Wetlands (2021) 41: 111
https://doi.org/10.1007/s13157-021-01479-6

WETLANDS IN THE DEVELOPING WORLD

Machine Learning Modeling Techniques for Forecasting the Trophic Level in a Restored South Mediterranean Lagoon Using Chlorophyll-*a*

Nadia Ben Hadid[1,2,3] · **Catherine Goyet**[2,3] · **Hatem Chaar**[4] · **Naceur Ben Maiz**[5] · **Veronique Guglielmi**[2,3] · **Abdessalem Shili**[1]

Received: 17 March 2021 / Accepted: 8 July 2021 / Published online: 16 November 2021

Abstract

An Artificial Neural Network (ANN), a Machine Learning (ML) modeling approach is proposed to predict the ecological state of the North Lagoon of Tunis, a shallow restored Mediterranean coastal ecosystem. A Nonlinear Auto Regressive with exogenous input (NARX) neural network model was fitted to predict Chlorophyll-*a* (Chl-*a*) concentrations in the North Lagoon of Tunis as an eutrophication indicator. The modeling is based on approximately three decades of monitoring water quality data (from January 1989 to April 2018) to train, validate and test the NARX model. The most relevant predictor variables used in this model were those having a high permutation importance ranking with Random Forest (RF) model, which simplified the structure of the network, resulting in a more accurate and efficient procedure. Those predictor variables are secchi depth, and dissolved oxygen. Various model scenarios with different NARX architectures were tested for Chl-*a* prediction. To verify the model performances, the trained models were applied to field monitoring data. Results indicated that the developed NARX model can predict Chl-*a* concentrations in the North Lagoon of Tunis with high accuracy ($R = 0.79$; $MSE = 0.31$). In addition, results showed that the NARX model generally performed better than multivariate linear regression (MVLR). This approach could provide a quick assessment of Chl-*a* variations for lagoon management and eco-restoration.

Keywords Machine Learning · NARX · Random Forest · Eutrophication · North Lagoon of Tunis · Forecasting

Introduction

Transitional water bodies, such as coastal lagoons, are situated at the interface between the continent and the sea. These are active areas delivering essential ecosystem services (Mooney et al. 2009; Newton et al. 2018), and they occupy approximately 13 % of the world's coastline (Barnes 1980). In these semi-closed water bodies, the gradient from fresh to saline water creates a rich biodiversity (Basset et al. 2013). In recent decades, Mediterranean coastal lagoon ecosystems have been especially exposed to anthropic eutrophication, primarily due to urbanization (Zaldívar et al. 2008; Souchu et al. 2010). Indeed, limited exports to the open sea and long periods of residence time in these water bodies have resulted in the accumulation of nutrients, mainly from anthropogenic activities (de Jonge et al. 2002; Newton et al. 2014). This excess of nutrient inputs contributed to the eutrophication of the water columns (Cloern 2001; Souchu et al. 2010).

Several studies have been conducted in coastal Mediterranean lagoons to assess the eutrophication level.

This article belongs to the Topical Collection: Wetlands in the Developing World.

✉ Nadia Ben Hadid
nadiabenhadid01@gmail.com

1 Département Génie Halieutique et Environnement, Unité de Recherche Ecosystèmes et Ressources Aquatiques, Université de Carthage, Institut National Agronomique de Tunisie, UR03AGRO1, 43 Avenue Charles Nicolle, 1082 Tunis, Tunisia

2 ESPACE-DEV, Univ Montpellier, IRD, Univ Antilles, Univ Guyane, Univ Réunion, Montpellier, France

3 Laboratoire IMAGES-ESPACE-DEV, Univ. Perpignan Via Domitia, Perpignan, France

4 Département Génie Rural, Eaux et Forêts (GREF), Université de Carthage, Institut National Agronomique de Tunisie, Laboratoire de Grande Culture (LGC-INRAT), 43 Avenue Charles Nicolle, 1082 Tunis, Tunisia

5 Société Al Buhaira de Développement et d'investissement, Les Berges du Lac, Rue de Rodrigo de Freitas, 1053 Tunis, Tunisia

García-Ayllón (2017) stated that the Mar Menor lagoon, located in the East of the Region of Murcia in Spain, has suffered from an important process of intense anthropization over the last five decades. One of the main indicators was the rapid population growth of new jellyfish species, reaching more than 100 million specimens each summer (Robledano et al. 2011). Thau Lagoon is another particularly interesting case of Mediterranean coastal lagoon eutrophication. Thau lagoon is an ecosystem located at the Mediterranean French coast, known by supporting traditional shellfish farming activities in France and which has been subject to eutrophication leading to large anoxic events associated with significant mortality of shellfish stocks (Derolez et al. 2020). In this context, the North Lagoon of Tunis, a South Mediterranean lagoon, located in the north of Tunisia, provides a good example for the diagnosis and study of eutrophication in Mediterranean coastal ecosystems. In fact, the North Lagoon of Tunis is one of the most important restored lagoons in Tunisia, which has known a critical ecological state substantially due to urban development (Harbridge et al. 1976).

Chlorophyll-*a* (Chl-*a*) is the most essential pigment in aerobic photosynthetic organisms and its measurement is used as an indicator of the phytoplankton biomass present in water and thus, the degree of eutrophication of the ecosystem (Tian et al. 2017). The probable presence of algae blooms that have a major effect on the physical, chemical and biological processes of the lagoon can be interpreted as elevated Chl-*a* levels (Tian et al. 2017).

Cyanotoxins generated by cyanobacteria in lagoon water could present a risk to human health (Watzin et al. 2006; Mc Quaid et al. 2011; Kalaji et al. 2016). In situations where the current cyanotoxin concentration is not available, Chl-*a* is also widely recognized as a surrogate indicator of cyanobacterial density (Wheeler et al. 2012). It is therefore important to monitor Chl-*a* concentration and, in turn, to provide information on the management of water quality.

The ability to automatically monitor water quality is very useful, especially in vulnerable areas where (1) there is a high threat of potential pollution events and (2) related socio-economic activities that involve preventive actions are carried out (Jimeno-Sáez et al. 2020). However, as far as we know, there is no automated system that reliably measures the concentration of Chl-*a* in real time (Jimeno-Sáez et al. 2020). Measurements of Chl-*a* must be performed in laboratories, which implies high latency and high cost (Jimeno-Sáez et al. 2020). To avoid such inconveniences, most ecologists have recently been using modeling techniques.

In terms of model building, there are typically two approaches in ecological modeling: (1) physically-based (or conceptual) and (2) data-driven-based models (Babovic et al. 2005; He et al. 2014; Zhang et al. 2016). Physical models are complex and require complex mathematical methods, sufficient physical data and high expertise and experience in their implementation (Aqil et al. 2007; He et al. 2014). While data-driven models are easier to implement, not so complex, and remove the need for specialized knowledge of physical processes controlling the transport of pollutants (Bowden et al. 2006), which make them popular and widely used for modeling complex natural processes, mainly in predictive modeling. They can be helpful and valuable for modeling and predicting eutrophication episodes in any natural ecosystem (Nayak et al. 2005; He et al. 2014).

The variables that influence Chl-*a* content in water bodies are numerous and complex (Jimeno-Sáez et al. 2020). In the existing literature, different statistical approaches have been used to predict Chl-*a* concentrations based on regression analyses (Su et al. 2015). However, these traditional methods of data processing usually use a linear relationship to simplify complex problems, resulting in unsatisfactory results because they are not sufficiently efficient to deal with complex non-linear relationships between the variables involved (Su et al. 2015).

Machine learning (ML) algorithms have been shown to be more efficient than conventional data processing methods in assessing water quality (Abba et al. 2017), as they are well suited for predicting non-linear and complex functions. Previous studies have confirmed the superiority of ML over traditional approaches in modeling water quality parameters (Charulatha et al. 2017). Among ML techniques, we can mention Artificial Neural Networks (ANNs). ANNs imitate learning processes of a human brain through training and calibration of the network. This skill makes ANNs useful tools for analyzing complex situations that are difficult to explain with traditional methods (Daliakopoulos et al. 2005; Samarasinghe 2007).

The ability to capture system dynamics and non-linearities makes ANNs especially appropriate for the investigation of natural systems, which usually have distinctive spatial-temporal variability (ASCE 2000).

ANNs algorithms have also been used to study Chl-*a* dynamics, since it is one of the variables representing algal biomass, and has been considered one of the early warning preventive approaches to avoid the occurrence of potential algal blooms (Chen et al. 2015). In a recent study, Li et al. (2017), applied different types of ANNs to estimate the concentration of Chl-*a* in 27 lakes in China, respectively. In recent study, Tian et al. (2017) used an ANN to predict Chl-*a* concentrations to an estuary reservoir in East China. Considering that the ANNs achieved the best result for different water quality parameters in several studies (Abba et al. 2017; Keller et al. 2018), including Chl-*a*, it can be expected that these models will obtain satisfactory Chl-*a* estimation results in this study.

There are also many classifications in ANNs, such as the Back Propagation Networks, Radial Basis Function

Networks, etc. The Back Propagation network is the most commonly used learning algorithm (Rajaee et al. 2019). In this study, a recent ANN approach namely, a nonlinear autoregressive with exogenous inputs (NARX) neural network was developed. NARX is neural network that belongs to the non-linear back propagation dynamic neural networks (Hayken 1999).

In terms of time and cost, reducing the number of variables to be measured is very important. For this reason, it is of a great interest to select specific variables that are most related to Chl-*a* levels. ML provides an efficient technique to do that. It is called the Random Forest model (RF).

RF has been applied in many studies. Indeed, in 2016, Béjaoui et al. have investigated with the RF model the most important predictor variables for Chl-*a* variations in the lagoon of Bizerte, located in the north of Tunisia. In another more recent research study, Béjaoui et al. (2018) have used the RF model to study the dynamic of the plankton in Ghar Melh lagoon, located in the north of the Tunisian Mediterranean coast.

It is known that an early-warning proactive approach of the Ch-*a* content is essential to prevent or mitigate the occurrence of eutrophication episodes, especially in sensitive ecosystems (Jimeno-Sáez et al. 2020) like the North Lagoon of Tunis. For this reason, the use of NARX is essential to perform a one-step ahead forecasting of Chl-*a* values in our study of the North Lagoon of Tunis.

Given the superiority of the ML algorithms, this study has been conducted to achieve the following objectives: (1) to select the specific variables that are the most related to Chl-*a* concentrations in the North Lagoon of Tunis, using, especially, RF model. To do so, different variables combinations were tested, (2) to develop an ANN network to estimate and forecast one step ahead of Chl-*a* concentrations based on NARX neural network and to (3) validate the performance of the model.

Several studies have been carried out on the eutrophication process and water quality parameters of the North Lagoon of Tunis (Ben Charrada 1992; Rezgui et al. 2008; Trabelsi-Bahri et al. 2013). However, to the very best of our knowledge, there is no previous research using ML models to predict water quality parameters in this lagoon, specifically Chl-*a* concentrations.

