Article

Data-Driven Models' Integration for Evaluating Coastal Eutrophication: A Case Study for Cyprus

Ekaterini Hadjisolomou [*] , Maria Rousou , Konstantinos Antoniadis , Lavrentios Vasiliades ,
Ioannis Kyriakides , Herodotos Herodotou , Michalis Michaelides [*]

Article

Data-Driven Models' Integration for Evaluating Coastal Eutrophication: A Case Study for Cyprus

Ekaterini Hadjisolomou [1, 2, *], Maria Rousou [3], Konstantinos Antoniadis [3], Lavrentios Vasiliades [3], Ioannis Kyriakides [2, 4], Herodotos Herodotou [1] and Michalis Michaelides [1,*]

[1] Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, 3036 Limassol, Cyprus; herodotos.herodotou@cut.ac.cy, michalis.michaelides@cut.ac.cy

[2] University of Nicosia Research Foundation, 46 Makedonitissas Avenue, CY-2417 P.O. Box 24005, CY-1700 Nicosia, Cyprus.

[3] Department of Fisheries and Marine Research, Ministry of Agriculture, Rural Development and the Environment, 2033 Nicosia, Cyprus; MRousou@dfmr.moa.gov.cy; kantoniadis@dfmr.moa.gov.cy, lvasiliades@dfmr.moa.gov.cy

[4] Cyprus Marine & Maritime Institute, CMMI House, Vasileos Pavlou Square, P.O.Box 40930, 6023 Larnaca, Cyprus; Ioannis.Kyriakides@cmmi.blue

* Correspondence: e.hadjisolomou@cut.ac.cy (EH), michalis.michaelides@cut.ac.cy (MM)

Abstract: Eutrophication is a major environmental issue with many negative consequences, such as hypoxia and harmful cyanotoxins production. Monitoring coastal eutrophication is crucial, especially for island countries like the Republic of Cyprus, which are economically dependent on the touristic sector. Additionally, the open-sea aquaculture industry in Cyprus has been exhibiting an increase in the last decades and environmental monitoring to identify possible signs of eutrophication is mandatory according to the legislation. Therefore, in this modelling study, two different types of Artificial Neural Networks (ANNs) are developed based on in-situ data collected from stations located in the coastal waters of Cyprus. Theses ANNs aim to model the eutrophication phenomenon based on two different data-driven modelling procedures. Firstly, the self-organizing map (SOM) ANN examines several water quality parameters (specifically water temperature, salinity, nitrogen species, ortho-phosphates, dissolved oxygen and electrical conductivity) interactions with the Chlorophyll-a parameter. The SOM model enables us to visualize the monitored parameters relationships and to comprehend complex biological mechanisms related to Chlorophyll-a production. A second feed-forward ANN model is also developed for predicting the Chlorophyll-a levels. Based on this ANN model, several scenarios associated to eutrophication-related water quality parameters can be extracted. The combination of these two ANNs models is considered a holistic modelling approximation for the identification of eutrophication scenarios, since it enables not only the prediction of the Chlorophyll-a parameter levels, but also the "capturing" of hidden biological mechanisms associated with algal production.

Keywords: data-driven models; eutrophication; coastal water quality; self-organizing map; feed-forward neural network

1. Introduction

The ocean is responsible for regulating the Earth's climate and provides humans with valuable resources, like energy and food [1]. Therefore, the sustainable usage of marine resources is an emerging concern. Marine environmental pollution related with human activities is a historically identified problem, but it only received the necessary attention during the recent years, when the anthropogenic pressure on the aquatic ecosystems and organisms reached a dangerous ecological threshold [2]. This intense anthropogenic pressure on the coastal environment is the result of the doubling of the human population and rapid industrial development [1]. Some of these anthropogenic activities impacting the coastal zones are related with inputs of excessive nutrients [3], heavy metals, and other pollutants originating from the land, like microplastics [4]. It is estimated that globally about 80% of marine pollution is land-derived [5]. The environmental degradation of

the coastal water results in harmful effects for the marine organisms and negatively impacts human wellbeing.

Eutrophication is considered as a key local stressor for coastal marine ecosystems. According to a study by Smith [6], which examined 92 coastal ecosystems, the coastal Chlorophyll-a (Chl-a) production was found to be related with two nutrients, Nitrogen (N) and Phosphorus (P). Furthermore, climate change and anthropogenic eutrophication have resulted in large variations in microalgae assemblage composition globally, like increase of Harmful Algal Blooms (HABs) or biomass increase [7]. The main impacts of these changes in algal composition include hypoxia/anoxia [8] with catastrophic side effects on aquatic organisms (e.g., declining fisheries stock). Additionally, eutrophication may trigger harmful bacterial production, which negatively affects the corals and other marine organisms [9]. Another side effect of eutrophication is related to nuisance blooms, which are having negative economic and societal impacts because of water aesthetic degradation, like water discoloration or foam [10].

The eutrophication of the coastal waters is addressed by several EU Directives including the Water Framework Directive (WFD) 2000/60/EC, the Marine Strategy Framework Directive (MSFD) 2008/56/EC and the Nitrates Directive 91/676/EEC; as well as Regional Sea Conventions, such as the Barcelona Convention for the protection of the Mediterranean Sea. The assessment of surface water bodies and the examination of their physicochemical status for the identification of anthropogenic pressure and possible changes, is a crucial issue for the associated environmental authorities [11]. Traditional methodologies include the analysis of data by using statistical methods, such as cluster analysis and ordination. Modelling studies have demonstrated that the application of suitable models, like the Artificial Neural Networks (ANNs), enables to examine the association/impact of several environmental parameters on water quality problems, like eutrophication [8].

Most of the water quality studies, that are using ANNs as modelling tools, are dealing with limnological or riverine applications, while ANN applications examining coastal systems are considerably fewer [12]. Specifically, for eutrophication related problems and their catastrophic effects, the use of ANN's predictions can prevent or minimize the effects of any possible HABs [13]. The ANN's well known abilities to model complex and non-linear relationships make them ideal for eutrophication modelling. As stated by Yussef et al. [14], in contrast to some other modelling techniques (e.g., statistical methods), ANNs are not affected by non-linearities or the complex interdependencies of interlayer connections.

Besides the predicting ability of ANNs to simulate algal productivity with a good accuracy, ANNs are good at examining the effect of the related water quality parameters based on their ability to associate them with the function of the algal biomass [15]. A category of ANNs broadly used in the eco-hydrological field are the unsupervised ANNs, known as the Kohonen Self-Organizing Maps (SOMs) [16]. These SOMs are mainly used for clustering [17] and for exploratory data analysis (data mining) of the investigated environmental data set [18]. As stated by Park et al. [19], the method of multivariate analysis is mainly applied for ecological patterning; however, ANNs are more suitable for this task because of the nonlinear and complex possible interactions between the various parameters in the modeled data set (which many times consists of many different species and sampling areas).

An example of a SOM model applied in eco-hydrological modelling is found in the study of Lu and Lo [20], where a trophic state classifier was constructed based on a SOM model, aiming to diagnose the water quality of the Fei-Tsui Reservoir (Taiwan) during the monitoring period, between 1987-1995, and compared the simulated SOM results with those of the Carlson Index. In the study of Li et al. [21], the SOM model is applied to evaluate the spatiotemporal variations of groundwater quality data in Northeast Beijing, where based on SOM's clustering, different pollution sources (like industrial and agricultural activities, domestic-sewage-discharge sources) were identified for the related sampling sites.

Another category of ANNs are the multilayer feed-forward neural networks, which are supervised learning-based ANNs. This type of ANNs is capable of predicting the Chl-a levels based on several water quality parameters associated to algal production [22]. These environmental

parameters, which are used as the ANN's inputs, may differ among modelling studies of coastal eutrophication. For example, in their study Salami et al. [23] created a feed-forward back propagation ANN for predicting coastal Chl-a values near Grant Line Canal, California, USA, based on the Electric Conductivity (EC), water temperature (WT) and pH parameters. Even though only three monitoring parameters were used as the model's inputs, the created ANN managed to predict the Chl-a levels with a satisfactory accuracy rate (75.9%). While, in another study by Melesse et al. [24], the coastal Chl-a levels at Florida Bay were modelled with the use of a back propagation ANN. Specifically, the authors examined various combinations of seven candidate input parameters (total phosphate, nitrite, ammonium, turbidity, WT, DO and antecedent Chl-a), and it was concluded that the ANN performed better when using all the above input parameters.

