MDPI

Article

Data-Driven Models 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
2   University of Nicosia Research Foundation, 1700 Nicosia, Cyprus; ioannis.kyriakides@cmmi.blue
3   Department of Fisheries and Marine Research, Ministry of Agriculture, Rural Development and the Environment, 2033 Nicosia, Cyprus; mrousou@dfmr.moa.gov.cy (M.R.); kantoniadis@dfmr.moa.gov.cy (K.A.); lvasiliades@dfmr.moa.gov.cy (L.V.)
4   Cyprus Marine and Maritime Institute, 6023 Larnaca, Cyprus
*   Correspondence: e.hadjisolomou@cut.ac.cy (E.H.); michalis.michaelides@cut.ac.cy (M.M.)

Abstract: Eutrophication is a major environmental issue with many negative consequences, such as hypoxia and harmful cyanotoxin production. Monitoring coastal eutrophication is crucial, especially for island countries like the Republic of Cyprus, which are economically dependent on the tourist sector. Additionally, the open-sea aquaculture industry in Cyprus has been exhibiting an increase in recent decades and environmental monitoring to identify possible signs of eutrophication is mandatory according to the legislation. Therefore, in this modeling 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. These ANNs aim to model the eutrophication phenomenon based on two different data-driven modeling 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 (Chl-a) parameter. The SOM model enables us to visualize the monitored parameters' relationships and to comprehend complex biological mechanisms related to Chl-a production. A second feed-forward ANN model is also developed for predicting the Chl-a levels. The feed-forward ANN managed to predict the Chl-a levels with great accuracy (MAE = 0.0124; R = 0.97). The sensitivity analysis results revealed that salinity and water temperature are the most influential parameters on Chl-a production. Moreover, the sensitivity analysis results of the feed-forward ANN captured the winter upwelling phenomenon that is observed in Cypriot coastal waters. Regarding the SOM results, the clustering verified the oligotrophic nature of Cypriot coastal waters and the good water quality status (only 1.4% of the data samples were classified as not good). The created ANNs allowed us to comprehend the mechanisms related to eutrophication regarding the coastal waters of Cyprus and can act as useful management tools regarding eutrophication control.

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 to human activities is a historically identified problem, but it has only received the necessary attention in recent years, when the anthropogenic pressure on aquatic ecosystems and organisms has reached a dangerous ecological threshold [2]. This intense anthropogenic pressure on the coastal environment is a result of the doubling of the human population and rapid

industrial development [1]. Some of these anthropogenic activities impacting the coastal zones are related to 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 coastal water results in harmful effects for marine organisms and negatively impacts human wellbeing.

Eutrophication is considered to be 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 to two nutrients, nitrogen (N) and phosphorus (P). Furthermore, climate change and anthropogenic eutrophication have resulted in large variations in microalgae assemblage composition globally, like increases in harmful algal blooms (HABs) or biomass [7]. The main impacts of these changes in algal composition include hypoxia/anoxia [8] with catastrophic side effects on aquatic organisms (e.g., declining fishery stocks). Additionally, eutrophication may trigger harmful bacterial production, which negatively affects corals and other marine organisms [9]. Another side effect of eutrophication is related to nuisance blooms, which have 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 the 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 are crucial issues for the associated environmental authorities [11]. Traditional methodologies include the analysis of data by using statistical methods, such as cluster analysis and ordination. Modeling studies have demonstrated that the application of suitable models, like artificial neural networks (ANNs), enables us to examine the association/impact of several environmental parameters on water quality problems, like eutrophication [8].

The majority of ANN-based hydrological modeling studies refer to freshwater applications compared to maritime studies [12]. Specifically, for the eutrophication phenomenon, the adaptation of data-driven models like ANNs has a beneficial role for environmental control and prevention [13]. As stated by Yussef et al. [14], in contrast to some other modeling techniques (e.g., statistical methods), ANNs are not affected by nonlinearities or the complex interdependencies of interlayer connections.

The eutrophication process and the role of the related environmental parameters can be evaluated by utilizing ANN techniques [15]. The Kohonen Self-Organising Maps (SOMs), which are unsupervised models [16], have clustering [17] and data mining abilities regarding dataset analysis [18]. Multivariate analysis is mainly applied for ecological patterning [19]; however, ANNs are more suitable for this task because of the nonlinear and complex possible interactions between the various environmental parameters.

Based on the Kohonen SOM model utilization, Lu and Lo [20] created a eutrophication status classifier and examined the environmental quality of the Fei-Tsui Reservoir. In another SOM application of Li et al. [21], the SOM model was applied to evaluate the groundwater quality of spatial data, and based on SOM clustering, several anthropogenic activities were identified for the related sampling sites.

Another category of ANNs are multilayer feed-forward neural networks, which are supervised-learning-based ANNs. This type of ANN is capable of predicting Chl-a levels based on several water quality parameters associated with algal production [22]. These environmental parameters, which are used as the ANN's input, may differ among modeling studies of coastal eutrophication. Salami et al. [23] created a back-propagation ANN for predicting coastal Chl-a values near Grant Line Canal, California, USA, based on electric conductivity (EC), water temperature (WT), and pH parameters. Even though only three monitoring parameters were used as the model's input, the model calculated the Chl-a

variable with a satisfactory accuracy rate (75.9%). In another study by Melesse et al. [24], the coastal Chl-a levels in Florida Bay, USA were also modeled with the use of a supervised ANN. Specifically, various combinations of seven candidate input parameters (total phosphate, nitrite, ammonium, turbidity, WT, dissolved oxygen (DO), and antecedent Chl-a were examined, and it was concluded that the ANN performed better when all aforementioned parameters were used as inputs to the ANN.