Materials and Methods

Study Area

The North Lagoon of Tunis is a well-mixed shallow coastal seawater lagoon located in north of Tunisia (36°45′–36°52′ N and 10°10′–10°20′ E) and Southern Mediterranean Sea (Fig. 1). Covering about 22 Km^2 with an average depth of 2 m (range between 0.5 and 3.5 m). This lagoon is one of Tunisia's main shallow water bodies (Trabelsi-Bahri et al. 2013). It is connected to the open sea at the Gulf of Tunis, where water is exchanged with the Mediterranean Sea through Kheireddine channel, which measures 800 m in length and 40 m in width and has a mean depth of approximately 2.5 m (Ben Charrada 1992).

In 1985, a major restoration project has been undertaken in this lagoon to stop pollution and eutrophication (Van Berk and Oostinga 1992). The ultimate goal of this project was to achieve a good chemical and ecological status in this eutrophicated lagoon and to achieve extensive land reclamation around it (Trabelsi-Bahri et al. 2013). The goal was also to reduce the water retention time in the lagoon (Trabelsi-Bahri et al. 2013). These aims have been successfully accomplished by creating an inlet/outlet tide-driven circulation system following the construction of a longitudinal (East-West) separation dam across the lagoon and inlet/outlet gates at the entrance of the canal connecting the lagoon to the open sea (Trabelsi-Bahri et al. 2013). The gates and the separating dam permitted a good circulation (Fig. 1) of the lagoon's water (Van Berk and Oostinga 1992). In addition, the shoreline was straightened to prevent water stagnation (Trabelsi-Bahri et al. 2013).

The restoration project resulted in a clear improvement of the biodiversity (Trabelsi-Bahri et al. 2013). As a matter of fact, in January of 2013 the North Lagoon of Tunis was deemed a "wetland of international importance", Ramsar site (Mdaini et al. 2019). However, being aware of the importance and fragility of this ecosystem, it must always remain under observation.

Data Analysis

The monthly concentrations of Chl-*a* data along with physicochemical parameters of water quality of the North Lagoon of Tunis for the period from January 1989 to April 2018 were collected.

In the present study, a set of seven environmental variables known to affect Chl-*a* concentrations were monitored: Secchi depth, dissolved oxygen, total phosphorus, total nitrogen, pH, water temperature and practical salinity (Sp; IOC et al. 2010) called salinity in this paper. Sampling campaigns have been carried out from February 2014 through April 2018 at five sampling stations (Fig. 1). Water samples were collected about 10–20 cm below the water surface according to standard methods. In addition, Al-Buhaira Invest company, which is in charge of the ecosystem, provided us with long monthly time series, as a part of the monitoring program for the lagoon, in order to gather information on the physical and chemical characteristics of the ecosystem.

All ML techniques in addition to linear model building were performed using the MATLAB software MATLAB® software (version 9.3.0.948333 (R2017b), The Mathworks, MA, USA).

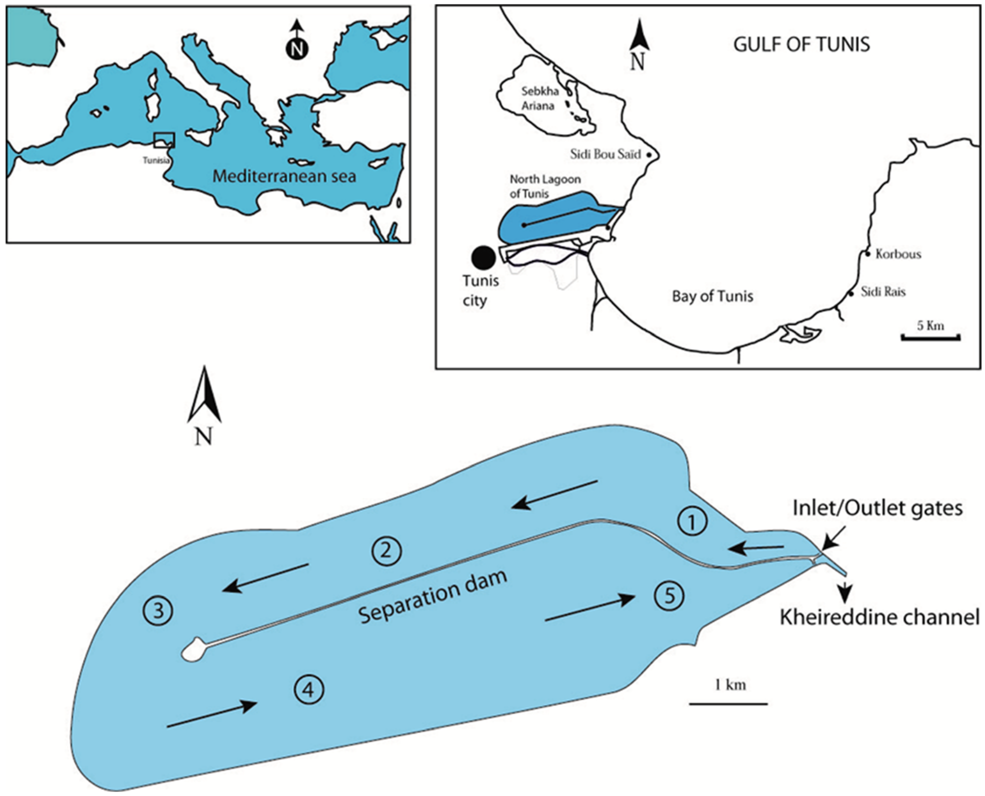

Fig. 1 Location of the study area and water quality monitoring stations (1–5). Arrows inside the map represent the unidirectional inlet/outlet water circulation system

Chl-*a* Concentrations

As above-mentioned, Chl-*a* concentrations were used as the eutrophication indicator, thus to assess the ecological status of the studied area. Water samples of 500 mL for Chl-*a* measurements were filtered through a 0.45 µm pore-size membrane (Millipore). Chl-*a* was extracted in 10 mL of 90 % acetone for 24 h in the dark at -20 °C following the protocols published by Parsons et al. (1984). The extract concentration was analyzed spectrophotometrically according to the method of Lorenzen (1967).

Figure 2 illustrates that the temporal variability of the Chl-*a* in the North Lagoon of Tunis, is relatively low and approximately similar for stations 1, 2, 3, and 4, and is significantly higher at station 5, which is the furthest from the sea water inlet gate.

Physico-chemical Variables

Physico-chemical variables, including water temperature and salinity were measured *in situ* using a WTW LF325 conductivity meter. pH was measured by a pH 330i WTW pH meter.

The Secchi depth of the lagoon was measured with a 25 cm diameter Secchi disc. the Secchi depth is the visibility of the Secchi disc. In fact, the Secchi depth is a parameter indicator of the transparency of the water column and it is the depth of disappearance of the Secchi disc.

Since the end of the restoration works, the visibility of the lake bottoms has improved significantly. In the absence of strong wind the rapport (transparency / depth) generally exceeds 90 % (Shili 1995).

Dissolved oxygen was measured by OXY 197 oxymeter. Figure 3 shows that the temporal (seasonal and interannual) variations of dissolved oxygen is relatively similar at all five stations in the lagoon. This figure further illustrates that the lagoon remains well oxygenated from the inlet to the outlet gates. Thus, suggesting that the dissolved oxygen air-sea exchanges are efficient and that the lagoon ecological state remains healthy.

Surface water samples were analyzed in the laboratory for total phosphorus and total nitrogen. Samples for nutrient determination were collected in 1000 mL acid-washed polypropylene bottles and kept on ice until use.

Nutrient analyses were performed using a spectrophotometric method (Strickland and Parsons 1972) with a UV-visible spectrophotometer. Total phosphorus and total nitrogen were determined after alkaline peroxodisulfate digestion in an autoclave using unfiltered water. For the

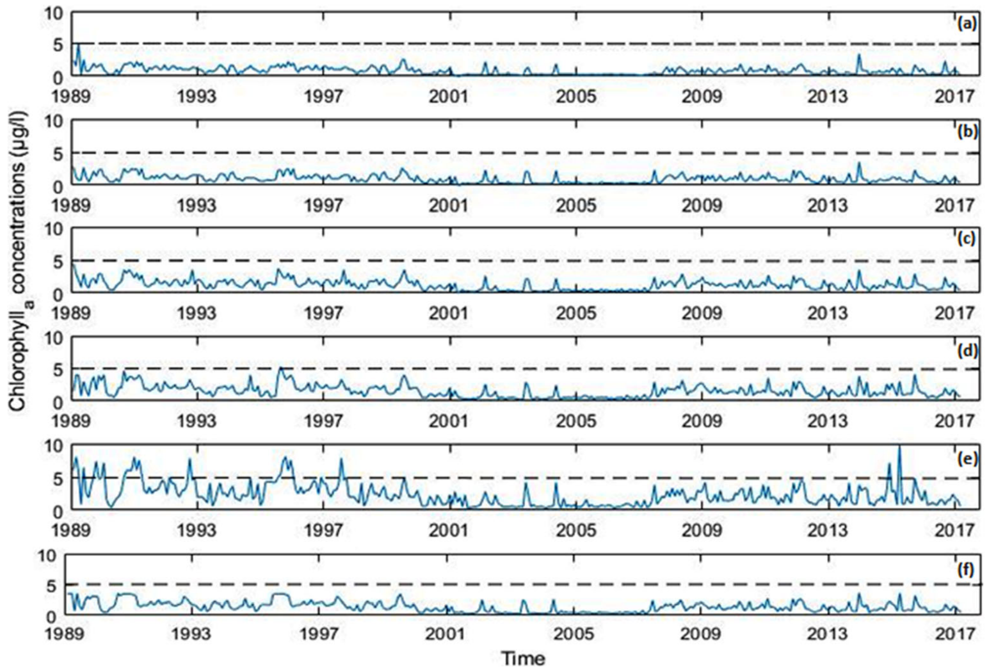

Fig. 2 Temporal variability of Chl-*a* in the North Lagoon of Tunis at: (**a**) station 1; (**b**) station2; (**c**) station 3; (**d**) station 4; (**e**) station 5; (**f**) mean concentration of the five sampling stations

determination of total phosphorus, the phosphorus compounds were mineralized to orthophosphate ions in an autoclave at 100 °C using a solution of sulfuric acid and potassium persulfate.

Determination of total nitrogen compounds required high oxidation of the nitrogenous ions into nitrates in an autoclave using an alkaline solution of persulfate, then by the reduction of nitrates to nitrites by passing through a cadmium column. The nitrites formed were determined using sulfanilamide and N-naphthyl-ethylene.