Data-driven models based on ANNs algorithms can be used to support the development of eutrophication control management tools, since ANNs are able to reveal the underlying mechanisms associated with algal productivity and the related environmental parameters [25]. Additionally, as stated by Georgescu et al. [26], the application of AI methods for water quality modelling saves time and resources in lab analysis, while the generated statistical data are important for the relevant authorities/managers. The above practical reasons and the fact that no other similar modelling study based on SOM models currently exists for the Cyprus coastal waters, motivated the current modelling study. The proposed SOM model enables us to comprehend to a greater extent possible hidden mechanisms and interactions between the Chl-a parameter and the rest of the eutrophication-related parameters. In the proposed modelling study, we are focusing on the role/interactions of water quality parameters associated to eutrophication and the impact of anthropogenic activity for several coastal stations near the Republic of Cyprus. The land use of the different regions near the sea catchment area is reflected based on the nutrients' concentration in the nearby coastal stations, while it is well documented that excessive amounts of nutrients in the surface water may lead to eutrophication [27]. In our case, it was found that the water quality status of Cyprus is good and practically not impacted by anthropogenic activities. Nevertheless, the created data-driven models can act as advisory/management tools for assessing the expected pressure from planned anthropogenic activities or even environmental changes, like global warming.

2. Materials and Methods

2.1. Study Area and Data Aquisition

The Republic of Cyprus is an islandic country, located in the Eastern Mediterranean area, specifically in the Levantine Basin. According to Tselepides et al. [28], the Levantine Sea is considered as one of the most oligotrophic seas worldwide, therefore Cyprus marine waters are having very low primary production, as a result of the limited nutrient availability [29]. In addition to its ultra-oligotrophism, Levant's Sea is characterized by high temperatures ranging yearly from 16 °C in the winter up to 26 °C in the summer period [22]. Moreover, the evaporation and salinity are high (yearly average salinity of Eastern Mediterranean exceeds 37.5 psu, while average salinity of coastal waters of Cyprus is 39.1 psu); and the inflow of fresh water is very limited because of extensive damming and the absence of large rivers [30].

The Department of Fisheries and Marine Research (DFMR) of the Ministry of Agriculture, Rural Development and Environment, of the Republic of Cyprus, as part of the implementation of the WFD, MSFD, Nitrates Directive and the Barcelona Convention, carries out a monitoring programme to collect, among others, water-column data. A total of 49 coastal stations are monitored along the Cyprus coastline, some of which are located near anthropogenic activities such as aquaculture facilities and industrial units (Figure 1). Water column samples are collected and analysed, and the data are included in DFMR's "Thetis" database.

Figure 1. Satellite map of the Republic of Cyprus, which is located in the Eastern Mediterranean region (green colored markers are used for indicating the sampling sites).

For the scope of the current study, a total of 1552 water column data were provided for the development of the ANNs. The data samples were collected sporadically (having no regular time intervals) from the 49 coastal stations between the years 2000-2020. Specifically, the following environmental parameters were investigated: (i) nitrogen species (NH_4^+, NO_2^-, NO_3^-); (ii) ortho-phosphates (PO_4^{3-}); (iii) salinity; (iv) dissolved oxygen (DO); (v) pH; (vi) electrical conductivity (EC); (vii) water temperature (WT); and (viii) Chl-a. More details regarding the stations and the data sampling process are found in Antoniadis et al. [30].

2.2. Multilayer feed-forward ANNs

ANNs are inspired by the function of the biological brain, where a neuron receives a signal, processes it, and then transmits an output signal to other interconnected neurons or nodes [31]. Multilayer feed-forward ANNs are supervised machine learning models, and are capable of processing non-linear phenomena [14, 24]. According to Kohonen and Kaski [32], the multilayer feed-forward ANN is an efficient non-linear "general-purpose" function approximator. The multilayer perceptron (MLP) architecture is a layered feedforward ANN, in which the neurons are arranged in fully-connected successive layers: the input layer, the hidden layer(s) and the output layer [33]. A synaptic weight is associated with each neuron, which is connected with all the neurons of the next layer.

The output value of the j-th neuron (o_j) is calculated by the following equations [31]:

$$o_j = f(u_j), \tag{1}$$

$$u_j = \sum w_{ij}\, x_i + z_j\,, \tag{2}$$

where f is the transfer function, x_i is the input from the i-th neuron belonging to the immediate previous layer, w_{ij} is the synaptic weight that connects x_i with the j-th neuron and z_j a bias term. The output of each neuron is computed and propagated through the next layer until the last layer, and this procedure is repeated until the calculated output starts to converge to a desired target-output [8], while the goal of the training process is finding a set of synaptic weights that minimizes the loss function.

Data standardization/normalization is an important step before ANN model development. The data normalization eliminates dimensional differences among the different variables [34], since the input variables may have values of different orders of magnitude [33]. The ANN's performance is measured based on several statistical performance indices (metrics) for the test set data, like the Root

Mean Square Error (*RMSE*), the Mean Absolute Error (*MAE*), and the Pearson's Correlation Coefficient (*R*).

The MLP ANN's sensitivity analysis can be examined based on several methodologies. The Perturb algorithm, which demonstrates how the trained network reacts to a small change/perturbation of each input, is one of the most widely applied sensitivity analysis algorithms. The Perturb sensitivity is calculated by the following [35]:

$$\text{Sensitivity (\%)} = \frac{1}{N_P} \sum_{i=1}^{N_p} \left(\frac{change\ in\ output\ (\%)}{change\ in\ input\ (\%)} \right)_i \times 100 \quad , \tag{3}$$

where the parameter N_p represents the number of patterns (samples number).

2.3. Self-Organizing Map (SOM)

SOM is an unsupervised learning type of ANNs, meaning that no human intervention (supervision) is required during its learning process [36]. The term self-organizing is given because of the SOM's ability to learn and organize information without being given the associated output values for the corresponding input data, while the desired output is not known a priori [37]. The SOM can project high-dimensional data into a low-dimension space, most commonly two-dimensional [36].

The SOM consists of an input layer and an output layer, which are connected with computational weights [38, 39]. The SOM algorithm's procedure [21, 40] is summarized by the following steps:
1. Weight vector initialization with random values.
2. Use of a distance measure -usually the Euclidean distance- to find the best-matching unit (BMU).
3. Move closer to the input vector by updating the weight vector of the BMU and the neighboring neurons.

The Euclidean distance (D_i), calculates the distance measure between the input vector and the i-th weight vector [39] and is given by the following:

$$D_i = \sqrt{\sum_{j=1}^{L} (p_{ij} - w_{ij})^2} \quad ; \ i = 1,2, \dots S \quad , \tag{4}$$

where S is the number of output neurons, L is the dimension of the input vectors, p_{ij} represents the j-th element of the input vector, and w_{ij} symbolizes the j-th element of the i-th weight vector. The term BMU is defined as the neuron with the weight vector closest to the input variable x, meaning the weight vector that has the shortest distance to the input vector [41] and is calculated by the equation:

$$|x - m_c| = min(|x - m_i|) \quad , \tag{5}$$

where $|\cdot|$ symbolizes the distance measure, x the input vector, m the weight vector, and c the subscription of the weight vector for the winning neuron.

A very common rule of thumb for finding the SOM's optimum map size [38] is the one proposed by Vesanto and Alhoniemi [42] using the following formula:

$$M \approx 5\sqrt{n} \quad , \tag{6}$$

where n is the data sample number and M is the number of SOM's neurons.

SOM's output space is visualized by using a unified distance matrix (U-Matrix). The U-Matrix calculates distances between neighboring map units (neurons) [40]. The SOM's Component Planes (CPs) are an important visual feature of the SOM map and are defined as the values of a single vector component in all map units [43].

The SOM can automatically group (cluster) and typify data according to different properties of the data set variables [44]. The data can be clustered either manually as determined by the U-matrix; or can be automated by a clustering algorithm implemented in the SOM, by applying hierarchical (e.g., a dendrogram) and partitive (e.g., k-means algorithm) approaches [42].

3. Results

3.1. SOM's Results

For the needs of this modelling study, a SOM with **20 × 10** neurons was created. The SOM's topology, which is associated with the number of SOM's neurons, was calculated after applying Equation 6. The data simulations were based on the SOM Toolbox for MatLab [45]. The created SOM's U-matrix and the CPs are visualized as in Figure 2.

The CPs revealed a strong positive relationship between the EC, pH and salinity parameters, since they have very similar CPs. Not surprisingly, the CPs for the NH_4^+, NO_2^-, PO_4^{3-} parameters are associated with the Chl-a parameter, with a strong positive relationship; the highest values of NH_4^+, NO_2^-, PO_4^{3-} parameters correspond to increased values of Chl-a. This observation derived from the SOM's CPs agrees with the eutrophication production mechanism, since eutrophication is associated with an excessive increase in nutrients [46]. Regarding the rest of the parameters, no clear conclusions can be derived by the CPs observation. So, in an additional step, to reveal hidden relationships/mechanisms between the parameters the SOM's clusters statistical properties are investigated.