Data-driven models based on ANN 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 production and related environmental parameters [25]. Additionally, as stated by Georgescu et al. [26], the application of artificial intelligence (AI) methods for water quality modeling saves time and resources in lab analysis, while the generated statistical data are important for the relevant authorities/managers. Motivated by the above practical reasons, a new feed-forward ANN model, which is suitable for regression purposes, is designed to predict Cypriot coastal Chl-a levels at several locations. Furthermore, a new SOM model is proposed for the first time for Cypriot coastal waters, which enables us to comprehend to a greater extent possible hidden mechanisms and interactions between Chl-a and the rest of the eutrophication-related parameters.

This modeling study focuses on the role/interactions of water quality parameters associated with eutrophication and the impact of anthropogenic activity for several coastal areas near the Republic of Cyprus. Land uses of the different regions near the sea catchment area are reflected in the nutrients' concentrations in the nearby coastal areas, while it is well-documented that excessive amounts of nutrients in the surface water may lead to eutrophication [27]. Based on the SOM's clustering abilities, the association between the water quality near the sampling stations and the related anthropogenic activities can be extracted by observing/interpreting the SOM's results and by making associations among the water quality parameters, with a focus on the nutrients. In our case, it was found that the water quality status of Cyprus is good and practically unimpacted 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. It is also important to note that no other similar modeling study based on SOM models exists for the Cypriot coastal waters.

2. Materials and Methods

2.1. Study Area and Data Acquisition

The Republic of Cyprus is an island country, located in the Levantine Basin (Eastern Mediterranean area). The Levantine Sea is considered as one of the most oligotrophic seas worldwide [28] and, therefore, Cypriot marine waters have very low primary algal production due to the limited nutrient availability [29]. In addition, the Levantine Sea has high temperatures fluctuating annually from 16 °C (winter season) up to 26 °C (summer season) [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). Additionally, freshwater's inflow is very limited due to 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 program to collect, among others, water column data. A total of 49 coastal stations are monitored along the Cypriot coastline, some of which are located near anthropogenic activities such as aquaculture facilities and industrial units (Figure 1). Water column samples are collected and analyzed, and the data are stored in DFMR's "Thetis" database.

Figure 1. Satellite map of the Republic of Cyprus (where 1cm: 50 km), which is located in the Eastern Mediterranean region (green colored markers are used to indicate the sampling sites). For more details, please see the study of Antoniadis et al. [30].

　　For this modeling study, 1552 in situ samples were collected by the DFMR from several monitoring sites, based on which the ANN models were created. The data samples were collected sporadically (having no regular time intervals) from the 49 coastal stations between the years 2000 and 2020. The sampling frequency varied from monthly to yearly, depending on different monitoring programs applied to the 49 coastal stations during the sampling period. Bad meteorological conditions were also a limiting factor, causing discontinuities in the sampling process. Specifically, the water quality parameters that were measured/monitored are (i) nitrogen species (NH_4^+, NO_2^-, NO_3^-); (ii) ortho-phosphates (PO_4^{3-}); (iii) salinity; (iv) DO; (v) pH; (vi) EC; (vii) WT; and (viii) Chl-a. More information/details about the sampling stations and the data monitoring process are in the technical report of Antoniadis et al. [30]. Table 1 provides a statistical description of the measured environmental parameters.

Table 1. Statistical description of the measured environmental parameters.

Variable	Units	Min.	Max.	Mean $\pm$ SD
Salinity	psu	0.86	44.05	37.74 $\pm$ 2.88
pH	-	0.25	8.79	7.88 $\pm$ 0.48
Dissolved Oxygen	mg/L	1.6	8.61	7.84 $\pm$ 0.39
Chlorophyll-*a*	µg/L	0.01	10.80	0.22 $\pm$ 0.59
Water Temperature	°C	21.68	27.69	22.25 $\pm$ 0.32
Electrical Conductivity	mS/cm	0.8	67.15	56.34 $\pm$ 4.79
Orthophosphates	µmol/L	0.00	17.77	0.46 $\pm$ 0.98
Ammonium	µmol/L	0.99	65.39	4.03 $\pm$ 3.51
Nitrite	µmol/L	0.00	24.21	0.49 $\pm$ 1.31
Nitrate	µmol/L	0.01	448.74	11.74 $\pm$ 24.94

2.2. Multilayer Feed-Forward ANNs

　　ANNs are inspired by the function of the biological neuron system, where a signal is received and processed by a neuron, and then an output signal is transmitted to the other interconnected neurons or nodes [31]. Multilayer feed-forward ANNs are supervised machine learning models and are capable of processing nonlinear phenomena [14,24]. According to Kohonen and Kaski [32], the multilayer feed-forward ANN is an efficient,

nonlinear, "general-purpose" function approximator. The multilayer perceptron (MLP) architecture is a layered feed-forward 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 node/neuron, which is connected with all the nodes/neurons found in the next layer.