Modeling Approaches

Analysis of Variance (ANOVA)

Analysis of variance ANOVA was performed to ascertain if there were any significant difference in physico-chemical conditions and in Chl-*a* concentrations among the sampling stations in the North Lagoon of Tunis. ANOVA was also conducted to verify any significant relationship between Chl-*a* and the physico-chemical parameters.

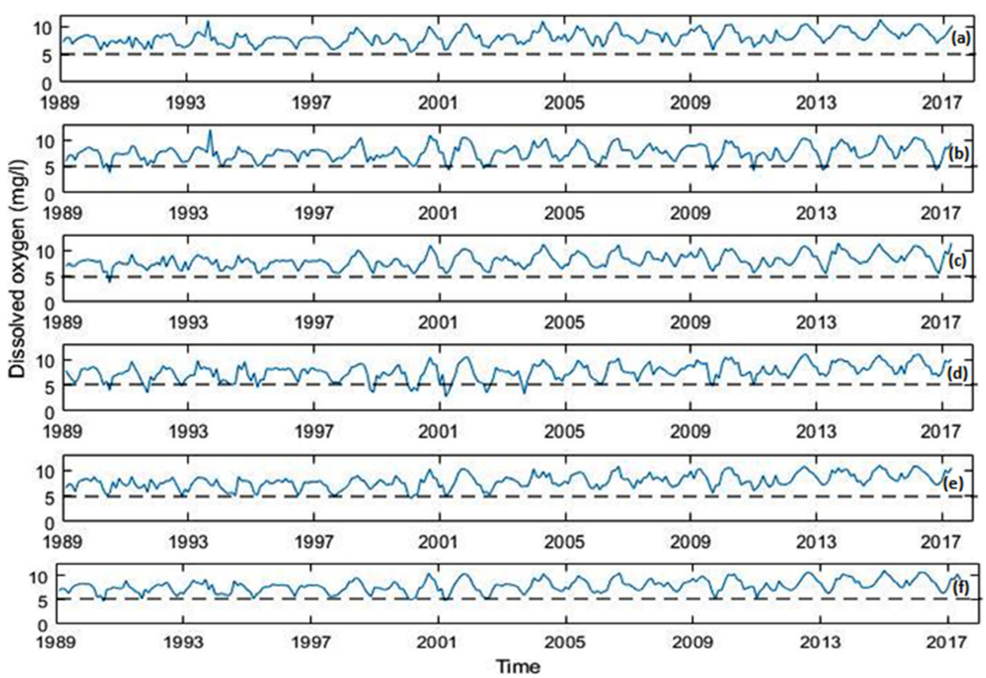

Fig. 3 Temporal variability of the dissolved oxygen in the North Lagoon of Tunis at: (**a**) station 1; (**b**) station 2; (**c**) station 3; (**d**) station 4; (**e**) station 5; (**f**) mean concentration the five sampling stations

Random Forest

ML algorithms are typically implemented with a set of predictor variables (input variables) and one or more target variables (output variables) represented as a continuous value for regression problems (Kohavi and Jhon 1997) where maximizing accuracy is the main goal of predictive modeling (Motoda and Liu 2002). To estimate a parameter of water quality we can use all available predictor variables, or select a smaller number of them. This can result in the inclusion of too few or too many inputs to the model, both of which are undesirable (Maier et al. 2010). To address this issue, a predictor variable selection stage had been considered in this study to eliminate redundant data. The aim of reducing the number of predictor variables in ML is to speed up the process of the learning algorithm to boost predictive accuracy and increase the comprehensibility of learning results (Motoda and Liu 2002). There are many parameters that influence the concentration of Chl-*a*. This study performed a RF model to determine the appropriate predictor variables that are the most important to the Chl-*a* concentration. RF modeling is a relatively recent ML approach based on decision trees and trained on a collection of input predictor variables in order to obtain an accurate prediction of the output variable (Breiman 2001).

RF approach has many advantages. First, no probability distribution of predictor variables is assumed. Second, it can handle a large number of variables, selecting the most useful ones among them (Mulia et al. 2013; Park et al. 2015). Third, RF predictions are exceptionally accurate, since they come from the average set of many simple models, thereby avoiding the over-fitting problem typical of many non-linear regression techniques (Phillips et al. 2008; Huang et al. 2015). Fourth, Because each tree is built on a random subset of the original data, no separate independent dataset or cross-validation approach is required to test the predictive performance of the model (Were et al. 2015). Finally, RF technique is adequate for natural ecosystems where there is a large amount of physico-chemical and biological variables that have complex relationships.

An interesting aspect of the RF model is that it can provide a quantitative measure of the importance of the various predictor variables in the final result, which can be useful in choosing the most important ones.

The method used to evaluate the ranking of the most important predictor variables from the RF model is the out-of-bag (OOB) technique by permutation; a technique that measures how influential the predictor variables in the model are at predicting the response variable (Chl-*a*). The effect of the predictor variable increases with the value of this measure.

If a predictor variable has an effect on the prediction, then the permutation of its values should have an impact on the model error. If a predictor variable is not influential, the permutation of its values should have little to no effect on the model error (Mitchell 2011). It consists of calculating the gain in the mean square error, which is computed by permuting OOB data: for each tree, the prediction error on the OOB portion of the data is recorded; the same is done by permuting each predictor variable (Mitchell 2011). The differences between the two OOB errors are then averaged over all trees and normalized by the standard deviation of the differences (Mitchell 2011).

The mean of squared residuals (MSE) and coefficient of correlation (R) were used to evaluate RF model performance. MSE is a quantitative measure of the error acquired by the model when a prediction for the target variable is made. It can be sensitive to outliers and is best used in conjunction with other metrics when outliers are present to evaluate a given model (Cutler et al. 2007). If the MSE is close to 0, it indicates a very close approximation to the actual values. The MSE is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2 \tag{1}$$

where:

y_i and $\widehat{y_i}$ denote the modeled concentration and the observed concentration of Chl-*a*, respectively and n is the number of data in each data set.

Prediction accuracy R represents the degree of correlation between the prediction values and the observed values, and a high R value (close to 1) means the prediction is close to the observed value (Xu et al. 2019).

$$R = \left(1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y_i}\right)^2}\right)^{\frac{1}{2}} \tag{2}$$

where:

$\bar{y}_i$ denotes the average of the observed values of Chl-*a*.

It is to be mentioned that the coefficient of determination (R^2) was calculated from R to contribute in the assessing and comparing the performance of the models. Thus, useful information can be obtained concerning the relative importance of all variables and their capability of forecasting Chl-*a* concentrations.

The RF model was simulated twice. First, considering only the seven physico-chemical predictors we had for predicting Chl-*a* concentrations. Second, we checked whether for an observable spatial or seasonal dependence among Chl-*a* predictors. In other words, we checked whether the predictions of Chl-*a* concentrations could be improved by considering the sampling stations and the seasons as observed variables. This was done by including two new categorical predictor variables representing the five stations and the four seasons, respectively. This is possible because RF models can handle both quantitative and qualitative predictor variables.

Nonlinear AutoRegressive with eXogenous Inputs (NARX) Neural Network

ANN is a massively parallel-distributed information-processing tool that aims to simulate the functioning of brain neurons via a network of artificial neurons organized into layers (Haykin 1999). The network receives a stimulus and converts this input into a signal output via a transfer function (Jimeno-Sáez et al. 2018). Because of its ability to assign meaning to input parameters and to map the inputs to the outputs, the ANN model is an effective modeling technique when the relationships between the variables of the underlying physical processes are complex or uncertain (Wei et al. 2001). These neural networks are a non-linear modeling tool that can manage a large number of inputs to determine one or more outputs (Fogelman et al. 2006).

There are many types of ANNs for different applications. Generally, in ANNs, the direction of information flow between nodes or neurons, from the input to the output layer, and each node in a layer is connected to each of the nodes in the next layer (ASCE 2000). The neurons are linked to other neurons via links that have an associated weight that reflects their strength of connection and stores network information (Jimeno-Sáez et al. 2018). The important feature of the transfer function is that as input values changes, it provides a smooth, distinguishable transition. In other words, a minor change in the data, produces a minor change in the output (Jimeno-Sáez et al. 2020). This transfer function is used within a certain proper range to constrain the outputs of neurons (Xu et al. 2019). The most common threshold transfer functions include a linear function, nonlinear gradient descent function, stepwise function, and S-shaped function (Xu et al. 2019). In NARX, the transfer function for the hidden layer is an S-shaped (e.g. sigmoid) function, and the transfer function for the output is a linear function (Xu et al. 2019).

Then, the propagation function computes the input to a neuron from the outputs of its predecessor neurons and their connections as a weighted sum (Rajaee et al. 2019). A bias term can be added to the result of the propagation (Rajaee et al. 2019).

Therefore, the connection weights, biases, transfer and propagation functions parameterize the mathematical relationship between inputs and outputs of the network (Nguyen et al. 2007). These weights and biases need to be adjusted in the training process of the networks to minimize the model error (Jimeno-Sáez et al. 2020).

NARX model is a dynamic recurrent neural network that encloses several layers with feedback connections (Hayken 1999). It has previously been applied by many researchers to model nonlinear processes. In 1996, Lin et al. stated that NARX network is a powerful modeling and validation tool with a much faster convergence that generalizes much better than other ANNs. There are many applications for the NARX network in representing nonlinear dynamic behaviors.

A NARX model is defined as follows:

$$y(t) = f\left(y(t-1), y(t-2), , y(t-n_y), u(t-1), u(t-2), , u(t-2), , u(t-n_u)\right) \quad (3)$$

where: f is generally unknown and can be approximated, $u(n)$ and $y(n)$ denote the input and output of the model at discrete time step n, respectively.

The first step in a NARX model is to determine the input and output variables. In our study, the output is the Chl-*a* variable and the input variables are those having the highest permutation importance according to RF model. To ascertain that, different predictor variables combinations were tested. The next step is to set up the network configuration, which consists of determining the number of neurons in the hidden layer and the number of time delays in the input layer to maximize modeling ability. The prediction accuracy (weights and biases) can be improved by adjusting these two parameters (Xu et al. 2019). There is no default criterion to determine the best structure, therefore we evaluated the performance of the network with different structures, after a training phase mostly based on the size of errors, such as checking the MSE and R, as in the RF model. In addition, the error autocorrelation function and input-error cross-autocorrelation function were also checked to evaluate the NARX performance. The autocorrelation error function describes how prediction errors are related in time. For a perfect model of prediction, the difference between the two errors should be small enough to be statistically insignificant. This would mean that the prediction errors are entirely uncorrelated with each other. This also means that the values of error autocorrelation should mostly be within a certain confidence interval of 95 % (Xu et al. 2019).

The input-error cross-correlation function describes how the errors are correlated with the input sequence. For the ideal prediction model, all correlations should be zero, except for the one at zero lag (Xu et al. 2019).

Three training algorithms, which are the fastest and most commonly adopted in NARX training were tested: Levenberg-Marquardt algorithm, the backpropagation algorithm and the Bayesian regularization algorithm.