The U-matrix is often used to explore the parameters interactions between the SOM's formed groups (clusters) [47]. The U-matrix visualization (Figure 2) is indicating a tendency for the data to be grouped into three clusters; however, this is not clearly observed here (see Figure 2). Therefore, the *k*-means clustering algorithm was implemented in the SOM to calculate the optimal number of SOM's clusters. The Davies–Bouldin index is used to compute a minimum value for the SOM's optimal number of clusters [42]. In our case, the optimal number of clusters was three, as it is shown in Figure 3. The clustering of the SOM based on the *k*-means algorithm and the percentage of SOM's hits for each cluster are illustrated in Figure 4.
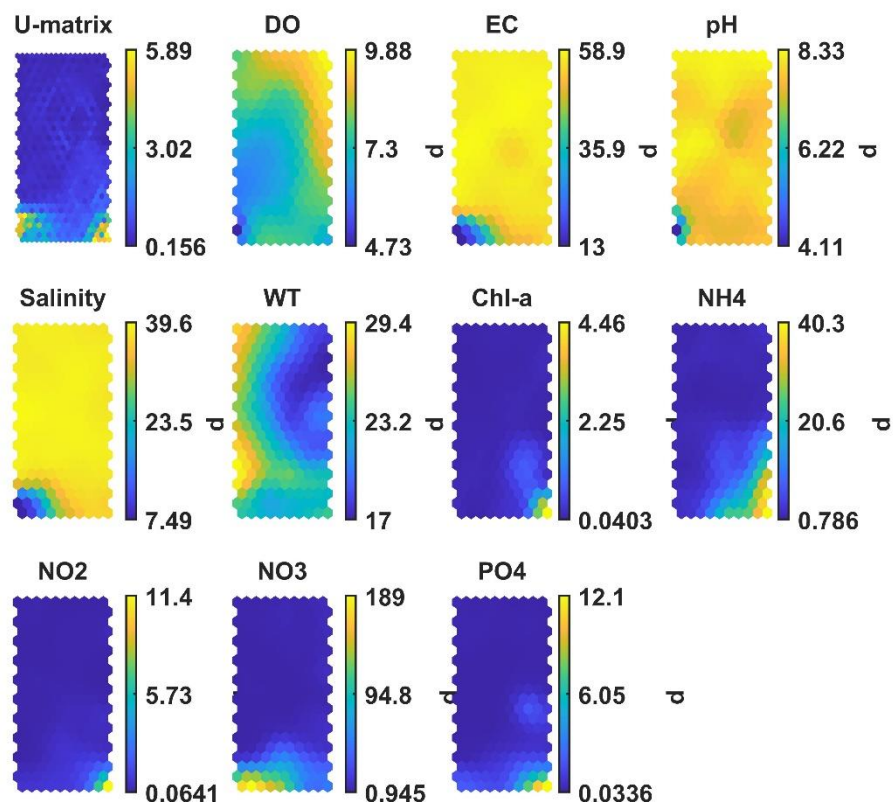
Figure 2. Visualization of the SOM's component planes (CPs) for each environmental parameter, where the colorbars indicate the mapping of the data values.
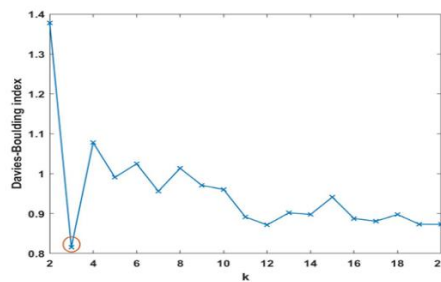

Figure 3. Procedure for calculating the optimal number of clusters based on the minimization for the Davies–Bouldin index when the SOM is implemented by the *k*-means algorithm. The minimum number of the Davies-Boulding index (*k*=3) is indicated in a red circle.

As indicated by the CPs and SOM's clustering (Figures 2 and 4), the Cluster 2 (C2) has the worst water quality. The nutrients (except the NO_3^- parameter) and the Chl-a have the highest concentrations for data belonging to the Cluster 2. Regarding Cluster 1 (C1), the parameters NO_3^-, EC and salinity, are having significant influence on it, while the pH seems to be associated but in a lesser extend. Finally, Cluster 3 (C3) has the best water quality since it is characterized by low concentration of Chl-a and nutrients; however, no clear associations can be made regarding the water quality parameters interactions. Nevertheless, it must be noted that based on the SOM's clustering, 95% of the sample hits are grouped into C3 (n3=1475), 3.6% of the sample hits are grouped into C1 (n1=56) and 1.4% of the sample hits are grouped into C2 (n2=21).
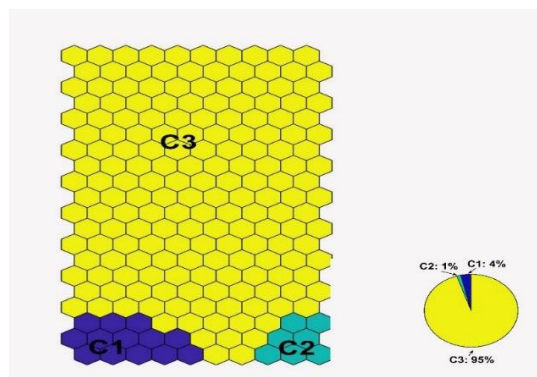

Figure 4. Clustering of the SOM based on the k-means algorithm (where Cluster 1: C1 is symbolized with blue, Cluster 2: C2 is symbolized with green, Cluster 3: C3 is symbolized with yellow). The three formed clusters of the SOM are observed. The pie chart is presenting the percentage of SOM's hits for each cluster.

The boxplots' utilization is synopsizing the basic statistical properties (e.g., median value, outliers) of the data belonging to each of the three formed SOM's groups (C1, C2, C3) and for each SOM's input environmental parameter (Figure 5). Based on the boxplots, the comparison between the data belonging to each group/cluster is enabled by examining their statistical properties. The NH_4^+, NO_3^-, EC, salinity, Chl-a and PO_4^{3-} parameters are having clear differences between the three SOM' s groups. The rest of the parameters (DO, pH, WT, NO_2^-) seem to have more similar statistical properties, however the smaller magnitude of their value range should be taken into consideration. From the DO, pH, WT, NO_2^- parameters, the NO_2^- is the only one without overlapped notches of its boxplots, indicating a differentiation between the three SOM's groups.
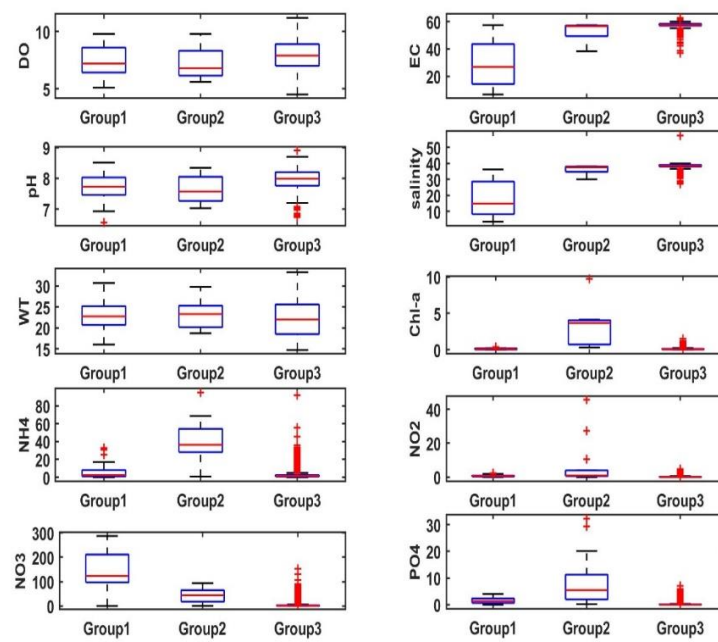
Figure 5. Boxplot graphical representation of the SOM's groups/clusters (Group1, Group2, Group3) found by using the *k*-means algorithm for each input environmental parameter (where the red horizontal line symbolizes the group's median value; the box gives the 25–75% percentile; the whiskers give the valid range; red marks are associated with extreme values/ outliers).

3.1. Feed-Forward ANN's Results

For prediction/regression purposes regarding the Chl-a values, a feed-forward ANN was created. Initially, the variables before being presented to the ANN, were transformed based on the min–max normalization, which projected the data to the range [0, 1], ensuring that feature variables have similar scales [48]. The ANN's optimal topology was found to be *9-6-1*, after following the trial-and-error procedure. The ANN was trained with the Levenberg-Marquardt training algorithm, since it is considered most effective for medium-sized networks [49]. The EC, pH, salinity, NO_3^-, NH_4^+, NO_2^-, PO_4^{3-}, DO and WT parameters served as the ANN's inputs.

