkey

Received: 12 November 2013
Accepted: 28 February 2014

Abstract

Chlorophyll-*a* (chl-*a*) concentration is considered to be the main measure of phytoplankton biomass. The location and intensity of the surface chl-*a* maximum in a coastal area are governed by daylight hours, air and seawater temperatures, and nutrient availability in the euphotic zone. The aim of this study is to model a back-propagation neural network (BP-ANN) for estimating chlorophyll-*a* concentrations from obtained input values. In this study an ANN structure of 3 input neurons and 1 output neuron is used. The 3 inputs represent sea surface temperature (SST), air temperature, and daylight hours, while the output represents chl-*a* concentration respectively and hidden layers number which is dependent to the application is determined as 20. The ANN structure, which is simulated in MATLAB, estimated the data of the experiments. When compared to current data, it can be said that these are successful results and they provide ANN for estimating chl-*a*. In our ANN approach, the effects of all input/output parameters can be evaluated and various outputs can be obtained for different environments and predicted maximum chl-*a* data.

Keywords: back-propagation, artificial neural networks, chlorophyll-*a*, estimation

Introduction

Among the studies carried out in sea-based areas pertaining to the measurement, prevention, and control of increasing pollution within recent years, it has become extremely important to determine both the locations of increased primary production and the durations of this increases. High levels of phytoplankton biomass in areas where primary production rates are high tend to confirm the existence of a linear relationship. With the increase in urbanization, the weight imposed on the shoreline ecosystems, nutrient entry, and the establishments of optimal physical conditions, observable increases in primary productions during late spring and early fall have become inevitable. Not only during these periods but also during intermittent periods, different methods other than on-site sampling have been used to determine the increasing algae biomass during mid-periods other than the previously mentioned periods.

Chlorophyll-*a* (chl-*a*) concentration is considered to be the main measurement of phytoplankton biomass. There are a couple of important reasons why chl-*a* analyses are widely used:

(1) it is a major foundational pigment for all phytoplankton groups that plays an important role in photosynthesis
(2) its analysis is easy [1]

Because chl-*a* concentrations are an indicator of phytoplankton, environmental factors affecting phytoplankton also effect the chl-*a* concentrations in these organisms [2].

*e-mail: sevgens@istanbul.edu.tr

Determining the photosynthetic pigment concentrations in sea water will provide information on the phytoplankton population within the algae groups and in general, with information on the chl-*a* concentration, simple inferences can be made.

To accurately interpret the increase in the pytoplankton biomass, not only should the cellular increase be taken into account but also the factors that have an impact on this increase should be taken into consideration. Especially in areas with high nutrient concentrations and no light limit, studies have documented increases in phytoplankton forms mediated by their light adaptation mechanisms. Sea surface temperature and ambient temperature, daylight hours, and nutrients are among the most important abiotic factors that impact phytoplankton mass growth [3]. Daylight hours are especially effective in the development of vegetative cell growth. Changes in temperature and other factors impact photosynthetic production. In general, with the increase in temperature, so does the primary production; however, after it reaches a point, this production rapidly decreases. With heat, light also impacts the seasonal changes in phytoplankton production. Excessive changes in sea movement, changes in salinity and dissolved oxygen concentrations, or stability deprivation are among other important parameters that impact the increased growth.

Correspondingly to today's technological developments, experimental research plays a crucial role in research studies. Although experimental studies can give realistic results, they are limited by cost, time, the environment, and parameter properties. To avoid these limitations, experimental studies are modeled with the help of computer modeling or simulation techniques. Within the scope, in recent years these type of studies have been simulated by different methods like optimization methods, regression analysis, and artificial neural networks. But not every modeling technique gives accurate results in every experimental study and the properties must correspond to each other. In the estimation problems, a system that has the ability to be trained like artificial neural networks (ANN) would be successful. ANN aims to model the human nervous system with the properties of learning, estimating and making decisions. They have some advantages apart from conventional problem-solving techniques like nonlinearity and the ability to complete missing data. So they are used in different types of problems and because they got successful results, their usage areas increased daily. Especially in recent years ANN has become a popular solving technique and is used in many fields like accounting, civil engineering, mine engineering, environmental engineering, medicine, etc.