The output value of the k-th neuron (o_k) is calculated by using the following equations [31]:

$$o_k = h(u_k) \tag{1}$$

$$u_k = \sum w_{ik} x_i + z_k \tag{2}$$

where h is the transfer function, x_i is the input from the k-th node of the immediate previous layer, w_{ik} corresponds to the synaptic weight that connects the input x_i with the k-th neuron, and z_k is the term corresponding to the bias. 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]. The goal of the training process is to find 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 variables serving as inputs might differ in magnitude [33]. The ANN's performance can be calculated using several statistical performance metrics, including the root mean square error ($RMSE$), the mean absolute error (MAE), and Pearson's correlation coefficient (R).

The MLP's sensitivity analysis can be performing using several algorithms. The perturbation sensitivity analysis algorithm, which demonstrates how the trained network reacts to a small change/perturbation of each input, is one of the most commonly used sensitivity analysis algorithms. The perturb sensitivity is calculated by using the following [35]:

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

where the parameter N_m corresponds to the number of samples.

2.3. Self-Organizing Map (SOM)

SOM is an unsupervised ANN, which means that no human supervision/intervention is necessary for its learning process [36]. The characterization "self-organizing" is given to the SOM because it can learn and organize information without knowing the corresponding output values of the input data [37]. The SOM is able to project high-dimensional data into a lower dimension space, usually a two-dimensional space [36].

The SOM has 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. Initialize the weight vector with random values.
2. Utilize a distance measure (e.g., the Euclidean distance) to calculate 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_k) calculates the distance measure between the input vector and the k-th weight vector [39] and is given by the next equation:

$$D_k = \sqrt{\sum_{j=1}^{V} \left(p_{kj} - w_{kj} \right)^2}\ ;\ k = 1, 2, \dots N, \tag{4}$$

where N is the number of output neurons, V is the dimension of the input vectors, p_{kj} symbolizes the j-th element of the input vector, and w_{kj} represents the j-th element of the k-th weight vector. The term BMU corresponds to the neuron with the weight vector closest to the input variable x, i.e., the weight vector that has the shortest distance to the input vector [41], and is calculated by using the equation:

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

where $|\cdot|$ is the symbol for the distance measure, x corresponds to the input vector, m_k corresponds to a weight vector, and c gives the subscript 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], which uses the following formula:

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

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

The 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 and are defined as the values of a single vector component in all map units [43].

The SOM can automatically group (cluster) data according to different properties of the dataset variables [44]. The data can be clustered either manually as determined with the U-matrix or automatically by using a clustering algorithm implemented in SOM using 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 modeling study, a SOM with 20×10 neurons was created. The SOM's topology, which is associated with the number of SOM neurons, was calculated after applying Equation (6). The data simulations were based on the SOM Toolbox 2.0 for MATLAB (available by the Laboratory of Information and Computer Science in the Helsinki University of Technology, Finland) [45]. The created SOM's U-matrix and the CPs are visualized in Figure 2.

The CPs revealed a strong positive relationship between EC, pH, and salinity since they had very similar CPs. Not surprisingly, the CPs for NH_4^+, NO_2^-, and PO_4^{3-} were associated with Chl-a, with a strong positive relationship; the highest values of NH_4^+, NO_2^-, and PO_4^{3-} parameters corresponded 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 other parameters, no clear conclusions can be derived through the CPs. Hence, in an additional step, to reveal hidden relationships/mechanisms between the parameters, the SOM's clusters statistical properties were 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) indicated a tendency for the data to be grouped into three clusters; however, this was 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 clusters. The Davies–Bouldin index was 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 is shown in Figure 3. The clustering of the SOM based on the k-means algorithm and the percentage of SOM hits for each cluster are illustrated in Figure 4.
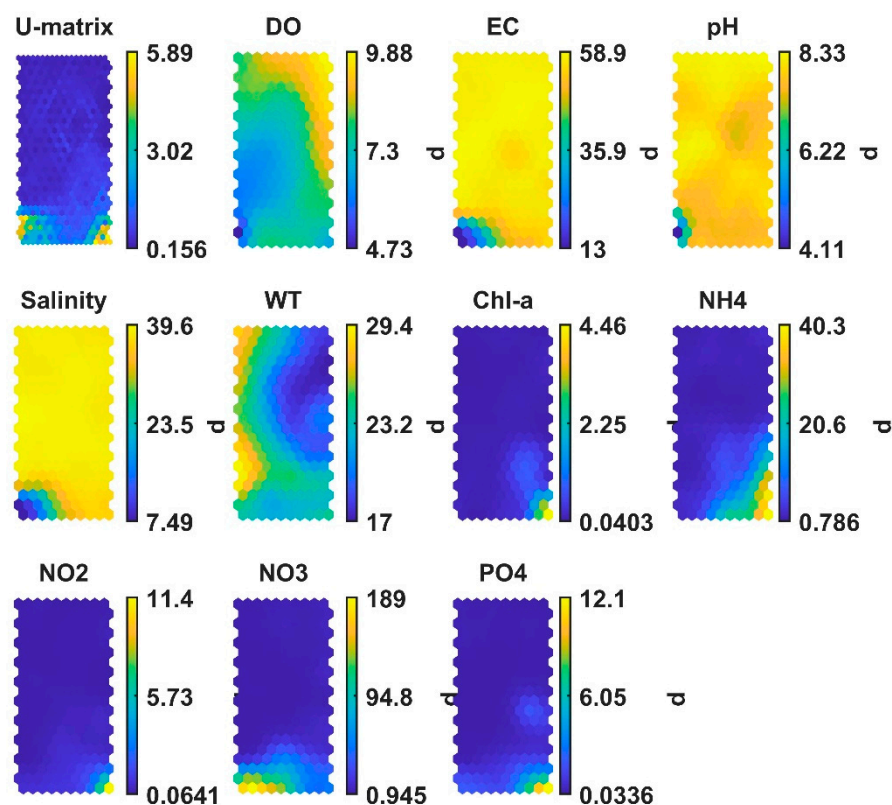
Figure 2. Visualization of SOM's component planes (CPs) for each environmental parameter, where the mapping of the data values is indicated by the colored bars.
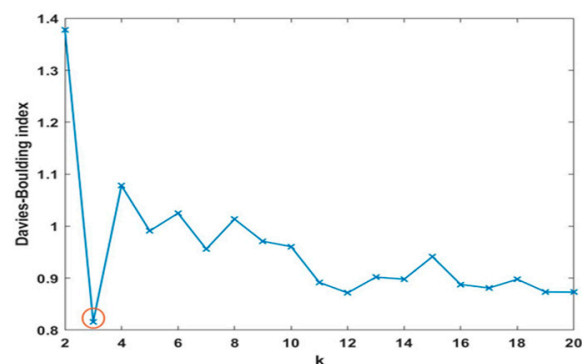