The data set (*n*= 1552) was divided into training set and test set by 80% and 20% respectively, while the ANN was evaluated on its test set data. The achieved performance metrics are *MAE*= 0.0124 and *R*=0.97, while the graphical illustration between the real and the predicted data of the test set is given in Figure 6. It is observed that the plots of the real and predicted Chl-a values are very similar, verifying the ANN's good performance. The Chl-a limits for different water quality statuses (high, good, moderate) regarding Cyprus are given in the embedded table in Figure 6. For Chl-a concentrations below 0.4 mg/l (moderate and good water quality status) the real and the predicted data have almost a perfect much. Regarding the moderate-status Chl-a values, the ANN managed to produce also good outputs, as can be observed from Figure 6, except for one point corresponding to the highest measured value of the Chl-a parameter.
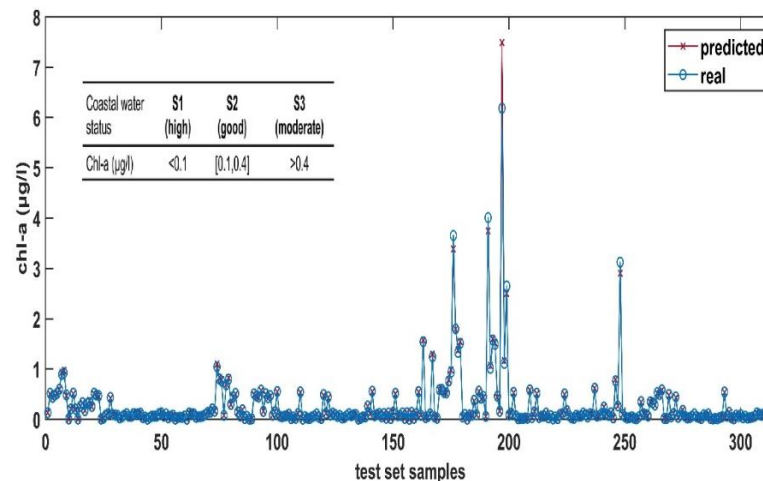
9

Figure 6. ANN's predicted values for Chlorophyll-a (Chl-a) levels regarding the test set data vs. the real Chl-a measurements, where the blue line is associated with the real data and the red line is associated with the predicted data. The embedded table is describing the Cyprus coastal water status for different Chl-a concentrations (where S1: high, S2: good, and S3: moderate).

Sensitivity analysis was carried to evaluate the input parameters' impact on the modelled Chl-a parameter. For that reason, the input parameters were increased (perturbed) based on the perturbation sensitivity analysis algorithm by +10%, and similarly decreased by -10%. The results of the sensitivity analysis are graphically illustrated in Figure 7. In the case of +10% increase of the input parameters, it was calculated that the nutrients (PO_4^{3-}, NO_2^-, NH_4^+, NO_3^-) are having a positive relationship with the Chl-a parameter production mechanism. Also, the pH, EC and salinity are positively related with the Chl-a parameter, while the WT and DO parameters are having a negative relationship with the algal production. In the case of -10% decrease of the input parameters, it was calculated that the Chl-a levels are decreased for the PO_4^{3-}, NO_3^-, NH_4^+, salinity, EC, pH parameters, while the Chl-a levels are increased for the NO_2^-, WT, DO parameters. The salinity (when negatively perturbated) and the WT (when positively perturbated) parameters are the most influential on the Chl-a production.
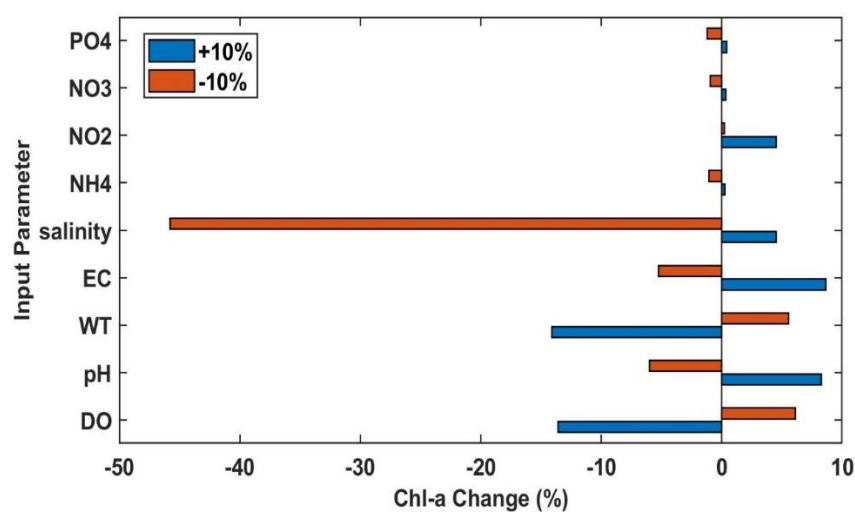

Figure 7. ANN's sensitivity analysis results for each of the input parameters. The fluctuation of each input parameter by an increase of + 10% and the associated Chl-a change is symbolized with blue colour, while the fluctuation of each input parameter by a decrease of -10% and the associated Chl-a change is symbolized with red colour.

4. Discussion

Eutrophication is an environmental issue closely related to anthropogenic activities. A vast number of monitoring studies are pointing out the negative impact of these anthropogenic activities, which are responsible for nitrogen and phosphorus release into the water environment [3]. For example, in a water quality study of Papastergiadou et al. [50], long term hydrological data and a GIS system were used for extracting land cover/use changes, while the authors concluded that anthropogenic activities are seriously affecting water quality and are promoting eutrophication. Therefore, understanding eutrophication related water quality parameters interactions and how each of these environmental parameters affects algal production is the keystone to developing sustainable management practices and restoration measures in eutrophication-affected areas [51]. As stated by Peppa et al. [52], many environmental studies are dealing with the prediction and analysis of eutrophication phenomena and the related parameters interactions in order to identify the possible causes and to provide possible solutions for the problem.

The maintenance/achievement of good water quality status is a goal for all the European Union member countries, including the Republic of Cyprus. For that reason, as indicated before, several Directives must be implemented, like the Water Framework Directive (WFD), the Nitrates Directive and the Marine Strategy Framework Directive (MSFD). In this modelling study, data-driven modelling techniques are applied aiming to model the coastal water quality in several areas of Cyprus. Based on the modelling outputs, the Chl-a levels can be predicted, but also the eutrophication-related water parameters and their contribution to Chl-a production can be evaluated. Specifically, two different types of ANNs were utilized for the needs of this modelling study. Firstly, an unsupervised type of ANN was created, specifically the SOM model. Secondly, another type of ANN, the feed-forward ANN, which is a supervised type was also developed. By combining the output information provided by these two types of ANNs, an in-depth investigation of the eutrophication phenomenon was enabled. In their study, Youssef et al. [14] state that ANNs have better performance in comparison to other machine learning and statistical methods, however, their black box nature makes ANNs' outcomes difficult to interpret and explain in practice. In our case, the parallel utilization of the SOM's results and the feed-forward ANN's sensitivity analysis outcomes, enabled us to unravel hidden complex mechanisms between the Chl-a parameter and the rest of the water quality parameters. As stated by Chon [53], the integration of the SOM and MLP models promotes the advanced information extraction from water quality data sets.

According to Kalteh et al. [44], the SOM can be characterized as a modelling technique suitable to investigate many types of aquatic systems and water resources processes. Also, the previous authors state that the SOM has the ability to group data into homogeneous areas, which is useful when needed to transfer information from gauged to ungauged sites (like geographically remote areas). Another useful property of the SOM comes from its clustering capabilities and the heat maps associated with the CPs, which allow visual qualification of relationships between input parameters properties [54]. The utilization of SOM is very beneficial when the correlation between the input parameters is non-linear and/or when dealing with noisy data; under those conditions the CPs can reveal relationships between the data that wouldn't be otherwise detected [55]. In their study, Astel et al. [56] are emphasizing the SOM's classification and visualization ability for large water quality data sets, while the authors are also mentioning the SOM's ability for simultaneous observation of the water quality parameters and their spatial and temporal changes based on the CPs visualization. Meanwhile, Varbiro et al. [57] argue the SOM's superiority against traditional multivariate statistical methods (like cluster analysis and ordination) because of the SOM's ability to simplify data's complex statistical relationships between the variables into simple geometric relationships represented into a 2-dimensonial space.