Multiple scenarios with different predictor variables combinations (inputs) were tested to simulate the NARX model, and the one with the best performance were used to develop the network.

Results and Discussion

Data Properties

ANOVA revealed no significant difference among the sampling stations in the lagoon for any of the physico-chemical variables or Chl-*a* concentrations, which is in general good agreement with the observed data sets (except for Chl-*a* at only station 5; Figs. 2 and 3). Therefore, all data were grouped by months to reconstruct the monthly dynamics (mean of the five stations) of each variable in the North Lagoon of Tunis (such as in Figs. 2f and 3f).

In general, temperature, salinity and dissolved oxygen measurements show typical trends of the southern Mediterranean coastal marine climate (De Casabianca et al. 1991; Armi et al. 2012; Dhib et al. 2016). The temperature values indicate an apparent seasonal pattern. The minimum and maximum temperatures recorded in winter (January) and summer (July and August) varies between 8 and 30 °C, with an annual average of 19 °C. salinity also shows a seasonal trend with a minimum of 32.75 in July and a maximum of 42 in November, a mean of about 36.5 in the lagoon.

The concentrations of dissolved oxygen display a strong seasonal pattern, ranging from 5 mg/l in summer (July) to 11.5 mg/l in winter (December) in the lagoon. Concerning the eutrophication indicator (expressed as Chl-*a*), its trend shows an increase from winter (January) to summer (June) reaching its maximum monthly concentration of about 3.65 µg/l in the lagoon. The total phosphorus concentrations ranged from 4 µg/l to 65 µg/l in the North lagoon of Tunis. Total nitrogen monthly concentrations in the lagoon ranges between 200 and 1100 µg/l. Nutrient concentrations (total phosphorus and total nitrogen) exhibit a common seasonal variation, being relatively high in autumn and winter and low in summer. The monthly pH values ranged from 7.2 to 8.7 during our study period, in the North Lagoon of Tunis. A pH range of 6.5 to 8.5 is acceptable for aquatic biota according to the APHA (1999). So we can state that the North Lagoon of Tunis is in a good agreement with these limits.

Random Forest

Chl-*a* is one of the most relevant markers of water bodies' presence and degree of eutrophication (Lu et al. 2016). In the North Lagoon of Tunis, Chl-*a* monthly mean concentrations range from a minimum of 0.22 µg/L and maximum of 3.65 µg/L. Chl-*a* levels can have complex relationships with both nutrient components (total phosphorus, total nitrogen) and water quality variables (salinity, pH, temperature, dissolved oxygen and Secchi depth) in coastal ecosystems (Jimeno-Sáez et al. 2020). The RF is a suitable technique from ML algorithms, when it comes to dealing with complex relations between variables. The RF model was trained on the North Lagoon of Tunis data-352 samples of 7 predictor variables (Secchi depth, dissolved oxygen, temperature, salinity, total nitrogen, total phosphorus and pH) and one target variable, the Chl-*a*. The implementation gave an R^2 measure of about 0.62 and MSE equal to 0.28. Figure 4 shows the ranking of predictor variables according to their importance by OOB technique by permutation. Only a few descriptors contributed noticeably to the estimation of the Chl-*a* content namely, Secchi depth followed by the dissolved oxygen and pH.

The most important pigment in aerobic photosynthetic organisms is Chl-*a*. The depth will impact the strength of sunlight in water and thus the photosynthesis of most algae (Frolov et al. 2012), explaining the strong correlation between the depth and Chl-*a*.

Algae produce oxygen during the day and absorb it during the night. Oxygen absorption also occurs during the process of algae death and decay (Béjaoui et al. 2016). In agreement with that, our findings have shown that dissolved oxygen is also associated with Chl-*a* concentrations.

In addition, several studies have demonstrated the strong correlation between Chl-*a* and pH (Menendez et al. 2001; Zang et al. 2010; Wallace et al. 2016).

In decreasing order of importance, the other predictor variables included in the RF model were: total phosphorus, total nitrogen, salinity and temperature.

A direct comparison (scatter plot) of the observed and predicted Chl-*a* concentrations is shown in Fig. 5. The fitted RF model was much better than the one reported by Béjaoui et al. (2016) for Bizerte lagoon ($R^2 = 0.51$), and similar to the one reported by Béjaoui et al. (2018) for Ghar el Melh lagoon ($R^2 = 0.64$). Both lagoons are located in the Mediterranean coast of north Tunisia. Hence, for the North Lagoon of Tunis, the observed Chl-*a* concentrations were more accurately predicted than those of Bizerte lagoon. It is known that for predictive modeling, the number of the observed data is very important for the accuracy of the model (Béjaoui et al. 2016). We used long-term of monthly observations that lasted approximately three decades in the North Lagoon of Tunis, which makes the results of the RF accurate in the studied ecosystem.

For comparison, a Multivariate Linear Regression (MVLR) model was fitted in addition to the RF model. The linear model parameters (Estimate) between predictor variables and Chl-*a* concentrations were almost consistent with the relationships found with the above RF model (Table 1). Both Secchi depth and dissolved oxygen were the two most important predictors explaining Chl-*a* concentration levels. Thus, the linear model quantitatively confirmed the outcomes of the RF model.

MVLR gave an R^2 of about 0.29. It's obvious that the RF model captures more efficiently the dependency of Chl-*a* concentrations on other variables than the MVLR. The quality of the results is ensured by using the OOB procedure by permutation. We can thus confirm that the RF model could be used to better understand more complex dependencies among variables since it has several advantages over traditional

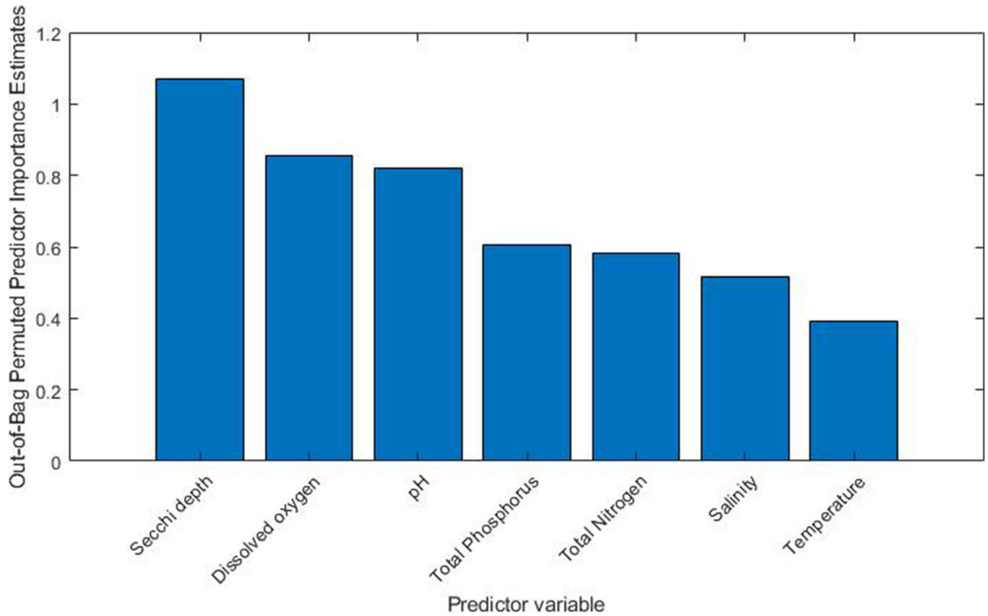

Fig. 4 Predictors importance ranking for the "first" RF model to predict Chl-*a* content in the North Lagoon of Tunis. The importance of each predictor is measured using the OOB technique by permutation due to each predictor

correlative analyses (i.e. a decrease in outlier sensitivity, no implicit assumptions on data distribution).

According to the MVLR, the Chl-*a* concentrations had a significant correlation with water quality variables, as secchi depth, dissolved oxygen followed by total phosphorus in the study area. However, the weakness of its performance suggests that the use of traditional regression methods in the modeling of such a complex process is meaningless, so there is a great need to use more effective techniques (Mjalli et al. 2006).

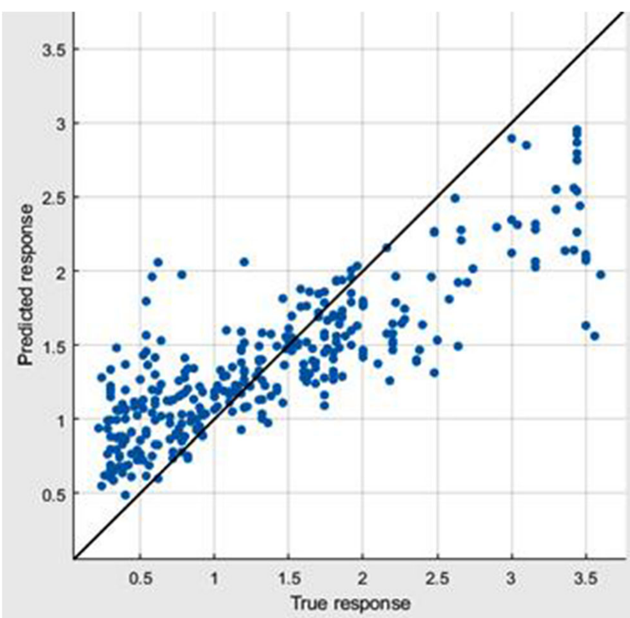

Fig. 5 RF Prediction of the Chl-*a* concentrations using the physicochemical predictor variables data in the North Lagoon of Tunis. Predicted response is Predicted Chl-*a* values and True response is Observed Ch-*a* values

This fact may support the conclusion drawn by Maier et al. (2010) that using a linear approach to define which of the potential input variables have a significant relationship with the model output is not sufficient for the development of ANN models.

All variables, were transformed, to normalize their distribution prior to any modeling analyses. However, the transformations did not improve the performance of the MVLR.

The relationship between all variables is strictly nonlinear, which is expected. Natural ecosystems are governed by a number of complex processes due to the impacts of hydroclimatic variables such as evaporation, temperature, precipitation, etc. and anthropogenic contribution (Schramm 1999; Viaroli et al. 2008).

In recent studies of similar lagoons located in the north of Tunisia, Chl-*a* concentrations were found linearly not related to the physico-chemical parameters (Béjaoui et al. 2016, 2018).

We performed all our modeling directly on the original data using ML techniques, known for their abilities to deal with non-linear complex time series processes. Fitting a model directly without transformation is advantageous for forecasting, because forecasts are returned on the original scale.

A second RF model was also fitted by adding two new categorical variables: station (for an observable spatial dependence) and season (for an observable seasonal dependence). The performance of the RF slightly decreases and attains an R^2 of about 0.59.