In the ecological studies, evaluation of bio-factors interaction with each other, and hence to estimate the missing data to obtain the result, is an important field of study. For this purpose, the aim of this study, which is primarily used in the literature of artificial intelligence methods are often utilized for artificial neural networks. In this literature, using artificial neural networks such topics as for weather forecasting [4-10], prediction of material in an area [11-13], and determining quality [14-16]. In this study, the chl-*a* concentration is calculated with a BP-ANN structure that has 3 inputs as sea surface temperature, (SST), air temperature and daylight hours. According to obtained values, BP-ANN is seen as an appropriate method for estimating chl-*a* concentrations.

Material and Method

Study Area

In this study, ten sampling stations were selected among coastal stations on the southwestern coast of İstanbul (Fig. 1). The sampling was carried out monthly between January 2008 and December 2012. Air temperature and sea surface temperature were measured in situ during the sampling.

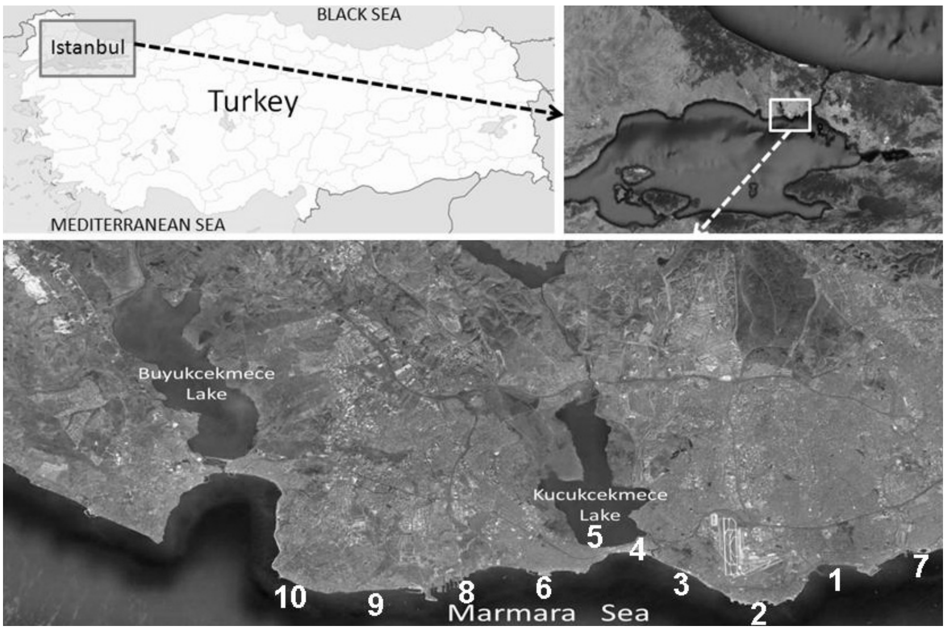

Fig. 1. Study area and the locations of the sampling stations on the southwestern coast of Istanbul.

Also, daylight hours data were taken from the Turkish State Meteorological Service (TSMS) (*http://www.mgm.gov.tr*).

Chl-*a* Measurements

The chl-*a* samples were collected from the sea surface. Seawater (1 L) was filtered through filter paper with pore size of 0.45-µm. Following filtration, the filter paper was folded and placed into a 20-ml centrifuge tube. Then, 10 ml of 90% acetone solution was added and the tube was left in the refrigerator overnight to ensure the extraction of chlorophyll by acetone. Following this procedure, the sample was centrifuged for 10 min at room temperature. The absorbance values at 750, 664, 647, and 630 nm were recorded using a Shimadzu T60U spectrophotometer. Then the absorbance values measured at 664, 647, and 630 nm were subtracted from the absorbance value at 750 nm in order to take the errors introduced by turbidity into consideration. Following these measurements, the amount of chl-*a* per liter (µg/L) of seawater was calculated using a previously reported calculation procedure [17].

Artificial Neural Networks

ANN is a heuristic method that has a purpose of modelling the biologic neural network with the properties of learning, completing missing data, predicting, and making decisions. ANN is used as a powerful tool to analyze the behaviour of an existing system in order to design new processes and to optimize the conditions, or to predict the behaviour of a given system with given parameters [18].