Figure 3. Calculation of the optimal number of clusters based on the minimization for the Davies–Bouldin index when the SOM is clustered using the *k*-means algorithm. The minimum number of the Davies–Boulding index (*k* = 3) is indicated in a red circle.

As indicated by the CPs and SOM's clustering (Figures 2 and 4), Cluster 2 (C2) has the worst water quality. The nutrients (except NO_3^-) and Chl-a have the highest concentrations for data belonging to Cluster 2. Regarding Cluster 1 (C1), the parameters NO_3^-, EC, and salinity have a significant influence on the cluster, while pH seems to be associated but to a lesser extent. Finally, Cluster 3 (C3) has the best water quality since it is characterized by low concentrations of Chl-a and nutrients; however, no clear associations can be inferred regarding the interactions among the water quality parameters. Nevertheless, it must be noted that based on the SOM's clustering, 95% of the samples are grouped into C3 (n3 = 1475), 3.6% of the samples are grouped into C1 (n1 = 56), and 1.4% of the samples are grouped into C2 (n2 = 21).
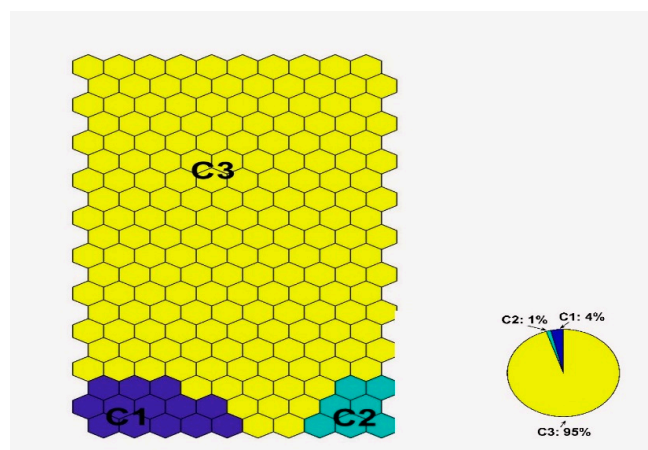
Figure 4. Clustering of the SOM based on the k-means algorithm (where Cluster 1: C1 is symbolized with blue, Cluster 2: C2 is symbolized with green, and Cluster 3: C3 is symbolized with yellow). The pie chart is presenting the percentage of SOM's samples for each cluster.

The boxplots in Figure 5 provide synopses for descriptive statistical properties (e.g., median value, percentile, and outliers) of the data belonging to each of the three formed SOM groups (C1, C2, C3) and for each SOM's input environmental parameter.
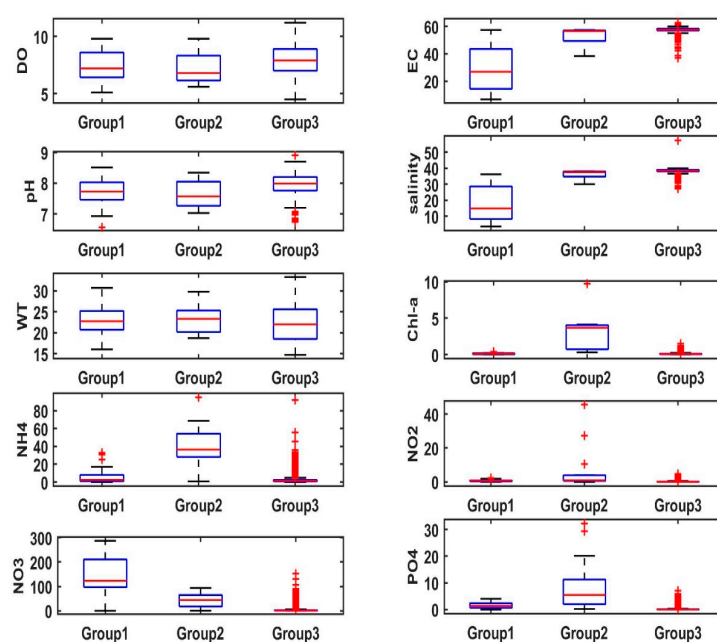
Figure 5. Boxplot graphical representation of the SOM's groups/clusters (Group 1, Group 2, Group 3) derived from the *k*-means algorithm for each input environmental parameter (where the red horizontal line denotes the group's median value; the blue box gives the 25–75% percentile range; the whiskers give the valid range; and red marks are associated with extreme values/outliers).