Regarding the second ANN implemented in this modelling study, the feed-forward ANN was chosen, which is a supervised type of ANN. The feed-forward ANNs are able to model non-linear complex environmental systems [58]. Additionally, as stated by Bushra et al. [59], the backpropagation ANNs have the merit of being simple to adapt and no tuning or learning is required for their parameter and function features. Furthermore, as it is stated by Brown et al. [60], ANN

models are giving more reliable outputs in comparison to other machine learning methods (e.g., decision trees or linear regression) when the data measurements number is relatively small, like in our case. Generally, feed-forward ANNs are considered reliable predictors of the Chl-a parameter and are widely used for Chl-a levels prediction [8].

As it was mentioned above, the created feed-forward ANN model managed to model the Chl-a levels with high accuracy, while the error between the real and the predicted data is very small, which is easily observed from the graphical illustrations. For the relatively low-medium values of the Chl-a parameter, the ANN produced almost identical outputs between the real and the simulated data. For the elevated Chl-a values, the ANN's error tends to increase, however, the calculated ANN's values are still near the measured ones, suggesting the ANN's good generalization ability. Despite these small errors, the ANN managed to correctly categorize the trophic status for all data samples.

The perturb sensitivity analysis algorithm was applied and each parameter was fluctuated by ±10% respectively. Based on the sensitivity analysis results, the basic trends between each input parameter and the Chl-a parameter were observed. When the parameters were increased/fluctuated by +10%, it was concluded that the salinity parameter was the most influential, since the Chl-a levels experienced the biggest modification.

The WT and DO parameters were also found to be significantly influential concerning Chl-a production. For the WT parameter, it was calculated that the WT and the Chl-a are negatively associated. This finding agrees with the fact that the coastal Chl-a levels near Cyprus reach their maximum values during the winter to early spring months, where cooler temperatures prevail, following the winter mixing and increase of phytoplankton production [61]. This is also recorded by Fyttis et al. [62] during a monitoring study of 12 consecutive months (January—December 2016), where the maximum coastal Chl-a levels of Cyprus were recorded during the winter. Regarding the salinity parameter, the Chl-a levels are significantly decreased when the salinity is decreased and vice versa. The upwelling phenomenon is again suggested to be related with this, since during the upwelling phenomenon nutrient rich water is emerging to the surface [63].

Regarding the strong negative relationship between the DO and the Chl-a parameters, the upwelling might also explain this. In a study of Georgiou et al. [64] in the Amvrakikos Gulf (Greece), low oxygen levels are reported during winter months. The above authors are attributing the anoxia to the strong winds and the resulting upwelling phenomenon. Therefore, the wintertime upwelling (and wind speed) is a factor that should be considered for future water quality modelling studies in Cyprus. As mentioned by Suursaar [65], the wintertime upwelling is a phenomenon, which has been ignored and not given the necessary attention, in contrast to the summer upwelling.

The rest parameters seem to be less contributing to the algal production. The feed-forward ANN captured the relationship between the phosphorus and the Chl-a parameters, where the increased values of phosphorus are positively related with increased algal production and vice versa. As stated by Ren et al. [66], the high levels dissolved inorganic phosphorus, mainly in the form of phosphate in the water column, could enhance the algal production. Regarding the DIN species, a less important relationship with the Chl-a parameters is found, which is having similar behaviour with the phosphorus parameter. A major source of DIN into coastal waters is associated with atmospheric deposition. Two main sources of DIN are related with anthropogenic activities, specifically riverine inputs and atmospheric deposition. In the study of Paerl et al. [67] contacted along the U.S coast and the eastern Gulf of Mexico, it was estimated that the nitrogen atmospheric deposition was responsible for a range of values between 10% and 40% of the new nitrogen loadings. While, according to Droge and Kroeze [68], riverine inputs are considered the main source of nitrogen for coastal waters and as estimated by the authors based on modelling studies, the DIN export will keep increasing in comparison to the pre-industrial era.

The development of data-driven models is a precious scientific tool for coastal water quality modelling. In our case, the integration of a supervised and an unsupervised ANN was proven to be a successful combination, not only for predicting the Chl-a levels, but also for examining the interactions of the eutrophication related parameters. The sensitivity analysis results provided the tendency regarding the parameters fluctuations (increased/decreased) and the analogous

negative/positive impact on the algal production mechanism. At the same time, the SOM model enabled an in-depth examination of the water quality parameters dataset. Specifically, in the SOM case, the resulted clustering of the data revealed biological mechanisms regarding algal production between the groups, which are not apparent if the data set is examined as a whole. The SOM's results revealed hidden relationships between the water quality parameters, which couldn't be easily identified or understood based on other modelling procedures. The visualization ability and the grouping of the SOM enabled us to make associations for specific value ranges for the parameters. As highlighted by Duarte et al. [69], complex patterns and interactions between the input parameters can be interpreted and understood based on the CPs visualization.

Regarding the nutrients based on the SOM's results, the Chl-a parameter and the NH_4^+, NO_2^- and PO_4^{3-} parameters are having similar box plots and CPs, suggesting a strong relationship between the Chl-a parameter and the impact of the NH_4^+, NO_2^- and PO_4^{3-}. While regarding the NO_3^- parameter, its moderate concentrations, based on the SOM clustering, are associated with the highest Chl-a values. The SOM's clustering of the data set (see Figure 4) verified the good water quality status of Cypriot coastal water, since only 1.4% of the total samples were characterized as problematic by the SOM results. In their study, Varbiro et al. [70] are applying the SOM to evaluate the Danube's tributaries based on diatom association, where the authors concluded that the upper stretch (German-Austrian region) is having better water quality than the lower stretch (Slovakian-Hungarian region). This SOM's visualization ability, which enables clustering the data samples and at the same time comparing the parameters' concentration levels for each cluster based on the analogous CP region, enables the extraction of conclusions about the different data sampling stations and their association with different water quality status. In our case, this finding can provide important information to the local authorities for the eutrophication, since it is indicated that not all the nutrients must have the same treatment regarding eutrophication control, as analysed above based on the box plots results (Figure 5).

Despite the limiting factor of the relatively small data set used in this modelling study, the created ANNs not only managed to perform well, but also managed to capture biological mechanisms/relationships and special characteristics describing the coastal algal production in Cyprus, like the winter upwelling phenomenon mentioned above. The issues related to ANN's poor performance because of limited data for learning are discussed in the study of Scardi [71], where the author suggests that new approaches could enhance ANNs to overcome this problem, like the adaptation of co-predictors. It must be noted that in a previous modelling study, Hadjisolomou et al. [72] developed a feed-forward ANN, which managed to predict the surface coastal Chl-a levels near Cyprus with a good accuracy (*R*=0.87 for the test). However, the data set was much smaller (*n*= 681) in comparison with the data set of this modelling study (*n*= 1552). For that reason, the previous model was validated by applying the *k*-fold method, while the used topology of that ANN was different (9-8-1). As explained by Hadjisolomou et al. [25], the application of *k*-fold method might include some concerns, related to the small data set for testing and therefore the evaluation might become less reliable and robust. Another, important detail related with the nature of the data set, which was analysed in Hadjisolomou et al. [72], was that only one sporadic measurement with elevated Chl-a value was recorded. As expected, the current ANN created for the needs of this modelling study has better performance (*R*=0.97 for the test set), while differences related with the parameter's sensitivity analysis results are also observed. These differences are mainly attributed to the fact that the current ANN is created based on a data set which contains a significant number of high/elevated Chl-a parameter measurements. Therefore, the current ANN, besides the fact that it performs better, it can generalize better in situations where the algal production is increased. So, the creation of updated ANNs models based on denser measurements and a bigger database would provide even more valuable information and could allow us to better understand the algal production mechanisms.

Often enough, in situ data collected from monitoring campaigns are usually few and demand high economical cost. As it is noted by Xu et al. [73], the development of ML models based on a small data set, may result into modelling complications (e.g., the model tends to overfit or underfit the data). So, water quality data originated from other data sources (besides monitoring cruises) is an

alternative. For example, in their study, Shan et al. [74] are demonstrating the need for more advanced modelling approaches (like the Long-Short-Term-Memory model) using online monitored data for simulating the complicate patterns of algal growth, since as the authors state no persuasive conclusions can be extracted only from statistical analysis of the data derived from monitoring campaigns. The laborious and costly nature of monitoring campaigns is also emphasized in the study of Silva et al. [75], where a modified Water Quality Index is developed based on sensor-derived data. Taking into consideration the above considerations and the generally accepted opinion that the ML models are data-hungry [76], the creation of data-driven models based on data collected from monitoring cruises fused with data from several other available sources (like satellite data, buoy data, historical databases) is recommended. Data-driven models created based on such a hybrid database would enable the development of even more specialized models, able to forecast ahead the Chl-a levels spatially as well as temporally. Additionally, management scenarios for economic activities (e.g., aquaculture) can be studied. For example, in their study, Giangrande et al. [77] are discussing the integration of different trophic levels with mariculture in the Mediterranean Sea and the role of restoration ecology. In their study, Eze at al. [78] developed a deep neural network based on real sensor data for the prediction of water quality parameters for a South African Aquaculture Farm.