In order to predict system behaviour, most popular structures of ANN architectures are feed-forward networks that are mostly trained from the input data using error back-propagation algorithm [19]. This ANN architecture has a three-layered structure. The input and output layers represent independent variables and dependent variables of the system, respectively, while the hidden layer is used to perform the transformations. The hidden layer is the most important part of ANN because it has the adjusting neuron number ability. The number of neurons is not definite and can be organized according to the problem structure and functions used. Previous studies could not find any certain rules or facts for determining the number of neurons in the hidden layer, but it is known that there is a relationship between this number and complexity of system. Too many hidden layer neurons provide successful training and memorizing, but unsuccessful testing and generalizing [20].

In this study, ANN is used for estimating outputs about chlorophyll-*a* by using the inputs of water temperature, air temperature and daylight hours. When ANN is used for an estimating problem, generally back-propagation typed neural networks (BP-ANN), which has a supervised learning method, are used. In these methods, input values are separated into groups as training, validation, and test [21]. A BP-ANN structure is shown in Fig. 2.

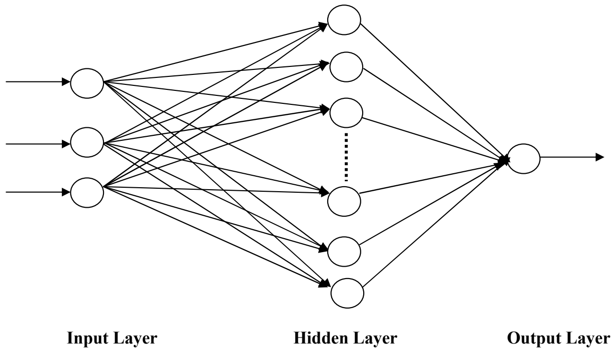

Fig. 2. Back-propagation ANN structure.

Results and Discussion

The SST and air temperature were measured in situation on a monthly basis from 2008 to 2012. The highest SST was measured as 26.7ºC in August 2008 and the lowest SST was measured as 5ºC in February 2012 during the sampling period (Fig. 3). A similar data of the ambient temperature yielded that the highest temperature was measured in June 2012 as 34.1ºC and the lowest temperature was measured in January 2012 as 3.8ºC (Fig. 4).

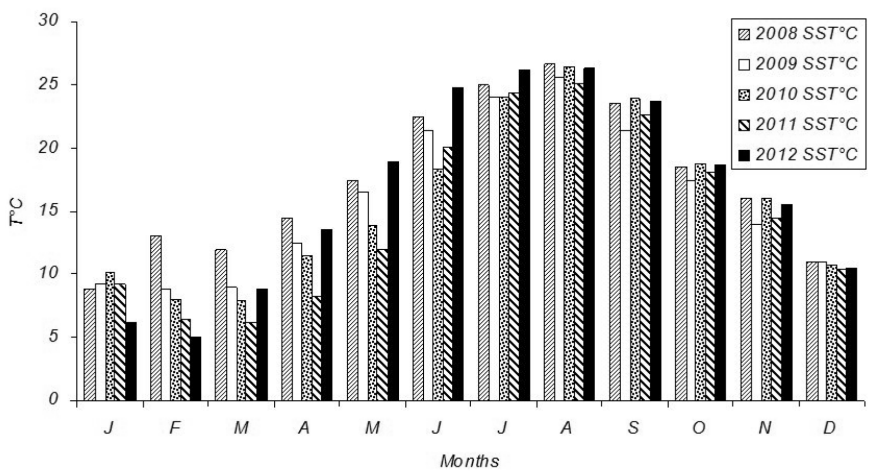

Fig. 3. Average SST for years 2008-12.

During daylight hours, sun angle plays a significant role. However, phytoplankton concentration is the critical factor controlling light attenuation and, therefore, light levels below the euphotic layer. In this study, the maximum value was found for the period of summer and the minimum values were found in winter months for each year. These values show that the surface water temperature data is significantly correlated with daylight hours (Fig. 5).

In this study, chl-*a* concentration data are estimated with BP-ANN. For this reason, a BP-ANN structure with 3 input neurons and 1 output neuron is used. The 3 input neurons represent sea surface temperature (SST), air temperature, and daylight hours, respectively. This BP-ANN system, which has 10 neurons in the hidden layer, is simulated by using a MATLAB program. To prevent MATLAB taking the input values randomly, we used divideblock parameter and arranged them according to the monthly values (from 2008 to 2012). After this process, the ordered input values are separated into three groups as usual: training 70%, validation 15%, and testing 15%. While updating the weight coefficients, the mean squared error (MSE) function whose formula is given below is used.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \qquad (1)$$

...where $\hat{Y}$ is a vector of n predictions, and Y is the vector of the true values [22].