The addition of a new predictor variable containing spatial information (station) appears to have little importance on the model simulating Chl-*a* concentrations (Fig. 6). The existence of a strong correlation between Chl-*a* concentrations and secchi depth showed that depth data itself might be sufficient

Table 1 Regression coefficients between Chl-*a* concentrations and physico-chemical variables in the North lagoon of Tunis using linear model (coefficients marked with (*) are statistically significant at p-value < 0.05)

	Estimate	Std. Error	t-value	p-value
Intercept	3.434	2.3318	1.4727	0.14175
Dissolved oxygen	-0.12818	0.05122	-2.5026	0.012792*
Total Phosphorus	0.012391	0.00753737	2.3059	0.021713*
Secchi depth	-0.47539	0.17532	-2.7116	0.007032*
Temperature	0.0077276	0.013261	0.58275	0.56045
pH	-0.19153	0.21548	-0.8882	0.37472
Salinity	0.027982	0.032362	0.86464	0.38784
Total Nitrogen	-0.00012694	0.0001341	-0.94663	0.34449

for demonstrating the spatial variations of Chl-*a* concentrations in the lagoon without including the categorical variable (station). Moreover, for the RF model, the categorical variable (season) did not really had an important effect on Chl-*a* concentrations. This finding was expected since the variable (temperature) can interfere with the variable (season). Additionally, we may state that this can be due to the climate of Tunis, which stays relatively warm all along the year, including winter.

RF is a good predictive technique to study the correlations between physico-chemical and/or biological variables in coastal ecosystems. Béjaoui et al. (2016), showed that mainly dissolved inorganic nitrogen (NO_3) along with dissolved oxygen are the greatest contributors to Chl-*a* content in Bizerte lagoon. Furthermore, in 2018, Bejaoui et al. reported that temperature and silicates are the two most strongly correlated variables to the plankton dynamics in Ghar Melh lagoon. Although, the influence of the predictor variables of Chl-*a* were different in several research works, the dissolved oxygen and secchi depth generally were among the main variables.

For example, Palani et al. (2008) applied the ANN model with location variables, orthophosphates (PO_4) dissolved oxygen and temperature as the explanatory variables to predict Chl-*a* concentration. Li et al. (2017) selected the concentration of total phosphorus and total nitrogen, temperature, secchi depth, and dissolved oxygen among the most influential input variables for Chl-*a*, using a genetic algorithm optimized back-propagation neural network. Furthermore, Kuo et al. (2007) defined the Chl-*a* model by the input of month, temperature, pH, secchi depth, suspended solids (SS), PO_4 and NO_3.

It is important to highlight that the difference in the RF results between the previous ecosystems, with the North Lagoon of Tunis can be explained by the ecosystem specificities as dimensions of water masses, different eutrophic states, water depth and communication with the sea. In addition, various modeling approaches, in addition to different field works and laboratory analysis techniques may have contributed to these differences.

Nonlinear AutoRegressive with eXogenous inputs (NARX) Neural Network

In this study, four scenarios with different input combinations of the predictor variables are tested for estimating and forecasting Chl-*a* concentration values in the North Lagoon of Tunis using the NARX network.

The first input scenario (S1) considered all parameters we had as inputs without selection. The second scenario (S2) included

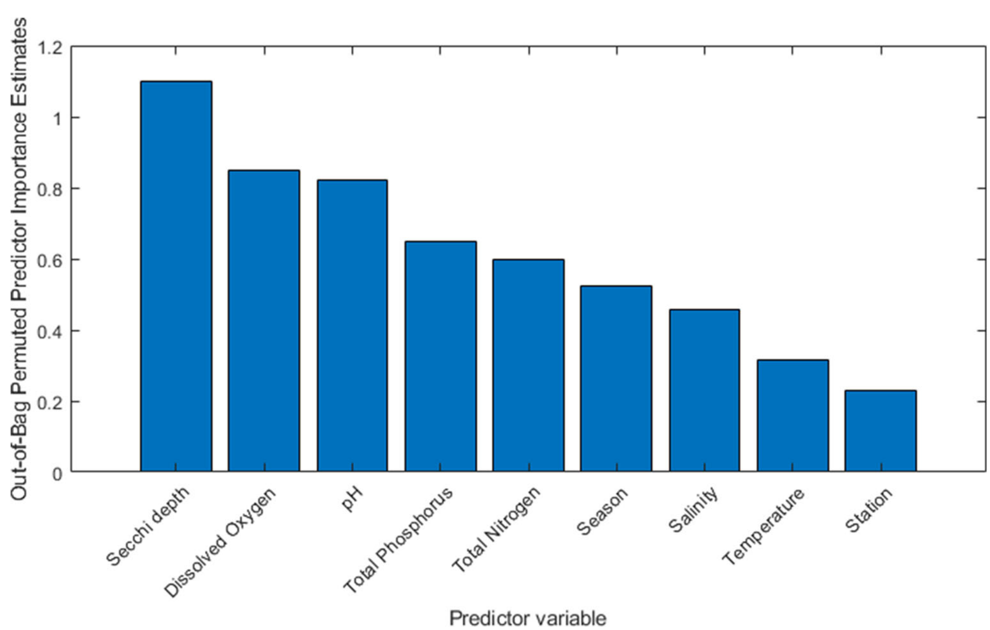

Fig. 6 Predictor importance ranking for the "second" RF model to predict Chl-a content in the North Lagoon of Tunis. The importance of each predictor is measured using the OOB technique by permutation due to each predictor

Table 2 Summarized results of predictor variables selection

Algorithm	N. of predictor variables (inputs) selected	Predictor variables selected	Input scenario
NARX without inputs selection	7	All	S1
NARX with three most important predictor variables selected with RF	3	Depth Dissolved oxygen pH	S2
NARX with the correlated predictor variables selected with MVLR	3	Depth Dissolved oxygen Total phosphorus	S3
NARX with two most important predictor variables selected with RF	2	Depth Dissolved oxygen	S4

only the three most important predictor variables according to the RF model. The third input scenario (S3) included only the most highly correlated parameters according to the MVLR. the last scenario (S4) included only the two most important predictor variables according to the RF model. Summarized results of predictor variables selection is shown in Table 2.

The NARX models with the four input scenarios described in Table 2 were developed to simulate the Chl-*a* concentrations. The four versions of each model represent four substantially different Chl-*a* models, due to the different combinations of variables used as predictors. Different ML models are compared based on statistical indices (Jimeno-Saez et al. 2020) such as R, R^2, MSE, etc. These performance measurements are summarized in Table 3 for the NARX network. Different parameters are tried for each NARX model and the best one; with the minimum MSE and maximum R and R^2 in selected for the forecast of Chl-*a* task.

The topology with 10 neurons in one hidden layer and 2 lags in the input variables provided the best performance in the prediction of Chl-*a* concentrations among all the scenarios. Our study considered that a proportion of 70 % training, 15 % validation and 15 % test is a favorable implementation.

The Levenberg–Marquardt algorithm, an extensively recognized training algorithm, was used for minimizing nonlinear functions. Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples (Xu et al. 2019).

The comparative results between the four versions of the NARX model reveal that the NARX with two inputs selected by the RF algorithm yielded the best accuracy among all the developed NARX models in term of higher R and R^2 and lower MSE values ($R = 0.79$; $R^2 = 0.62$; $MSE = 0.31$).

With three inputs, selected also according to the RF model, the S2 scenario is the second most accurate model with a performance close to the best one. Because simulating time and over-fitting risks increase with the number of predictor variables in predictive modeling, a good practice is to create a model using as few predictor variables as possible (Jimeno-Saez at al. 2020).

We now present one step-ahead (a month) forecasting results for the eutrophication indicator considering the three datasets: Chl-*a* concentrations as the target, using secchi depth and dissolved oxygen as external inputs.

Most of the NARX model errors were very close to zero and fall within the confidence interval (Fig. 7), therefore the autocorrelation errors were negligible. In general, the input-error cross-correlation plot (Fig. 8) showed that all the correlations fell within the confidence interval.

Furthermore, the obtained results (Fig. 9) show an overall correlation of $R = 0.79$ between the actual data (targets) and the predicted values (the outputs). The Error histogram was checked. It presents a closely bell-shaped normal distribution of the errors (Fig. 10). Given that, we can conclude that, residuals of the NARX model are uncorrelated and normally distributed.

The model's fitness is described in Fig. 11 and a visual comparison of Chl-*a* concentrations predictions with respect to the observed data is shown. There is a fairly good match between the observed values and the fitted values. The NARX network was able to predict the high variability of Chl-*a* concentrations, therefore, the fitted model seems to be mathematically accurate and the NARX could be used on a new data set. Given its effectiveness, multiple studies used neural networks techniques to model the Chl-*a* contents as an eutrophication indicator in coastal ecosystems. We can mention the study of Nazeer et al. (2017), who suggested, using ML methods, such as ANN for a more accurate and efficient routine monitoring

Table 3 Performance of Chl-*a* estimation from NARX models based on four different input scenarios

Scenarios	R	R^2	MSE
S1	0.68	0.46	0.42
S2	0.75	0.56	0.31
S3	0.74	0.55	0.40
S4	0.79	0.62	0.31

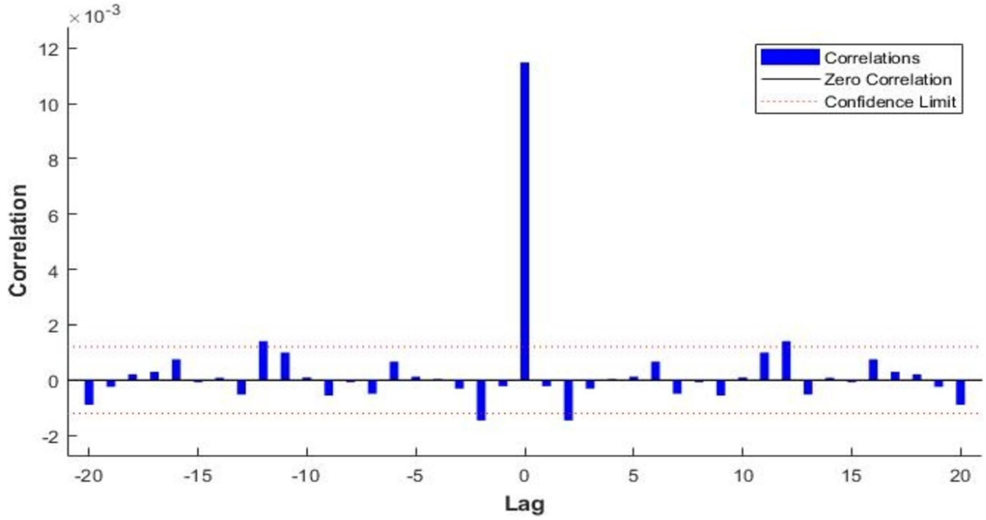

Fig. 7 The autocorrelation error plot

of coastal water quality parameters, particularly Chl-*a*, in a coastal area of Hong Kong. In another study carried out in the Mar Menor lagoon in Spain, a Multilayer Neural Networks have been used for the eutrophication modeling, considering Chl-*a* as the eutrophication indicator (Jimeno-Sáez et al. 2020). In 2003, Lee et al. used back-propagation learning algorithm for training the ANN to predict the algal bloom dynamics of the coastal waters of Hong Kong using a 4-year set of phytoplankton abundance data. Lu et al. (2016), used a back-propagation ANN model for the prediction of Chl-*a* concentrations in lake Champlain in China.