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 have clear differences between the three SOM groups. The rest of the parameters (DO, pH, WT, NO_2^-) seem to have similar statistical properties; however, the smaller magnitude of their value range should be taken into consideration. From the DO, pH, WT, and NO_2^- parameters, the NO_2^- is the only one without overlapped notches of its boxplots, indicating a clear differentiation between the three SOM groups.

3.2. 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 using min–max normalization, which projected the data to the range [0, 1], ensuring that feature variables had similar scales [48]. The ANN's optimal topology was found to be 9-6-1 after following a trial-and-error procedure. The ANN was trained with the Levenberg–Marquardt training algorithm since it is considered the 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 dataset ($n = 1552$) was divided into a training set and a test set at 80% and 20%, respectively, while the ANN was evaluated on the test set. The achieved performance metrics were $MAE = 0.0124$ and $R = 0.97$, whilea graphical illustration of the real and the predicted data of the test set is given in Figure 6. It can be 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 (namely, high, good, and moderate) for Cyprus are given in the embedded table in Figure 6. For Chl-a concentrations below 0.4 mg/L (high and good water quality status), the real and predicted data are almost a perfect match. Regarding the moderate-status Chl-a values, the ANN also managed to produce good outputs, as can be observed from Figure 6, except for one point corresponding to the highest measured value of the Chl-a parameter.
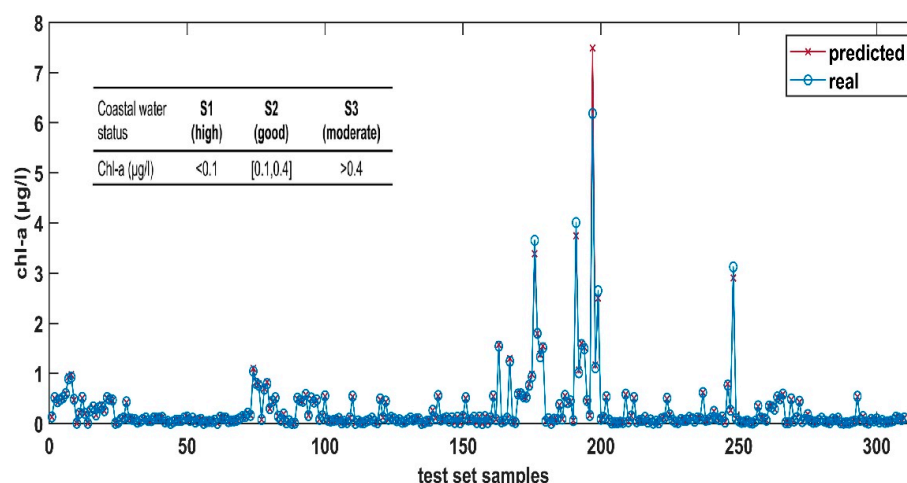
Figure 6. ANN's predicted values for Chlorophyll-a (Chl-a) levels regarding the test set data vs. the real Chl-a measurements, where the blue line is associated with the real data and the red line is associated with the predicted data. The embedded table is describing the Cypriot coastal water status for different Chl-a concentrations (where S1: high, S2: good, and S3: moderate).