Fig. 5 shows the MATLAB simulation of the BP-ANN structure. Respectively a, b, and c parts show training, validation, and all values.

Fig. 6 gives the comparison of experimental chl-*a* concentration results and BP-ANN estimated chl-*a* concentration results. It can be seen that most of the values, especially peak points at 60 months, are consistent for each other.

Because there chl-*a* concentration data in the last three months of the dataset were not calculated from experimental studies, from Figs. 6 and 7, it can be said that the training of BP-ANN is not 100% successful. But when the peak

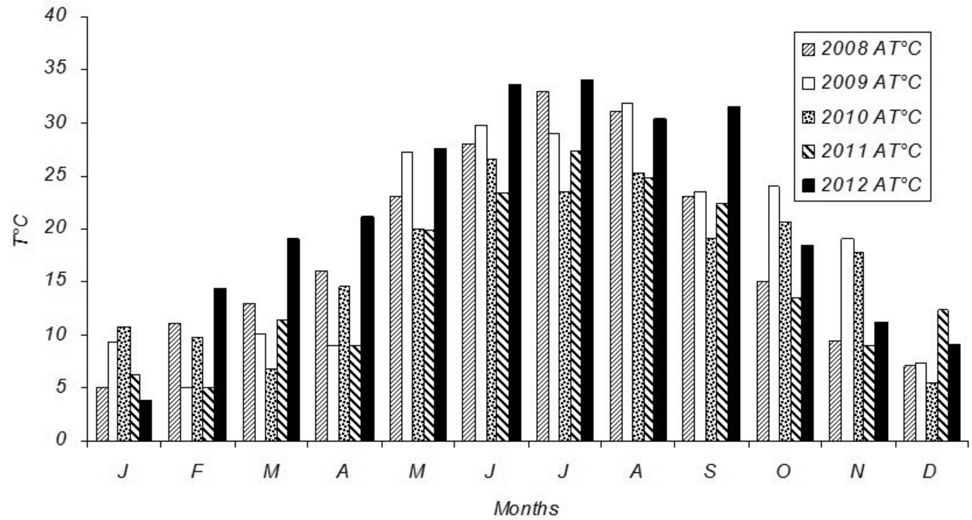

Fig. 4. Average ambient air temperature for years 2008-12.

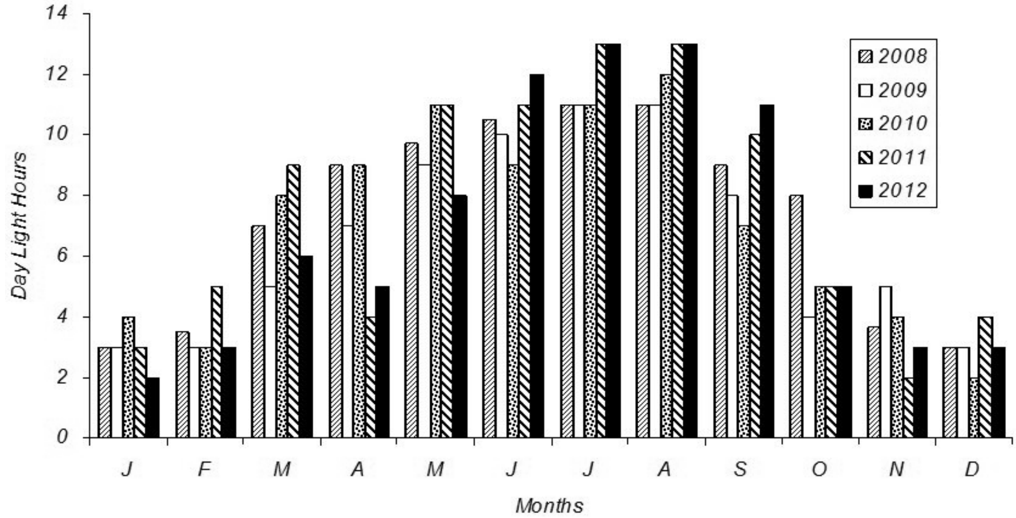

Fig. 5. Average daylight hours for years 2008-12.

values that take an important place of annual experiments are examined, it can be seen that they are determined successfully.