It is generally accepted that water quality monitoring is a time-consuming and expensive procedure [51]. Utilizing ANNs for the modelling of water quality parameters is considered the best practice, as compared to other experimental or monitoring methods, which are usually costly or take too long for the data gathering [79]. In the study by Ahmed et al [80], the various methods available for estimating the DO concentration are analysed and the authors state that most of these analytical methods are either time-consuming and/or expensive, while the conventional data processing techniques are inappropriate since they are affected by non-linearities; therefore, the above authors are proposing ML data-driven models for water quality modelling prediction purposes. The ML-data-driven models used for prediction are able to overcome modelling limitations related to complex and non-linear data sets, therefore are widely used in water quality modelling [31, 81]. To summarize, based on the results of our study, it is obvious that the utilization of ANNs for the identification of areas sensitive to eutrophication is of great important to the local authorities and policy makers allowing them to apply measures when needed for the protection of the marine environment, especially in areas where limited scientific knowledge might exist or because data availability/acquisition is difficult.

5. Conclusions

Eutrophication is well known to have a detrimental effect to the marine water quality. In our case, two data-driven models were developed for evaluating the impact of eutrophication-related water quality parameters. Despite the limiting factor of the small data set, the created ANNs not only managed to perform well, but also managed to capture biological mechanisms/relationships and special characteristics describing the coastal algal production in Cyprus. For example, the winter upwelling seems to have an important role to the eutrophication phenomenon and the cooler WT measurements are associated with higher Chl-a levels. Therefore, it is recommended that any implementation measures regarding eutrophication control must be assessed based on modelling scenarios, since data-driven models are proven to be reliable prediction tools. The created ANNs not only can predict Chl-a levels, but also can extract thresholds for the associated water quality parameters, like the phosphate and the nitrogen species. Therefore, the created ANNs for the needs of this modelling study can act as the basis for advisory tools, contributing not only for the Cyprus marine environmental protection, but the local economy as well, which is related to financial activities like coastal tourism, shipping, and aquaculture.

14

References

1. Visbeck, M. Ocean science research is key for a sustainable future. *Nat. Commun.* **2018**, *9*, 690. https://doi.org/10.1038/s41467-018-03158-3

2. Islam, S.; Tanaka, M. Impacts of pollution on coastal and marine ecosystems including coastal and marine fisheries and approach for management: a review and synthesis. *Mar. Pollut. Bull.* **2004**, *48*; 624-649. doi: 10.1016/j.marpolbul.2003.12.004.

3. Akiner, M.E. The problem of environmental pollution in the Mediterranean Sea along the coast of Turkey. *J. Eng. Stud. Res.* **2020**, *26*, 7-14. https://doi.org/10.29081/jesr.v26i2.161

4. He, Q.; Silliman, B.R. Climate Change, Human Impacts, and Coastal Ecosystems in the Anthropocene. *Curr Biol.* **2019**, *29*,1021-1035. doi: 10.1016/j.cub.2019.08.042.

5. Alam, M.W.; Xiangmin, X.; Ahamed, R. Protecting the marine and coastal water from land-based sources of pollution in the northern Bay of Bengal: a legal analysis for implementing a national comprehensive act. *Environ. Chall.* **2021**, *4*, 100154

6. Smith, V.H. Responses of estuarine and coastal marine phytoplankton to nitrogen and phosphorus enrichment. *Limnol. Oceanogr.* **2006**, *51*, 377–384.

7. Jiang, Z.B.; Liu, J.J.; Chen, J.F.; Chen, Q.Z.; Yan, X.J.; Xuan, J.L.; Zeng, J.N. Responses of summer phytoplankton community to drastic environmental changes in the Changjiang (Yangtze River) estuary during the past 50 years. *Water Res.* **2014**, *54*, 1–11.

8. Hadjisolomou, E.; Stefanidis, K.; Papatheodorou, G.; Papastergiadou, E. Assessing the Contribution of the Environmental Parameters to Eutrophication with the Use of the "PaD" and "PaD2" Methods in a Hypereutrophic Lake. *Int. J. Environ. Res. Public Health* **2016**, *13*, 764. https://doi.org/10.3390/ijerph13080764

9. Kline, D.; Kuntz, N.; Breitbart, M.; Knowlton, N.; Rohwer, F. Role of Elevated Organic Carbon Levels and Microbial Activity in Coral Mortality. *Mar. Ecol. Prog. Ser.* **2006**, *314*, 119-125. doi:10.3354/meps314119.

10. Tsikoti, C.; Genitsaris, S. Review of Harmful Algal Blooms in the Coastal Mediterranean Sea, with a Focus on Greek Waters. *Diversity* **2021**, *13*, 396. https://doi.org/10.3390/d13080396

11. Benkov, I.; Varbanov, M.; Venelinov, T.; Tsakovski, S. Principal Component Analysis and the Water Quality Index—A Powerful Tool for Surface Water Quality Assessment: A Case Study on Struma River Catchment, Bulgaria. *Water* **2023**, *15*, 1961. https://doi.org/10.3390/w15101961

12. Chau, K.-W. A review on integration of artificial intelligence into water quality modelling. *Mar. Pollut. Bull.* **2006**, *52*, 726-733

13. Devillers, J. Artificial Neural Network Modeling of the Environmental Fate and Ecotoxicity of Chemicals. In *Ecotoxicology Modeling*.; Devillers, J. Eds.; Springer-Verlag: Boston, USA, 2009.

14. Youssef, K.; Shao, K.; Moon, S.; Bouchard, L.-S. Landslide susceptibility modeling by interpretable neural network. *Commun. Earth Environ.* **4** 2023, *162*. https://doi.org/10.1038/s43247-023-00806-5

15. Kilic, H.; Soyupak, S.; Gurbuz, H.; Kivrak, E. Automata networks as preprocessing technique of artificial neural network in estimating primary production and dominating phytoplankton levels in a reservoir: An experimental work. *Ecol. Inform.* **2006**, *1*, 431-439.

16. Cereghino, R.; Park, Y.-S. Review of the Self-Organizing Map (SOM) approach in water resources: Commentary. *Environ. Model. Softw.* **2009**, *24*, 945-947. https://doi.org/10.1016/j.envsoft.2009.01.008.

17. Li, T.; Sun, G.; Yang, C.; Liang, K.; Ma, S.; Huang, L. Using self-organizing map for coastal water quality classification: Towards a better understanding of patterns and processes. *Sci. Total Environ.* **2018**, *628-629*,1446–1459. https://doi.org/10.1016/j.scitotenv.2018.02.163.

18. Peeters, L.; Dassargues, A. (2006) Comparison of Kohonen's self-organizing map algorithm and principal component analysis in the exploratory data analysis of a groundwater quality dataset. Proceedings of the 6th International Conference on Geostatistics for Environmental Applications. Rhodos, Greece, 25–27 October 2006; pp 1–12

19. Park, Y.-S.; Verdonschot, P.F.M.; Chon, T.-S.; Lek, S. Patterning and predicting aquatic macroinvertebrate diversities using artificial neural network. *Water Res.* **2003**, *37*, 1749-1758.

20. Lu, R.S.; Lo, S.L. Diagnosing reservoir water quality using self-organizing maps and fuzzy theory. *Water Res.* **2002**, *36*, 2265-2274. doi: 10.1016/s0043-1354(01)00449-3.

15

21. Li, J.; Shi, Z.; Wang, G.; Liu, F. Evaluating Spatiotemporal Variations of Groundwater Quality in Northeast Beijing by Self-Organizing Map. *Water* **2020**, *12*, 1382. https://doi.org/10.3390/w12051382

22. Hadjisolomou, E., Antoniadis, K., Vasiliades L.; Rousou, M.; Thasitis, I.; Abualhaija, R.; Herodotou, H.; Michaelides, M.; Kyriakides, I. Predicting Coastal Dissolved Oxygen Values with the Use of Artificial Neural Networks: A Case Study for Cyprus. *IOP Conf. Ser.: Earth Environ. Sci* **2022**, *1123*. doi: 10.1088/1755-1315/1123/1/012083

23. Salami, E.S.; Salari, M.; Rastergarc, M.; Sheibani, S.N.; Ehteshami, M. Artificial neural network and mathematical approach for estimation of surface water quality parameters (case study: California, USA). *Desalin. Water Treat.* **2021**, *213*, 75-83, https://doi.org/10.5004/dwt.2021.26709

24. Melesse, A.; Krishnaswamy, J.; Zhang, K. Modeling Coastal Eutrophication at Florida Bay using Neural Networks. *J. Coast. Res.* **2009**, *24*, 190-196. https://doi.org/10.2112/06-0646.1

25. Hadjisolomou, E.; Stefanidis, K.; Herodotou, H.; Michaelides, M.; Papatheodorou, G.; Papastergiadou, E. Modelling Freshwater Eutrophication with Limited Limnological Data Using Artificial Neural Networks. *Water* **2021**, *13*, 1590. https://doi.org/10.3390/w13111590

26. Georgescu, P.L.; Moldovanu, S.; Iticescu, C.; Calmuc, M.; Calmuc, V.; Topa, C.; Moraru, L. Assessing and forecasting water quality in the Danube River by using neural network approaches. *Sci. Total Environ.* **2023**, *879*, 162998. doi: 10.1016/j.scitotenv.2023.162998.