By comparing the NARX results ($R = 0.79$; $R^2 = 0.62$) with the MVLR including just the two most important predictors for the Chl-*a* concentrations, the performance of the final linear model decreases further and attains an R^2 of about only 0.2. We can thus confirm that the relationship between the variable predictors and Chl-*a* concentrations is obviously non-linear and the use of the NARX, RF and generally ML techniques, is adequate for forecasting Chl-*a* contents in the studied lagoon.

The forecasting of the Chl-*a* content one month ahead gave a value about 0.51 µg/L, which was close to the observed value (0.5 µg/L). These values are very close, which indicates a relative normal Chl-*a* level of monthly variation in the lagoon and the accuracy of the developed NARX model.

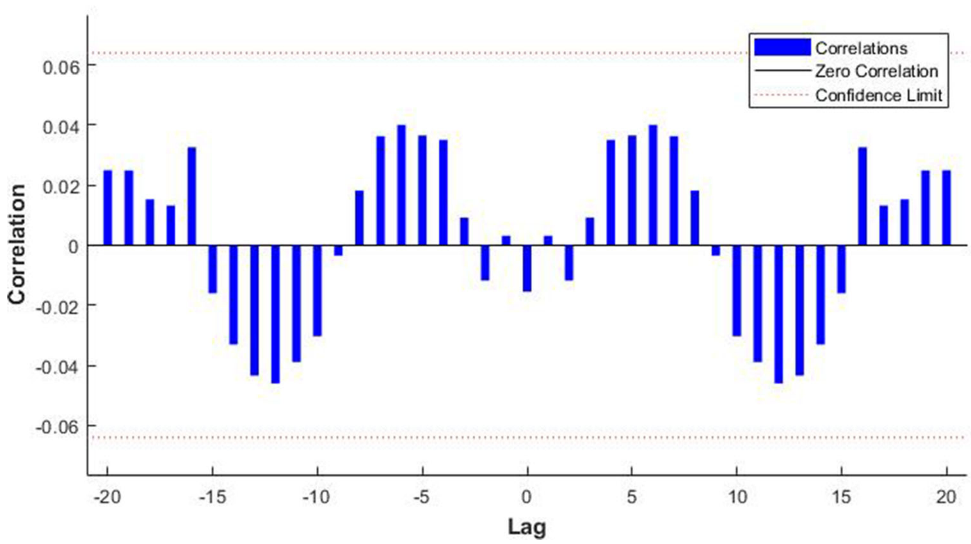

Fig. 8 The input-error cross-correlation plot

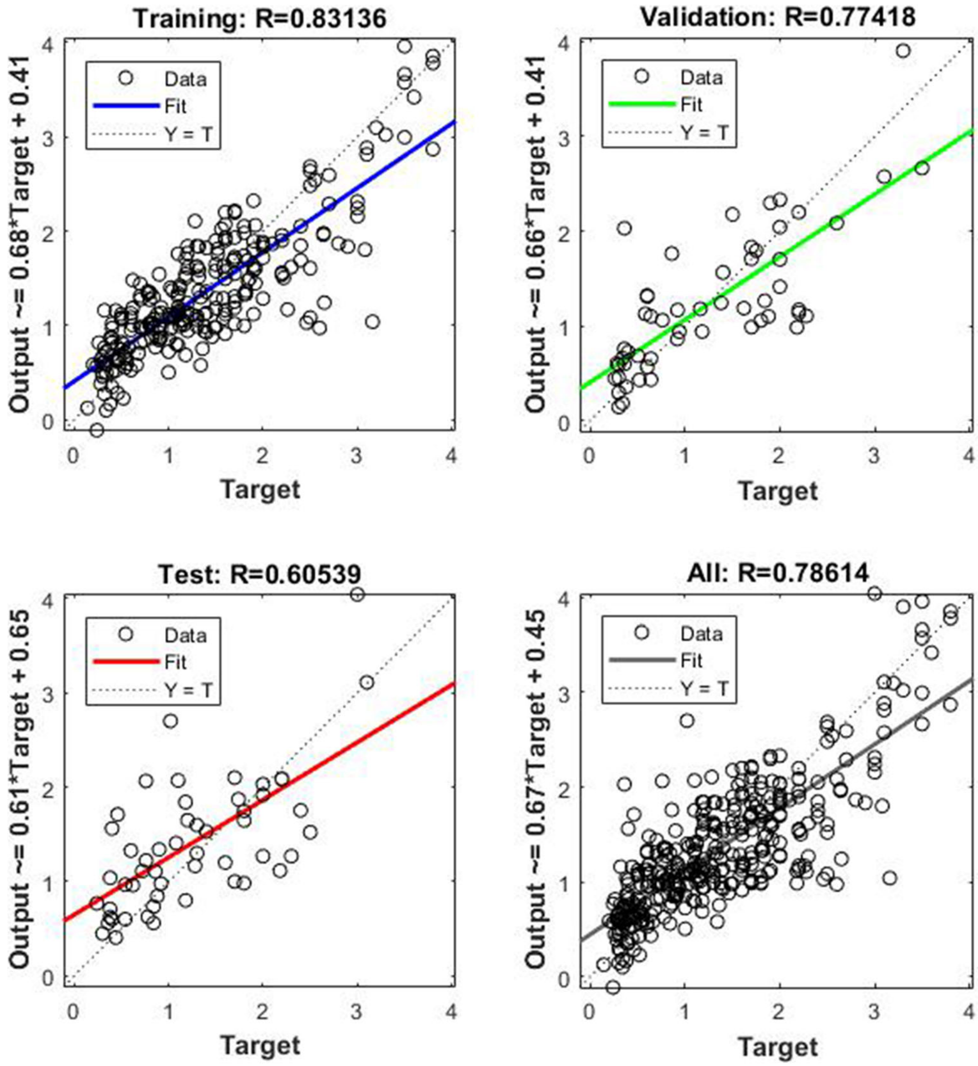

Fig. 9 Correlation between original (target) and predicted (output) Chl-a values obtained with the NARX network

Conclusion

The ecological stability of the North Lagoon of Tunis makes it of a significant socio-economic and ecological values. Multiples services are provided in this ecosystem, such as in tourism (water sports), in fisheries, and in the conservation of sea birds. Thus, it is necessary to improve our understanding of the eutrophication process and of the interactions among the water quality parameters in the lagoon, to adopt sustainable management strategies. One of the most important indicator of the presence and degree of eutrophication in water bodies is the Chl-*a* content.

The approach proposed in the current study relies on a combination of ML methods, using NARX neural network and RF model to predict and forecast Chl-*a* concentration in the North Lagoon of Tunis. Mainly secchi depth along with dissolved oxygen are the greatest contributors to this eutrophication assessment and forecasting. Our results agree well with findings from other studies carried out on Mediterranean coastal lagoons. It's worth mentioning that, secchi depth and dissolved oxygen are very practical variables to measure, without the need of extra laboratory analysis.

The NARX developed was able to predict Chl-*a* concentration dynamics fairly well using minimal input predictor variables. Our results show that complex behavior in the eutrophication process could be modeled using the NARX technique and, some extreme values were successfully estimated.

The results confirmed the significance and usefulness of intelligent modeling as a tool that is simple, rapid, easy to operate and not costly. The developed model can be used to (1) estimate Chl-*a* concentrations when the real value is not available and to (2) simulate different water quality scenarios for extreme ranges of input and output parameters.

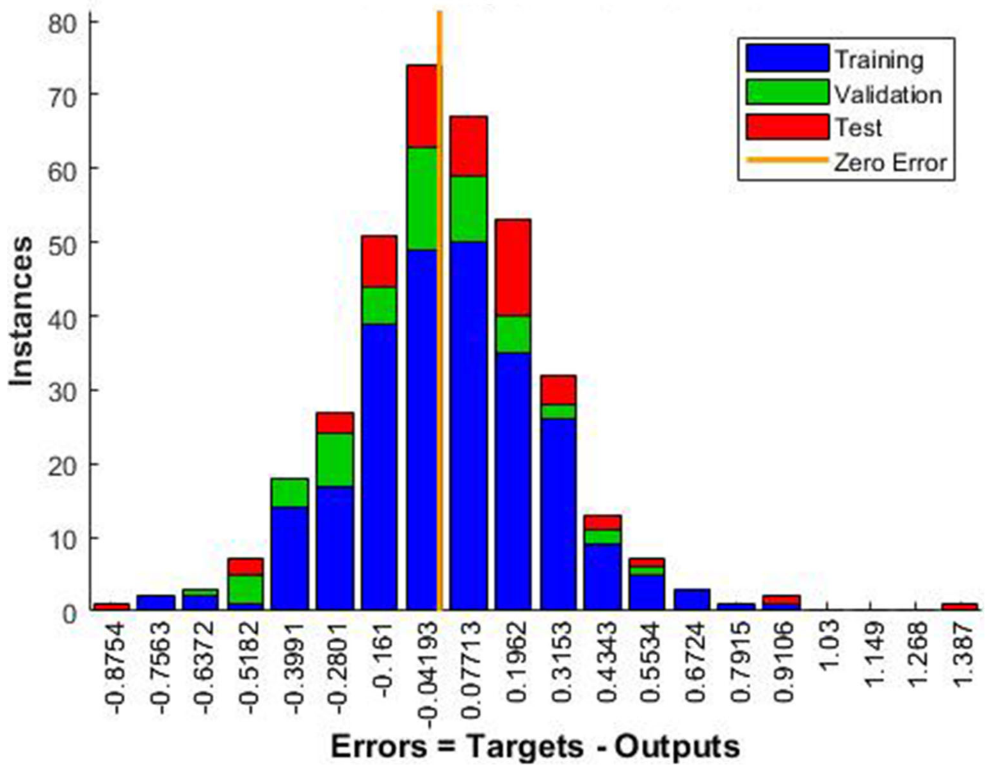

Fig. 10 Error histogram of the NARX network

It is important to mention that despite the important amount of the observed data (approximately three decades) used for developing the NARX, it has a very short computational time.

In the wider context of the study of coastal lagoons and other transitional ecosystems, our approach could be used to assess and predict the eutrophication process of these natural environments and help in decision-making by civil authorities, as well as by engineers, economists, investors and other interested stakeholders.