Conclusions

The measurement and distribution of microscopic living plant matter commonly referred to as phytoplankton or algae have been of interest to scientists, researchers, and aquatic resource managers for decades. An understanding of the phytoplankton population and its distribution enables researchers to draw conclusions about a water body's health, composition, and ecological status. For instance, nutrient pollution is a process whereby water bodies such as lakes, estuaries, or slow-moving streams receive excess nutrients that stimulate excessive growth of microorganisms, including algae and cyanobacteria. This enhanced growth often results in a bloom [23]. The major problems of eutrophication come with anthropogenic enrichment of the environment and the formation of dense blooms of toxic Dinoflagellates (principally marine) and colonial blue green algae (fresh water). The ability of blue-green algae to out-compete other members of phytoplankton at a time of year when certain environmental aspects (light, temperature) are at an optimum, is a key feature of the success of these organisms in bloom formation [24]. An overall summary regarding the Chlorophyll-*a* studies conducted in the Marmara Sea is provided in Table 1. The highest chl-*a* concentration was measured in the upper estuary of the Golden

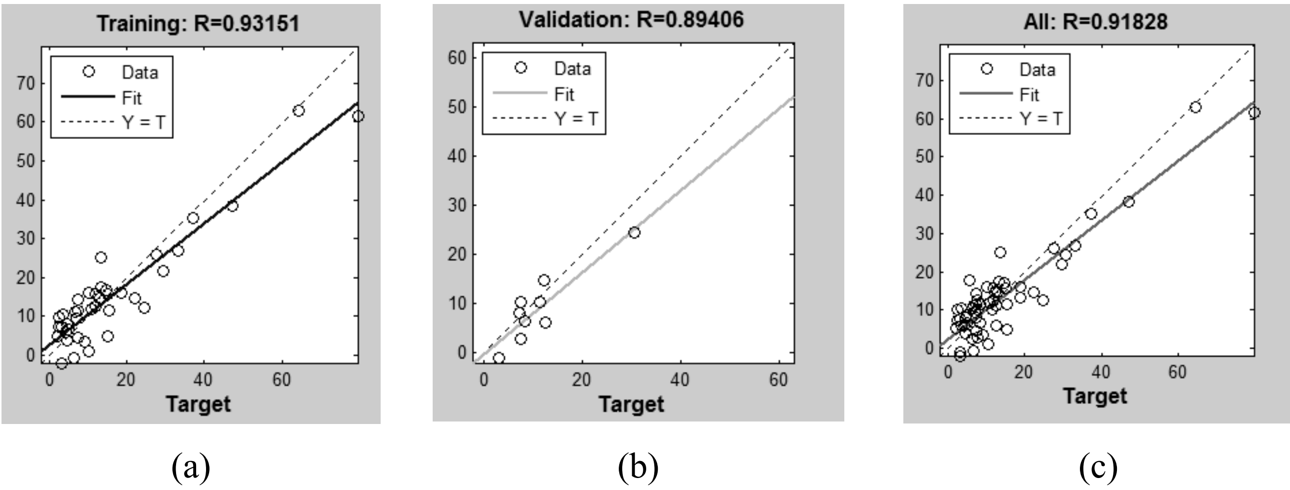

Fig. 6. The results of ANN simulation in MATLAB: a) training results of ANN simulation, b) validation results of ANN simulation, c) all results of ANN simulation.