27. Moiseenko, T.I. Surface Water under Growing Anthropogenic Loads: From Global Perspectives to Regional Implications. *Water* **2022**, *14*, 3730. https://doi.org/10.3390/w14223730

28. Tselepides, A.; Papadopoulou, N.; Podaras, D.; Plaiti, W.; Koutsoubas, D. Macrobenthic community structure over the continental margin of Crete (South Aegean Sea NE Mediterranean) *Prog. Oceanogr.* **2000**, *46*, 401–428. https://doi.org/10.1016/S0079-6611(00)00027-6

29. Azov, Y. Eastern Mediterranean—a marine desert? *Mar. Pollut. Bull.* **1991**, *23*, 225 – 232.

30. Antoniadis, K.; Rousou, M.; Markou, M.; Stavrou, P.; Vasileiou, E.; Vasiliades, V.; Iosiphides, M.; Papadopoulos, V.; Argyrou, M. Review-update report of the coastal waters in accordance with Article 5 of the Water Framework Directive (WFD) 2000/60/EC for the period 2013-2019. Department of Fisheries and Marine Research, Ministry of Agriculture, Rural Development and the Environment, Cyprus [In Greek] 2020. Retrieved from: http://www.moa.gov.cy/moa/dfmr/.

31. Kuo, Y.-M.; Liu, C.-W.; Lin, K.-H. Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan. *Water Res.* **2004**, *38*, 148-158. https://doi.org/10.1016/j.watres.2003.09.026.

32. Kohonen, T.; Kaski, S. Exploratory Data Analysis by The Self Organizing Maps: Structure of Welfare and Poverty in the World. Proceedings of the Third International Conference on Neural Networks in the Capital Markets, London, England, 11-13 October 1995.

33. Dedecker, A.P.; Goethals, P. L.M.; Gabriels, W.; De Pauw, N. Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). *Ecol. Model.* **2004**, *174*, 161-173, https://doi.org/10.1016/j.ecolmodel.2004.01.003

34. Hu, Z.; Zhang, Y.; Zhao, Y.; Xie, M.; Zhong, J.; Tu, Z.; Liu, J. A Water Quality Prediction Method Based on the Deep LSTM Network Considering Correlation in Smart Mariculture. Sensors **2019**, *19*, 1420. https://doi.org/10.3390/s19061420

35. Lee, J.H.W.; Huang, Y.; Dickman, M.; Jayawardena, A.W. Neural networking modelling of coastal algal blooms *Ecol. Model.* **2003**, *159*, 179-201. https://doi.org/10.1016/S0304-3800(02)00281-8

36. Kohonen, T. *Self-organising maps*; Springer-Verlag: Berlin Heidelberg, Germany, 2001

37. Al-Mudhaf, H.F., A.M. Astel, M.I. Selim, and A.I. Abu-Shady. 2010. Self-organizing map approach in assessment spatiotemporal variations of trihalomethanes in desalinated drinking water in Kuwait. *Desalination* 252:97–105. doi: 10.1016/j.desal.2009.10.017.

38. Park, Y.-S.; Tison, J.; Lek, S.; Giraudel, J.-L., Coste, M.; Delmas, F. Application of a self-organizing map to select representative species in multivariate analysis: A case study determining diatom distribution patterns across France. *Ecol. Inform.* **2006**, *1*, 247-257.

39. An, Y.; Zou, Z.; Li, R. Descriptive Characteristics of Surface Water Quality in Hong Kong by a Self-Organising Map. *Int. J. Environ. Res. Public Health* **2016**, *13*, 115. https://doi.org/10.3390/ijerph13010115

40. Choi, J.-Y.; Kim, S.-K.; Jeng, K.-S.; Joo, G.-J. Detecting response patterns of zooplankton to environmental parameters in shallow freshwater wetlands: discovery of the role of macrophytes as microhabitat for epiphytic zooplankton. *J. Ecol. Environ.* **2015,** *38*, 133-143. doi: 10.5141/ecoenv.2015.015.

41. Kim, D.-K.; Kaluskar, S.; Mugalingam, S.; Arhonditsis, G.B. Evaluating the relationships between watershed physiography, land use patterns, and phosphorus loading in the bay of Quinte basin, Ontario, Canada. *J. Great Lakes Res.* **2016**, *42*, 972-984. https://doi.org/10.1016/j.jglr.2016.07.008.

42. Vesanto, J.; Alhoniemi, E. Clustering of the Self-Organizing Map. IEEE Trans. Neural Netw. **2000**, *11*, 586–600.

43. Vesanto, J. SOM-based data visualization methods. *Intell. Data Anal.* **1999**, *3*, 111-126.

44. Kalteh, A. M.; Hjorth, P.; Berndtsson, R. Review of the Self-Organizing Map (SOM) approach in water resources: analysis, modelling and application. *Environ. Model. Softw.* **2008**, *23*, 835-845. https://doi.org/10.1016/j.envsoft.2007.10.001

45. Vesanto, J.; Alhoniemi, E.; Himberg, J.; Parhankangas, J. SOM Toolbox for Matlab. 2000. Available online: http://www.cis.hut.fi/projects/somtoolbox/

46. Garcia-Avila, F.; Loja-Suco, P.; Siguenza-Jeton, C.; Jimenez-Ordonez, M.; Valdiviezo-Gonzales, L.; Cabello-Torres, R.; Aviles-Anazco, A. Evaluation of the water quality of a high Andean lake using different quantitative approaches. *Ecol. Indic.* **2023**, *154*, 110924. https://doi.org/10.1016/j.ecolind.2023.110924.

47. Bernard, J.; Landesberger, T.; Bremm, S.; Schreck, T. (2011). Multi-Scale Visual Quality Assessment for Cluster Analysis with Self-Organizing Maps. Proceedings of the SPIE Conference on Visualization and Data Analysis **2011**, *7868*, https://doi.org/10.1117/12.872545.

48. Wang, X.; Li, Y.; Qiao, Q.; Tavares, A.; Liang, Y. Water Quality Prediction Based on Machine Learning and Comprehensive Weighting Methods. *Entropy* **2023**, *25*, 1186. https://doi.org/10.3390/e25081186

49. Zhang, P.; Hong, B.; He, L.; Cheng, F.; Zhao, P.; Wei, C.; Liu, Y. Temporal and spatial simulation of atmospheric pollutant PM2.5 changes and risk assessment on population exposure to pollution using optimization algorithms of the back propagation-Artificial Neural Network model and GIS. *Int. J. Environ. Res. Public Health* **2015**, *12*, 12171–12195.

50. Papastergiadou, E.; Kagalou, I.; Stefanidis, K.; Retalis, A.; Leonardos, I. Effects of anthropogenic Influences on the trophic state, land uses and aquatic vegetation in a shallow Mediterranean Lake: Implications for restoration. *Water Resour. Manag.* **2010**, *24*, 415–435.

51. Hadjisolomou, E.; Stefanidis, K.; Papatheodorou, G.; Papastergiadou, E. Assessment of the Eutrophication-Related Environmental Parameters in Two Mediterranean Lakes by Integrating Statistical Techniques and Self-Organizing Maps. *Int. J. Environ. Res. Public Health* **2018**, *15*, 547. https://doi.org/10.3390/ijerph15030547

52. Peppa, M.; Vasilakos, C.; Kavroudakis, D. Eutrophication Monitoring for Lake Pamvotis, Greece, Using Sentinel-2 Data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 143. https://doi.org/10.3390/ijgi9030143

53. Chon, T.-S. Self-Organizing Maps applied to ecological sciences. *Ecol. Inform.* **2011**, *6*, 50-61. https://doi.org/10.1016/j.ecoinf.2010.11.002.

54. Qian, J.; Nguyen, N. P.; Oya, Y.; Kikugawa, G.; Okabe, T.; Huang, Y.; Ohuchi, F.S. Introducing self-organized maps (SOM) as a visualization tool for materials research and education. *Results Mater.* **2019**, *4*, 100020. https://doi.org/10.1016/j.rinma.2019.100020.