As perspectives, in order to improve the accuracy of the model we would suggest, adding more data either by simulation (in interpolating the available data), or ideally by performing daily or weekly measurements, at least for the most important parameters (Chl-*a*, dissolved oxygen, transparency, nutrients), maybe not at all the stations but at a

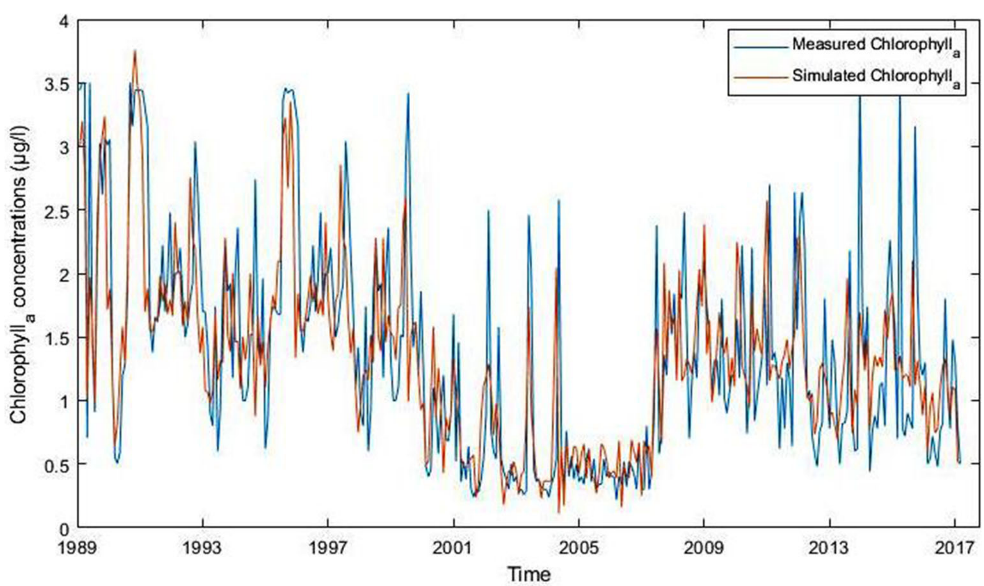

Fig. 11 Measured and simulated Chl-a concentrations using NARX network in the North Lagoon of Tunis

minimum at the two stations 2 (in the north zone of the lagoon) and 5 (in the south zone of the lagoon).

Acknowledgements We thank Al-Buhaira Invest company for providing the first author with the data.

Author Contribution NBH wrote the paper, collected data, conceived, designed and performed the analysis. CG helped in the correction of the paper and contributed in the designing of the analysis. HC helped in conceiving and the designing of the analysis and was a major contributor in writing the manuscript. NBM provided a big sequence of the monthly field investigation data. VG contributed in implementing and working with the software. AS helped in the supervision of this work and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Funding No funding has been provided. However, the field work and laboratory analyses were a part of an institutional activity (Al-Buhaira Invest company).

Data Availability The data provided by the Al-Buhaira Invest company to the first author are acknowledged.

Code Availability Neural Net Time Series Toolbox and Regression Learner Toolbox applications were used in the MATLAB® software (version 9.3.0.948333 (R2017b), The Mathworks, MA, USA).

Declarations

Conflicts of Interest/ Competing Interests No conflict of interest associated with this publication.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Ethical Approval Not applicable.

References

Abba SI, Hadi SJ, Abdullahi J (2017) River water modelling prediction using multi-linear regression, artificial neural network and adaptive neuro-fuzzy inference system techniques. Procedia Computer Science 120:75–82. https://doi.org/10.1016/j.procs.2017.11.212

APHA (1999) Standard methods for the examination of water and wastewater. American Public Health Association, Washington DC

Aqil M, Kita I, Yano A, Nishiyama S (2007) A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behavior of runoff. Journal of Hydrology 337:22–34. https://doi.org/10.1016/j.jhydrol.2007.01.013

Armi Z, Trabelsi E, Turki S, Ben Maïz N, Mahmoudi E (2012) Composition and dynamics of potentially toxic dinoflagellates in a shallow Mediterranean lagoon. Oceanological and Hydrobiological Studies 41(3):25–35

ASCE Task Committee (2000) Artificial neural networks in hydrology I. Preliminary concepts. Journal of Hydrology and Engineering 5: 115–123. https://doi.org/10.1061/(asce)1084-0699(2000)5:2(115)

Babovic V, Sannasiraj SA, Soon Chan E (2005) Error correction of a predictive ocean wave model using local model approximation. Journal of Marine Systems 53:1–17. https://doi.org/10.1016/j.jmarsys.2004.05.028

Barnes RSK (1980) Coastal lagoons: The natural history of a neglected habitat. Cambridge University Press, Cambridge

Basset A, Elliott M, West RJ, Wilson JG, Estuarine (2013) Estuarine and lagoon biodiversity and their natural goods and services. Estuarine, Coastal and Shelf Science 132:1–4. https://doi.org/10.1016/j.ecss.2013.05.018

Béjaoui B, Armi Z, Ottaviani., Barelli E, Gargouri-Ellouz E, Chérif R, Turki S, Solidoro C, Aleya L (2016) Random forest model and TRIX used in combination to assess and diagnose the trophic status of Bizerte Lagoon, southern Mediterranean. Ecological Indicators 7: 293–301. https://doi.org/10.1016/j.ecolind.2016.07.010

Béjaoui B, Ottaviani E, Barelli E, Ziadi B, Dhib A, Lavoie M, Gianluca C, Turki S, Solidoro C, Aleya L (2018) Machine learning predictions of trophic status indicators and plankton dynamic in coastal lagoons. Ecological Indicators 95:765–774. https://doi.org/10.1016/j.ecolind.2018.08.041

Ben Charrada R (1992) Le lac de Tunis après les aménagements. Paramètres physicochimiques de l'eau et relation avec la croissance des macroalgues. Marine Life 1:29–44

Bowden GJ, Nixon JB, Dandy GC, Maier HR, Holmes M (2006) Forecasting chlorine residuals in a water distribution system using a general regression neural network. Mathematical and Computer Modelling 44:469–484. https://doi.org/10.1016/j.mcm.2006.01.006

Breiman L (2001) Random forests. Machine Learning 45:5–32. https://doi.org/10.1023/A:1010933404324

Charulatha G, Srinivasalu S, Uma Maheswari O et al (2017) Evaluation of ground water quality contaminants using linear regression and artificial neural network models. Arabian Journal of Geosciences 10:128. https://doi.org/10.1007/s12517-017-2867-6

Chen Q, Guan T, Yun L, Li R, Recknagel F (2015) Online forecasting chlorophyll a concentrations by an auto-regressive integrated moving average model: Feasibilities and potentials. Harmful Algae 43: 58–65. https://doi.org/10.1016/j.hal.2015.01.002

Cloern JE (2001) Our evolving conceptual model of the coastal eutrophication problem. Marine Ecology Progress Series 210:223–253. https://doi.org/10.3354/meps210223

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. Ecology 88: 2783–2792. https://doi.org/10.1890/07-0539.1

Daliakopoulos IN, Coulibaly P, Tsanis IK (2005) Groundwater level forecasting using artificial neural networks. Journal of Hydrology 309:229–240. https://doi.org/10.1016/j.jhydrol.2004.12.001

De Casabianca ML, Samson-Kechacha FL, Bone C (1991) Etude spatiotemporelle des sels nutritifs et des principales variables hydrobiologiques dans une lagune méditerranéenne: le lac Mellah (Algérie). Mesogée 51:15–23

de Jonge VN, Elliott M, Orive E (2002) Causes, historical development, effects and future challenges of a common environmental problem: eutrophication. Hydrobiologia (incorporating JAQU) 475/476:1–19. https://doi.org/10.1023/A:1020366418295

Derolez V, Soudant D, Malet N, Chiantella C, Richard M, Abadie E, Aliaume C, Bec B (2020) Two decades of oligotrophication: Evidence for a phytoplankton community shift in the coastal lagoon of Thau (Mediterranean Sea, France). Estuarine, Coastal and Shelf Science 241:106 810. https://doi.org/10.1016/j.ecss.2020.106810

Dhib A, Denis M, Barani A, Turki S, Aleya L (2016) Ultra- and microplankton assemblages as indicators of trophic status in a Mediterranean lagoon. Ecological Indicator 64:59–71

Fogelman S, Zhao H, Blumenstein M, Zhang S (2006) Estimation of oxygen demand levels using UV-Vis spectroscopy and artificial neural networks as an effective tool for real-time, wastewater treatment control. In: Proceedings of the 1st Australian Young Water Professionals Conference, Sydney, Australia

Frolov S, Rayan JP, Chavez FP (2012) Predicting eutrophic-depth-integrated chlorophyll-a from discrete-depth and satellite-observable chlorophyll-a off central California. Journal of Geophysical Research 117:C05042. https://doi.org/10.1029/2011JC007322

García-Ayllón S (2017) Diagnosis of complex coastal ecological systems: Environmental GIS analysis of a highly stressed Mediterranean lagoon through spatiotemporal indicators. Ecological Indicators 83: 451–462. https://doi.org/10.1016/j.ecolind.2017.08.015

Harbridge W, Pilkey OH, Whaling P, Swetland P (1976) Sedimentation in the lake of Tunis: a lagoon strongly influenced by man. Environmental Geology 1:215–225. https://doi.org/10.1007/bf02407508

Haykin S (1999) Neural networks: a comprehensive foundation, 2nd edn. Prentice Hall International Editions Series, London

He Z, Wen X, Liu H, Du J (2014) A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. Journal of Hydrology 509:379–386. https://doi.org/10.1016/j.jhydrol.2013.11.054

Huang J, Gao J, Zhang Y (2015) Combination of artificial neural network and clustering techniques for predicting phytoplankton biomass of lake Poyang, China. Limnology 16:179–191. https://doi.org/10.1007/s10201-015-0454-7

IOC SCOR, IAPSO (2010) The international thermodynamic equation of seawater – 2010: Calculation and use of thermodynamic properties. Intergovernmental Oceanographic Commission, Manuals and Guides No. 56, UNESCO (English), 196 pp

Jimeno-Sáez P, Senent-Aparicio J, Pérez-Sánchez J, Pulido-Velazquez DA (2018) Comparison of SWAT and ANN Models for daily runoff simulation in different climatic zones of Peninsular Spain. Water 10: 192

Jimeno-Sáez P, Senent-Aparicio JM, Cecilia J, Pérez-Sánchez J (2020) Using Machine-learning algorithms for eutrophication modeling: case study of Mar Menor lagoon (Spain). International Journal of Environmental Research and Public Health 17:1189. https://doi.org/10.3390/ijerph17041189

Kalaji HM, Sytar O, Brestic M, Samborska IA, Cetner MD, Carpentier C (2016) Risk assessment of urban lake water quality based on in situ cyanobacterial and total chlorophyll-a monitoring. Polish Journal of Environmental Studies 25(2):655–661. https://doi.org/10.15244/pjoes/60895