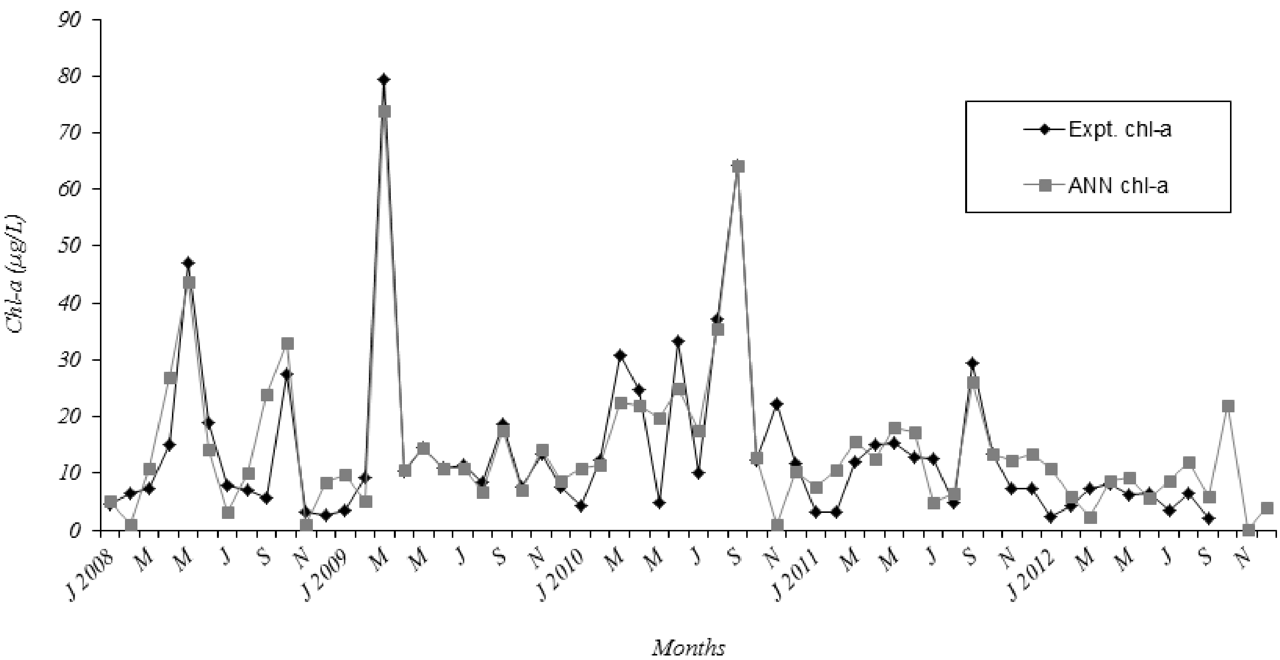

Fig. 7. The comparison of calculated and estimated chl-a values.

Table 1. Chl-*a* values (µg/L) measured at various stations in the Marmara Sea [3].

Stations	Min	Max	Citation
Bosphorus	0.54	6.13	Aktan et al., 2007
North-eastern of the Marmara Sea	0.26	11.72	Deniz & Taş, 2009
Golden Horn-in the lower estuary	1.10	27.20	Taş et al., 2009
Golden Horn-in the upper estuary	1.50	286.00	Taş et al., 2009
Büyükada	0.10	6.35	Balkıs et al., 2011

Horn as 286 µg/L, and this area deservers attention as a special ecosystem. The Golden Horn, which is an estuary under intense pollution stress, has become a region of dense population and high urbanization in recent years. Chl-*a* values were determined to be very high during warm periods in this area where currents and fresh water flow are very low.

In general, the amount of chlorophyll in a collected water sample is used as a measure of the concentration of suspended phytoplankton. The use of the measurement of phytoplankton is an indicator of water quality. Currently, chlorophyll determinations are made on lakes, rivers, reservoirs, and coastal and ocean waters across the globe. For this purpose, in this study we used a back-propagation neural network (BP-ANN) for estimating chl-*a* concentrations from obtained input values between January 2008 and December 2012.

These results are also estimated using BP-ANN structure and the results are compared. ANN is ideal for modeling temperature based on these relations. Artificial neural networks are well suited to modeling the nonlinear relationships between water quality (physical, chemical, etc.) and meteorological (air temperature, rainfall, etc) variables [10]. The results of the neural network training were considered to be very accurate and the validation test also indicated very satisfactory prediction accuracy [25]. In our ANN approach, the effects of all input/output parameters can be evaluated and various outputs can be obtained for different environments and predicted maximum chl-*a* data.