Sensitivity analysis was performed to evaluate the input parameters' impact on the modeled 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 in the input parameters, it was calculated that the nutrients (i.e., PO_4^{3-}, NO_2^-, NH_4^+, and NO_3^-) have a positive relationship with the Chl-a production mechanism. In addition, pH, EC, and salinity are positively related to the Chl-a parameter, while WT and DO have a negative relationship with the algal production. In the case of −10% decrease in the input parameters, it was calculated that the Chl-a levels are decreased for PO_4^{3-}, NO_3^-, NH_4^+, salinity, EC, and pH, while the Chl-a levels are increased for NO_2^-, WT, and DO. The salinity (when negatively perturbated) and the WT (when positively perturbated) parameters are the most influential on Chl-a production.
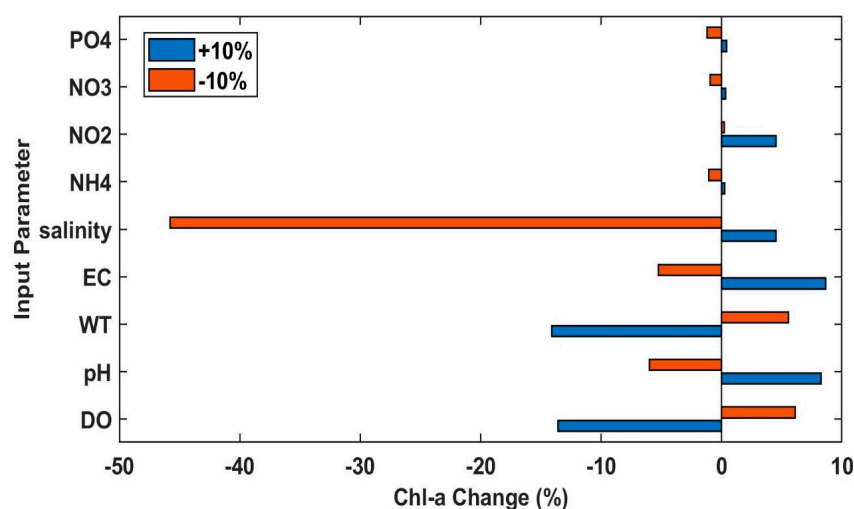
Figure 7. ANN's sensitivity analysis results for each of the input parameters. The fluctuation of each input parameter by an increase of +10% and the associated Chl-a change is symbolized with blue color, while the fluctuation of each input parameter by a decrease of −10% and the associated Chl-a change is symbolized with red color.

4. Discussion

The achievement and maintenance 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 modeling study, data-driven modeling techniques were applied, aiming to model the coastal water quality in several areas of Cyprus. Based on the modeling outputs, the Chl-a levels can be accurately predicted, while 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 modeling study. First, an unsupervised type of ANN was created, specifically the SOM model. Second, another type of ANN, the supervised feed-forward ANN, 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 enabled us to unravel hidden complex mechanisms between Chl-a and the rest of the water quality parameters. As stated by Chon [50], the integration of the SOM and MLP models promotes advanced information extraction from water quality datasets.

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 [51]. The utilization of the SOM is very beneficial when the correlation between the input parameters is nonlinear and/or when dealing with noisy data; under those conditions, the CPs can reveal relationships between the data that would not be otherwise detected [52]. In their study, Astel et al. [53] emphasized the SOM's classification and visualization ability for large water quality datasets, while the authors also mentioned the SOM's suitability for simultaneous observation of water quality parameters and their spatial and temporal changes based on the CP visualization. Meanwhile, Varbiro et al. [54] argued in favor of the SOM's superiority compared to 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 two-dimensonial space.

Regarding the second ANN implemented in this modeling study, the feed-forward ANN was chosen. Feed-forward ANNs are able to model nonlinear complex environmental systems [55]. Additionally, as stated by Bushra et al. [56], back-propagation 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. [57], ANN models give more reliable outputs in comparison to other machine learning methods (e.g., decision trees or linear regression) when the number of data measurements is relatively small, like in our case. Generally, feed-forward ANNs are considered reliable predictors of Chl-a and are widely used for the prediction of Chl-a levels [8].

As 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 was very small, which can be 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 tended to increase; however, the calculated ANN's values were 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%. Based on the results from the sensitivity analysis, 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 in phytoplankton production [58]. This was also recorded by Fyttis et al. [59] during a monitoring study of 12 consecutive months (January–December 2016), where the maximum coastal Chl-a levels for 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 suggested to be related to this, since during the upwelling phenomenon, nutrient-rich water emerges at the surface [60].

The upwelling phenomenon might also explain the strong negative relationship between DO and Chl-a. In a study by Georgiou et al. [61] in the Amvrakikos Gulf (Greece), low oxygen levels were reported during winter months. The above authors attribute 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 modeling studies in Cyprus. As mentioned by Suursaar [62], the wintertime upwelling is a phenomenon that has been ignored and not given the necessary attention, in contrast to the summer upwelling.

The rest parameters seem to contribute less to 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 to increased algal production and vice versa. As stated by Ren et al. [63], the high levels of dissolved inorganic phosphorus, mainly in the form of phosphate in the water column, could enhance algal production. Regarding the DIN species, a less important relationship with Chl-a is found, which has similar behavior to phosphorus. A major source of DIN in coastal waters is associated with atmospheric deposition. Two main sources of DIN are related to anthropogenic activities, specifically riverine inputs and atmospheric deposition. In the study of Paerl et al. [64] conducted 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. According to Droge and Kroeze [65], riverine inputs are considered the main source of nitrogen for coastal waters and, as estimated by the

authors based on modeling 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 modeling. In our case, the integration of a supervised and an unsupervised ANN proved 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 revealed the tendencies related to 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 parameter 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 dataset is examined as a whole. Furthermore, the SOM's results revealed hidden relationships between the water quality parameters, which could not be easily identified or understood based on other modeling 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. [66], complex patterns and interactions between the input parameters can be interpreted and understood based on the CP 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 have similar box plots and CPs, suggesting a strong relationship between Chl-a and the impact of NH_4^+, NO_2^-, and PO_4^{3-}. Regarding the NO_3^- parameter, its moderate concentrations are associated with the highest Chl-a values. The SOM's clustering of the dataset (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. [67] applied the SOM to evaluate the Danube's tributaries based on diatom association, where the authors concluded that the upper stretch (German–Austrian region) has better water quality than the lower stretch (Slovakian–Hungarian region). The 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 associations with different water quality statuses. In our case, this finding can provide important information to the local authorities relating to eutrophication, since it is indicated that not all nutrients must have the same treatment regarding eutrophication control, as analyzed above based on the box plot results (Figure 5).