Keller S, Maier PM, Riese FM, Norra S, Holbach A, Börsig N, Wilhelms A, Moldaenke C, Zaake A, Hinz S (2018) Hyperspectral data and machine learning for estimating CDOM, chlorophyll a, diatoms, green algae and turbidity. International Journal of Environmental Research and Public Health 15:1881

Kohavi R, John GH (1997) Wrappers for feature subset selection. Artificial Intelligence 97(1–2):273–324

Kuo JT, Hsieh MH, Lung WS, She N (2007) Using artificial neural network for reservoir eutrophication prediction. Ecological Modelling 200:171–177

Lee JHW, Huang Y, Dickmen M, Jayawardena AW (2003) Neural network modelling of coastal algal blooms. Ecological Modelling 159: 179–201. https://doi.org/10.1016/S0304-3800(02)00281-8

Li X, Sha J, Wang ZL (2017) Chlorophyll-A Prediction of lakes with different water quality patterns in China based on hybrid neural networks. Water 9:524

Lin T, Horne BG, Tino P, Giles CL (1996) Learning long-term dependencies in NARX recurrent neural networks. IEEE Transactions on Neural Networks 7(6):1329–1338. https://doi.org/10.1109/72.548162

Lorenzen CJ (1967) Determination of chlorophyll and pheopigments by spectrophotometric equations. Limnology and Oceanography 12: 34–346

Lu F, Chen Z, Liu W, Shao H (2016) Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin. Ecological Engineering 95:422–429. https://doi.org/10.1016/j.ecoleng.2016.06.072

Maier HR, Jain A, Dandy GC, Sudheer K (2010) Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environmental Modelling and Software 25:891–909. https://doi.org/10.1016/j.envsoft.2010.02.003

Mc Quaid N, Zamyadi A, Prevost M, Bird DF, Dorner S (2011) Use of in vivo phycocyanin fluorescence to monitor potential microcystin-producing cyanobacterial biovolume in a drinking water source. Journal of Environmental Monitoring 13:455–463. https://doi.org/10.1039/c0em00163e

Mdaini Z, El Gafsi M, Tremblay J, Pharand P, Gagné JP (2019) Spatio-temporal variability of biomarker responses and lipid composition of Marphysasanguinea, Montagu (1813) in the anthropic impacted lagoon of Tunis. Marine Pollution Bulletin 144:275–286. https://doi.org/10.1016/j.marpolbul.2019.04.065

Menendez M, Martınez M, Comın FA (2001) A comparative study of the effect of pH and inorganic carbon resources on the photosynthesis of three floating macroalgae species of a Mediterranean coastal lagoon. Journal of Experimental Marine Biology and Ecology 256:123–136

Mitchell MW (2011) Bias of the random forest Out-of-Bag (OOB) error for certain input parameters. Open Journal of Statistics 1:205–211. https://doi.org/10.4236/ojs.2011.13024

Mjalli FS, Al-Asheh S, Alfadala HE (2006) Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. Journal of Environmental Management 83: 329–338. https://doi.org/10.1016/j.jenvman.2006.03.004

Mooney H, Larigauderie A, Cesario M, Elmquist T, Hoegh-Guldberg O, Lavorel S, Mace GM, Palmer M, Scholes R, Yahara T (2009) Biodiversity, climate change, and ecosystem services. Current Opinion in Environment Sustainability 1:46–54. https://doi.org/10.1016/J.COSUST.2009.07.006

Motoda H, Liu H (2002) Feature selection, extraction and construction. Towards the foundation of data mining workshop. In: Proceedings of the Sixth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'02), Taipei, Taiwan; pp 67–72

Mulia IE, Tay H, Roopsekhar K, Tkalich P (2013) Hybrid ANN–GA model for predicting turbidity and chlorophyll a concentrations. Journal of Hydro-Environmental Research 7:279–299. https://doi.org/10.1016/j.jher.2013.04.003

Nayak PC, Sudheer KP, Rangan DM, Ramasastri KS (2005) Short-term flood forecasting with a neuro fuzzy model. Water Resources Research 41:2517–2530. https://doi.org/10.1029/2004WR003562

Nazeer M, Wong MS, Nichol JE (2017) A new approach for the estimation of phytoplankton cell counts associated with algal blooms. Science of the Total Environment 590–591:125–183. https://doi.org/10.1016/j.scitotenv.2017.02.182

Newton A, Icely J, Cristina S et al (2014) An overview of ecological status, vulnerability and future perspectives of European large shallow, semi-enclosed coastal systems, lagoons and transitional waters. Estuarine, Coastal and Shelf Science 140:95–122. https://doi.org/10.1016/J.ECSS.2013.05.023

Newton A, Brito AC, Icely JD et al (2018) Assessing, quantifying and valuing the ecosystem services of coastal lagoons. Journal for Nature Conservation 44:50–56. https://doi.org/10.1016/j.jnc.2018.02.009

Nguyen VD, Tan RR, Brondial Y, Fuchino T (2007) Prediction of vapor-liquid equilibrium data for ternary systems using artificial neural networks. Fluid Phase Equilibria 254:188–197

Palani S, Liong SY, Tkalich P (2008) An ANN application for water quality forecasting. Marine Pollution Bulletin 56:1586–1597

Park Y, Cho KH, Park J, Cha SM, Kim JH (2015) Development of early-warning protocol for predicting chlorophyll-*a* concentration using machine learning models in freshwater and estuarine reservoirs, Korea. Science of the Total Environment 502:31–41. https://doi.org/10.1016/j.scitotenv.2014.09.005

Parsons TR, Maita Y, Lalli CM (1984) A manual of chemical and biological methods for seawater analysis. Pergamon Press, Oxford Oxfordshire

Phillips G, Pietiläinen O, Carvalho L et al (2008) Chlorophyll–nutrient relationships of different lake types using a large European dataset. Aquatic Ecology 42:213–226. https://doi.org/10.1007/s10452-008-9180-0

Rajaee T, Ebrahim H, Nourani F (2019) A review of the artificial intelligence methods in groundwater level modeling. Journal of Hydrology 572:336–351. https://doi.org/10.1016/j.jhydrol.2018.12.037

Rezgui A, Maiz N, Moussa M (2008) Fonctionnement hydrodynamique et écologique du Lac Nord de Tunis par modélisation numérique. Revue des Sciences de l'Eau 21:349–361. https://doi.org/10.7202/018781ar

Robledano F, Esteve MA, Martínez-Fernández J, Farinós P (2011) Determinants of wintering waterbird changes in a Mediterranean coastal lagoon affected by eutrophication. Ecological Indicators 11:395–406. https://doi.org/10.1016/j.ecolind.2010.06.010

Samarasinghe S (2007) Neural networks for applied sciences and engineering. Auerbach Publications, New York

Schramm W (1999) Factors influencing seaweed responses to eutrophication: some results from EU-project EUMAC. Journal of Applied Phycology 11:69–78

Shili A (1995) Contribution à l'étude de Ruppia dans le Lac Nord de Tunis. Mémoire de Diplôme des Etudes Approfondies en biologie marine et océanographie, Faculté des Sciences de Tunis, p 1-128

Souchu P, Bec B, Smith VH, Laugier T, Fiandrino A, Benau L, Orsoni V, Collos Y, Vaquer A (2010) Patterns in nutrient limitation and chlorophyll-*a* along an anthropogenic eutrophication gradient in French Mediterranean coastal lagoons. Canadian Journal of Fisheries and Aquatic Sciences 67:743–753. https://doi.org/10.1139/F10-018

Strickland JDH, Parsons TR (1972) A Practical Handbook of Seawater Analysis. Fisheries Research Board of Canada, Ottawa, Bulletin 167. 1968

Su J, Wang X, Zhao S, Chen B, Li C, Yang Z (2015) A structurally simplified hybrid model of genetic algorithm and support vector machine for prediction of Chlorophyll a in reservoirs. Water 7(4): 1610–1627. https://doi.org/10.3390/w7041610

Tian W, Liao Z, Zhang J (2017) An optimization of artificial neural network model for predicting chlorophyll dynamics. Ecological Modelling 364:42–52. https://doi.org/10.1016/j.ecolmodel.2017.09.013

Trabelsi-Bahri EL, Armi Z, Trabelsi-Annabi N, Shili A, Ben Maiz N (2013) Water quality variables as indicators in the restoration impact assessment of the north lagoon of Tunis, South Mediterranean. Journal of Sea Research 79:12–19. https://doi.org/10.1016/j.seares.2013.01.003

Van Berk H, Oostinga H (1992) North Lake of Tunis and its shores: restoration and development. Terra et Aqua 49:23–32

Viaroli P, Bartoli M, Giordani G, Naldi M (2008) Community shifts, alternative stable states, biogeochemical controls and feedbacks in eutrophic coastal lagoons: a brief overview. Aquat Conserv Mar Freshw Ecosyst 18:105–117. https://doi.org/10.1002/aqc

Wallace J, Champagne P, Hall G (2016) Time series relationships between chlorophyll-a, dissolved oxygen, and pH in three facultative wastewater stabilization ponds. Environmental Science: Water Research & Technology 2:1032–1040. https://doi.org/10.1039/C6EW00202A

Watzin MC, Miller EB, Shambaugh AD, Kreider MA (2006) Application of the WHO alert level framework to cyanobacterial monitoring of Lake Champlain, Vermont. Environmental Toxicology 21:278–288. https://doi.org/10.1002/tox.20181

Wei B, Sugiura N, Maekawa T (2001) Use of artificial neural network in the prediction of algal blooms. Water Research 35:2022–2028

Were K, Bui DT, Dick ØB, Singh BR (2015) A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecological Indicators 52:394–403. https://doi.org/10.1016/j.ecolind.2014.12.028

Xu G, Schwarz P, Yang H (2019) Determining China's CO2 emissions peak with a dynamic nonlinear artificial neural network approach and scenario analysis. Energy Policy 128:752–762. https://doi.org/10.1016/j.enpol.2019.01.058

Zaldívar JM, Cardoso AC, Viaroli P, Wit RD, Ibañez C, Reizopoulou S, Razinkovas A, Basset A, Holmer M, Murray N (2008) Eutrophication in transitional waters: an overview. Transitional Waters Monographs 1:1–78. https://doi.org/10.1285/i18252273v2n1p1

Zang C, Huang S, Wu M et al (2010) Comparison of relationships between ph, dissolved oxygen and chlorophyll a for aquaculture and non-aquaculture waters. Water, Air, and Soil Pollution 219:157–174. https://doi.org/10.1007/s11270-010-0695-3

Zhang WZ, Wang H, Chai F, Qiu G (2016) Physical drivers of chlorophyll variability in the open South China Sea. Journal of Geophysical Research Oceans 121:7123–7140. https://doi.org/10.1002/2016JC011983