55. Krasznai, E.; Boda, P.; Csercsa, A.; Ficsor, M.; Varbiro, G. Use of self-organizing maps in modelling the distribution patterns of gammarids (Crustacea: Amphipoda). *Ecol. Inform.* **2016,** *31*, 39-48. https://doi.org/10.1016/j.ecoinf.2015.11.007.

56. Astel, A.; Tsakovski, S.; Barbieri, P.; Simeonov,V. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Res.* **2007**, *41*, 4566-4578. https://doi.org/10.1016/j.watres.2007.06.030.

57. Varbiro, G.; Acs, E.; Borics, G.; Erces, K.; Feher, G.; Grigorszky, I.; Japport, T.; Kocsis, G.; Krasznai, E.; Nagy, K.; Nagy-Laszlo, Z.; Pilinszky, Z.; Kiss, K.T. Use of Self-Organizing Maps (SOM) for characterization of riverine phytoplankton associations in Hungary. *Arch. Hydrobiol.* **2007**, *17*, 383-394.

58. Palani, S.; Liong, S.-Y.; Tkalich, P. An ANN application for water quality forecasting. *Mar. Pollut. Bullet.* **2008**, *56*, 1586–1597.

59. Bushra, B.; Bazneh, L.; Deka, L; Wood, P.J.; McGowan, S.; Das, D.B. Temporal modelling of long-term heavy metal concentrations in aquatic ecosystems. *J. Hydroinformatics* **2023**, *25*, 1188–1209. doi: https://doi.org/10.2166/hydro.2023.151

60. Brown, M.G.L.; Skakun, S.; He, T.; Liang, S. Intercomparison of Machine-Learning Methods for Estimating Surface Shortwave and Photosynthetically Active Radiation. *Remote Sens.* **2020**, *12*, 372. https://doi.org/10.3390/rs12030372

61. Petrou, A.; Kallianiotis, A.; Hannides, A. K.; Charalambidou, I.; Hadjichristoforou, M.; Hayes, D. R.; Lambridis, C.; Lambridi, V. et al. Initial Assessment of the Marine Environment of Cyprus: Part I–Characteristics. Ministry of Agriculture, Natural Resources, and the Environment, Department of Fisheries and Marine Research, Nicosia, Cyprus, 2012.

62. Fyttis, G.; Zervoudaki, S.; Sakavara, A.; Sfenthourakis, S. Annual cycle of mesozooplankton at the coastal waters of Cyprus (Eastern Levantine basin). *J. Plankton Res.* **2023**, *45*, 291–311. https://doi.org/10.1093/plankt/fbac075

63. Espinosa-Carreon, T.; Gaxiola-Castro, G.; Robles-Pacheco, J.; Najera-Martínez, S. Temperature, salinity, nutrients and chlorophyll a in coastal waters of the Southern California Bight. *Cienc. Mar.* **2001**, *27*, 397–422. https://doi.org/10.7773/cm.v27i3.490

64. Georgiou, N.; Fakiris, E.; Koutsikopoulos, C.; Papatheodorou, G.; Christodoulou, D.; Dimas, X.; Geraga, M.; Kapellonis, Z.G.; Vaziourakis, K.-M.; Noti, A.; et al. Spatio-Seasonal Hypoxia/Anoxia Dynamics and Sill Circulation Patterns Linked to Natural Ventilation Drivers, in a Mediterranean Landlocked Embayment: Amvrakikos Gulf, Greece. *Geosciences* **2021**, *11*, 241. https://doi.org/10.3390/geosciences11060241

65.   Suursaar, U. Winter upwelling in the Gulf of Finland, Baltic Sea. *Oceanologia* **2021**, *63*, 356-369. https://doi.org/10.1016/j.oceano.2021.04.001

66.   Ren, L.; Huang, J.; Zhu, H.; Jiang, W.; Wu, H.; Pan, Y.; Mao, Y.; Luo, M.; Jeong, T. Effects of Algal Utilization of Dissolved Organic Phosphorus by *Microcystis Aeruginosa* on Its Adaptation Capability to Ambient Ultraviolet Radiation. *J. Mar. Sci. Eng.* **2022**, *10*, 1257. https://doi.org/10.3390/jmse10091257

67.   Paerl, H.; Dennis, R.; Whitall, D. Atmospheric Deposition of Nitrogen: Implications for Nutrient Over-Enrichment of Coastal Waters. *Estuaries Coast.* **2002**, *25*, 677-693.

68.   Droge, R.; Kroeze, C. Critical load exceedance for nitrogen in the Ebrié Lagoon (Ivory Coast): a first assessment. *J. Integr. Environ. Sci.* **2007**, *4*, 5-19. doi: 10.1080/15693430601164949

69.   Duarte, I.; Ribeiro; M.C., Pereira, M.J.; Leite, P.P.; Peralta-Santos, A.; Azevedo, L. Spatiotemporal evolution of COVID-19 in Portugal's Mainland with self-organizing maps. *Int J Health aGeogr* **2023**, *22*, 4. https://doi.org/10.1186/s12942-022-00322-3

70.   Varbiro, G.; Borics, G.; Kiss, T.K.; Szabo, K.E.; Plenkovic-Moraj, A.; Acs, E. Use of Kohonen Self Organizing Maps (SOM) for the characterization of benthic diatom associations of the River Danube and its tributaries. *Arch. Hydrobiol.* **2007**, *17*, 395-403.

71.   Scardi, M. Advances in neural network modeling of phytoplankton primary production. *Ecol. Model.* **2001**, *146*, 33-45. https://doi.org/10.1016/S0304-3800(01)00294-0.

72.   Hadjisolomou, E.; Antoniades, K.; Thasitis, I.; Abu Alhaija, R.; Herodotou, H.; Michaelides, M. Exploring the Impact of Coastal Water Quality Parameters on Chlorophyll-a near Cyprus with the use of Artificial Neural Networks. Proceedings of the IAHR World Congress, Granada, Spain, 19-24 June 2022. doi:10.3850/IAHR-39WC2521711920221724.

73.   Xu, P.; Ji, X.; Li, M.; Lu, W. Small data machine learning in materials science. *npj Comput. Mater.* **2023**, *9*, 42. https://doi.org/10.1038/s41524-023-01000-z

74.   Shan, K.; Ouyang, T.; Wang, X.; Yang, H.; Zhou, B.; Wu, Z.; Shang, M. Temporal prediction of algal parameters in Three Gorges Reservoir based on highly time-resolved monitoring and long short-term memory network. *J. Hydrol.* **2022**, *605*, 127304. https://doi.org/10.1016/j.jhydrol.2021.127304.

75.   Silva, P.L.C.; Borges, A.C.; Lopes, L.S.; Rosa, A.P. Developing a Modified Online Water Quality Index: A Case Study for Brazilian Reservoirs. *Hydrology* **2023**, *10*, 115. https://doi.org/10.3390/hydrology10060115

76.   Chia, M.Y.; Huang, Y.F.; Koo, C.H. Resolving data-hungry nature of machine learning reference evapotranspiration estimating models using inter-model ensembles with various data management schemes. *Agric. Water Manag.* **2022**, *261*, 107343. https://doi.org/10.1016/j.agwat.2021.107343.

77.   Giangrande, A.; Gravina, M.F.; Rossi, S.; Longo, C.; Pierri, C. Aquaculture and Restoration: Perspectives from Mediterranean Sea Experiences. *Water* **2021**, *13*, 991. https://doi.org/10.3390/w13070991

78.   Eze, E.; Halse, S.; Ajmal, T. Developing a Novel Water Quality Prediction Model for a South African Aquaculture Farm. *Water* **2021**, *13*, 1782. https://doi.org/10.3390/w13131782

79.   Shah, M.I.; Alaloul, W.S.; Alqahtani, A.; Aldrees, A.; Musarat, M.A.; Javed, M.F. Predictive Modeling Approach for Surface Water Quality: Development and Comparison of Machine Learning Models. *Sustainability* **2021**, *13*, 7515. https://doi.org/10.3390/su13147515

80.   Ahmed, A.A.M.; Jui, S.J.J.; Chowdhury, M.A.I.; Ahmed, O.; Sutradha, A. The development of dissolved oxygen forecast model using hybrid machine learning algorithm with hydro-meteorological variables. *Environ Sci Pollut Res* **2023**, *30*, 7851–7873. https://doi.org/10.1007/s11356-022-22601-z

81.   Ahmed, A.A.M. Prediction of dissolved oxygen in Surma River by biochemical oxygen demand and chemical oxygen demand using the artificial neural networks (ANNs). *J. King Saud Univ. Eng. Sci.* **2017**, *29*, 151-158. https://doi.org/10.1016/j.jksues.2014.05.001.