Despite the limiting factor of the relatively small dataset used in this modeling 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 discussed above. It must be noted that in a previous modeling study, Hadjisolomou et al. [68] developed a feed-forward ANN that managed to predict the surface coastal Chl-a levels near Cyprus with a good accuracy ($R = 0.87$ for the test). However, the dataset was much smaller ($n = 681$) in comparison with the dataset of this modeling study ($n = 1552$). For that reason, the previous model was validated by applying the k-fold method, while the topology used for that ANN was different (9-8-1). As explained by Hadjisolomou et al. [25], the application of the k-fold method raises some concerns related to the small dataset for testing and, therefore, the evaluation might become less reliable and robust. Another important detail related to the nature of the dataset, which was analyzed in Hadjisolomou et al. [68], was that only one sporadic measurement with an elevated Chl-a value was recorded. As expected, the current ANN created for the needs of this modeling study has better performance ($R = 0.97$ for the test set), while differences related to 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 dataset that contains a significant number of high/elevated Chl-a measurements. Therefore, the current ANN, besides the fact that it performs better, can also generalize better in situations where algal production is increased. Thus, the creation of

updated ANN models based on denser measurements and a bigger database can provide information that is even more valuable and could allow us to better understand the algal production mechanisms.

It is generally accepted that water quality monitoring is a time-consuming and expensive procedure. Utilizing ANNs for the modeling of water quality parameters is considered the best practice compared to other experimental or monitoring methods, which are usually costly or take too long for data gathering [69]. In the study by Ahmed et al. [70], the various methods available for estimating the DO concentration are analyzed 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 nonlinearities. Therefore, the above authors propose using ML data-driven models for water quality modeling prediction purposes. The ML data-driven models used for prediction are able to overcome modeling limitations related to complex and nonlinear datasets and, therefore, are widely used in water quality modeling [31,71]. It must be noted that the eutrophication status can be evaluated directly based on measurable indicators like the nutrient (nitrogen and phosphorus) content, DO, turbidity, and Chl-a concentrations [72]. Additionally, some very simple modeling techniques dealing with eutrophication exist, for example, the linear regression method. However, as stated by Hadjisolomou et al. [25], such methods might be affected by nonlinearities, which commonly appear when examining the complex eutrophication mechanism and the associated parameters' interactions. 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 importance for 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

Two data-driven models were developed for evaluating the impact of eutrophication-related water quality parameters. The created ANNs managed to capture biological mechanisms/relationships and the special characteristics related to coastal algal production in Cyprus. The key findings from the ANNs are as follows:

- The feed-forward ANN, based on the sensitivity analysis results, revealed that the winter upwelling seems to have an important role in the eutrophication phenomenon, while the cooler WT measurements are associated with higher Chl-a levels.
- Based on the SOM clustering results, the water quality of Cypriot coastal waters is classified as good and only few data samples (1.4%) are classified as not good.

Therefore, it is recommended that any implementation measures regarding eutrophication control must be assessed based on modeling scenarios since data-driven models have been proven to be reliable prediction tools. The created ANNs cannot only predict Chl-a levels but can also extract thresholds for the associated water quality parameters, like the phosphate and the nitrogen species. Therefore, the ANNs created for the needs of this modeling study can act as a basis for advisory tools, contributing not only to Cypriot marine environmental protection but also to the local economy, as well, related to financial activities like coastal tourism, shipping, and aquaculture.

Author Contributions: Conceptualization, E.H.; methodology, E.H.; software, E.H.; data analysis, E.H., H.H. and M.M.; data curation, K.A., M.R. and L.V.; writing—original draft preparation, E.H.; writing—review and editing, E.H., K.A., M.R., L.V., H.H., M.M. and I.K.; supervision, H.H., M.M. and I.K. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Visbeck, M. Ocean science research is key for a sustainable future. *Nat. Commun.* **2018**, *9*, 690. [CrossRef]
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. [CrossRef]
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. [CrossRef]
4. He, Q.; Silliman, B.R. Climate Change, Human Impacts, and Coastal Ecosystems in the Anthropocene. *Curr. Biol.* **2019**, *29*, 1021–1035. [CrossRef] [PubMed]
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. [CrossRef]
6. Smith, V.H. Responses of estuarine and coastal marine phytoplankton to nitrogen and phosphorus enrichment. *Limnol. Oceanogr.* **2006**, *51*, 377–384. [CrossRef]
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. [CrossRef]
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. [CrossRef]
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. [CrossRef]
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. [CrossRef]
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. [CrossRef]
12. Chau, K.-W. A review on integration of artificial intelligence into water quality modelling. *Mar. Pollut. Bull.* **2006**, *52*, 726–733. [CrossRef]
13. Devillers, J. (Ed.) Artificial Neural Network Modeling of the Environmental Fate and Ecotoxicity of Chemicals. In *Ecotoxicology Modeling*; Springer: Boston, MA, USA, 2009.
14. Youssef, K.; Shao, K.; Moon, S.; Bouchard, L.-S. Landslide susceptibility modeling by interpretable neural network. *Commun. Earth Environ.* **2023**, *4*, 162. [CrossRef]
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. [CrossRef]
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. [CrossRef]
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. [CrossRef] [PubMed]
18. Peeters, L.; Dassargues, A. Comparison of Kohonen's self-organizing map algorithm and principal component analysis in the exploratory data analysis of a groundwater quality dataset. In 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. [CrossRef]
20. Lu, R.S.; Lo, S.L. Diagnosing reservoir water quality using self-organizing maps and fuzzy theory. *Water Res.* **2002**, *36*, 2265–2274. [CrossRef] [PubMed]
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. [CrossRef]
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*, 012083. [CrossRef]
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. [CrossRef]
24. Melesse, A.; Krishnaswamy, J.; Zhang, K. Modeling Coastal Eutrophication at Florida Bay using Neural Networks. *J. Coast. Res.* **2009**, *24*, 190–196. [CrossRef]