References

1. AGIRBAS E. Interaction of pigment concentration and primary production with environmental condition in the Southeastern Black Sea, PhD Thesis. Karadeniz Technical University. Trabzon, **2010** [In Turkish]
2. KAYAALP T., POLAT S. Estimation of Chlorophyll-a for Full Completed and Incompleted Regression Model, E.U. Journal of Fisheries & Aquatic Sciences, **18**, (3-4), 529, **2001** [In Turkish].
3. SIVRI N., SEKER D.Z., BALKIS N., ZAN A. Analysis Of Chlorophyll-*a* Distribution On The South-Western Coast Of Istanbul During 2008-2010 Using GIS. FEB. **21**, (11), 3233, **2012**.
4. MAQSOOD I., KHAN M. R., ABRAHAM A. An ensemble of neural networks for weather forecasting. Neural Comput. Appl. **13**, (2), 112, **2004**.
5. HAYATI M., MOHEBI Z. Application of artificial neural networks for temperature forecasting. World Academy of Science, Engineering and Technology. **28**, (2), 275, **2007**.
6. HAYATI M., MOHEBI Z. Temperature Forecasting Based on Neural Network Approach.World Applied Sciences Journal. **2**, (6), 613, **2007**.
7. BABOO S. S., SHEREEF I. K. An efficient weather forecasting system using artificial neural network. International Journal of Environmental Science and Development. **1**, (4), 2010, **2010**.
8. MIYANO T., GIROSI F. Forecasting Global Temperature Variations by Neural Networks, (No. AIM-1447), Massachusetts Institute Of Technology Cambridge Artificial Intelligence Lab, **1994**.
9. SIVRI N., KILIC N., UCAN O. N. Estimation of stream temperature in Firtina Creek (Rize-Turkiye) using artificial neural network model. J. Environ. Biol. **28**, (1), 67, **2007**.
10. SIVRI N., OZCAN H., UCAN O., AKINCILAR O. Estimation of Stream Temperature in Degirmendere River (Trabzon-Turkey) Using Artificial Neural Network Model. Turkish Journal of Fisheries And Aquatic Sciences. **9**, (2), 145, **2009**.
11. MOZEJKO J., GNIOT R. Application of Neural Networks for the Prediction of Total Phosphorus Concentrations in Surface Waters. Pol. J. Environ. Stud. **17**, (3), 363, **2008**.
12. GENCOGLU M. T., CEBECI M. Investigation of pollution flashover on high voltage insulators using artificial neural network. Expert Syst. Appl. **36**, (4), 7338, **2009**.
13. WU G. D., LO S. L. Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system. Eng. Appl. Artıf. Intel. **21**, (8), 1189, **2008**.
14. SUDHEER K. P., CHAUBEY I., GARG V. Lake Water Quality Assessment From Landsat Thematic Mapper Data Using Neural Network: An Approach To Optimal Band Combination Selection. J. Am. Water Resour. As. **42**, (6), 1683, **2006**.
15. BARAI S. V., DIKSHIT A. K., SHARMA S. Neural network models for air quality prediction: a comparative study. In Soft Computing in Industrial Applications. Springer: Berlin, Heidelberg, pp. 290-305, **2007**.
16. NAGENDRA S. S., KHARE M. Modelling urban air quality using artificial neural network. Clean Technologies and Environmental Policy. **7**, (2), 116, **2005**.
17. PARSONS T. R., MAITA Y., LALLI C. Manual of Chemical and Biological Methods for Sea Water Analysis, Pergamon Press, Great Britain, **1984**.
18. SINHA K., CHOWDHURY S., DAS SAHA P., DATTA S. Modeling of microwave-assisted extraction of natural dye from seeds of Bixa orellana (Annatto) using response sur-

face methodology (RSM) and artificial neural network (ANN). Ind. Crop. Prod. **41**, 165, **2013**.
19. BHATTI M.S., KAPOOR D., KALIA R.K., REDDY A.S., THUKRAL A.K. RSM and ANN modeling for electrocoagulation of copper from simulated wastewater: Multi objective optimization using genetic algorithm approach. Desalination. **274**, 74, **2011**.
20. CALLAN R. The essence of neural networks. Southampton Institute, Prentice Hall Europe, **1999**.
21. YEGNANARAYANA B. Artificial Neural Networks. Prentice-Hall of India, **2006**.
22. WACKERLY D., SCHEAFFER W. Mathematical Statistics with Applications. Thomson Higher Education, **2008**.
23. OKAFOR N. Environmental Microbiology of Aquatic and Waste Systems, Springer, New York, 307, **2011**.
24. SIGEE D.C. Freshwater Microbiology, Biodiversity and Dynamic Interactions of Microorganisms in the Aquatic Environ-ment, John Wiley & Sons Inc., USA, 541, **2005**.
25. SAILA S., CHEESEMAN M., POYER D. Maximum Stream temperature estimation from air temperature data and its relationship to Brook Trout (*Salvelinus fontinalis*) Habitat requirements in Rhode Island. Wood Pawcatuck watershed association (WPWA), **2004**.