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. [CrossRef]
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. [CrossRef] [PubMed]
27. Moiseenko, T.I. Surface Water under Growing Anthropogenic Loads: From Global Perspectives to Regional Implications. *Water* **2022**, *14*, 3730. [CrossRef]
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. [CrossRef]
29. Azov, Y. Eastern Mediterranean—A marine desert? *Mar. Pollut. Bull.* **1991**, *23*, 225–232. [CrossRef]
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. 2020. Available online: http://www.moa.gov.cy/moa/dfmr/ (accessed on 12 May 2022). (In Greek)
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. [CrossRef]
32. Kohonen, T.; Kaski, S. Exploratory Data Analysis by The Self Organizing Maps: Structure of Welfare and Poverty in the World. In Proceedings of the Third International Conference on Neural Networks in the Capital Markets, London, UK, 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. [CrossRef]
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. [CrossRef] [PubMed]
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. [CrossRef]
36. Kohonen, T. *Self-Organising Maps*; Springer: Berlin/Heidelberg, Germany, 2001.
37. Al-Mudhaf, H.F.; Astel, A.M.; Selim, M.I.; Abu-Shady, A.I. Self-organizing map approach in assessment spatiotemporal variations of trihalomethanes in desalinated drinking water in Kuwait. *Desalination* **2010**, *252*, 97–105. [CrossRef]
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. [CrossRef]
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. [CrossRef]
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. [CrossRef]
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. [CrossRef]
42. Vesanto, J.; Alhoniemi, E. Clustering of the Self-Organizing Map. *IEEE Trans. Neural Netw.* **2000**, *11*, 586–600. [CrossRef]
43. Vesanto, J. SOM-based data visualization methods. *Intell. Data Anal.* **1999**, *3*, 111–126. [CrossRef]
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. [CrossRef]
45. Vesanto, J.; Alhoniemi, E.; Himberg, J.; Parhankangas, J. SOM Toolbox for Matlab. 2000. Available online: http://www.cis.hut.fi/projects/somtoolbox/ (accessed on 5 April 2023).
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. [CrossRef]
47. Bernard, J.; Landesberger, T.; Bremm, S.; Schreck, T. Multi-Scale Visual Quality Assessment for Cluster Analysis with Self-Organizing Maps. In Proceedings of the SPIE Conference on Visualization and Data Analysis, San Francisco, CA, USA, 23–27 January 2011; SPIE: Bellingham, WA, USA, 2011; Volume 7868. [CrossRef]
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. [CrossRef] [PubMed]
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. [CrossRef]
50. Chon, T.-S. Self-Organizing Maps applied to ecological sciences. *Ecol. Inform.* **2011**, *6*, 50–61. [CrossRef]
51. 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. [CrossRef]
52. 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. [CrossRef]

53. 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. [CrossRef]

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

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

56. 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. [CrossRef]

57. 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. [CrossRef]

58. Petrou, A.; Kallianiotis, A.; Hannides, A.K.; Charalambidou, I.; Hadjichristoforou, M.; Hayes, D.R.; Lambridis, C.; Lambridi, V.; Loizidou, X.I.; Orfanidis, S.; 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.

59. 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. [CrossRef] [PubMed]

60. 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. [CrossRef]

61. 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. [CrossRef]

62. Suursaar, U. Winter upwelling in the Gulf of Finland, Baltic Sea. *Oceanologia* **2021**, *63*, 356–369. [CrossRef]

63. 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. [CrossRef]

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

65. 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. [CrossRef]

66. 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 Geogr.* **2023**, *22*, 4. [CrossRef]

67. 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.

68. 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. In Proceedings of the IAHR World Congress, Granada, Spain, 19–24 June 2022. [CrossRef]

69. 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. [CrossRef]

70. 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. *Env. Sci Pollut. Res.* **2023**, *30*, 7851–7873. [CrossRef] [PubMed]

71. 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. [CrossRef]

72. Kitsiou, D.; Karydis, M. Coastal marine eutrophication assessment: A review on data analysis. *Environ. Int.* **2011**, *37*, 778–801. [CrossRef] [PubMed]