



Article

Using Machine-Learning Algorithms for Eutrophication Modeling: Case Study of Mar Menor Lagoon (Spain)

Patricia Jimeno-Sáez ^{1,*}, Javier Senent-Aparicio ¹, José M. Cecilia ²
and Julio Pérez-Sánchez ¹

¹ Department of Civil Engineering, Universidad Católica San Antonio de Murcia, Campus de los Jerónimos s/n, 30107 Guadalupe, Murcia, Spain; jsenent@ucam.edu (J.S.-A.); jperez058@ucam.edu (J.P.-S.)

² Department of Computer Engineering, Universitat Politècnica de València, Camí de Vera, s/n, 46022 Valencia, Spain; jmcecilia@disca.upv.es

* Correspondence: pjimeno@ucam.edu; Tel.: +34-968-278-818

Received: 16 January 2020; Accepted: 9 February 2020; Published: 13 February 2020



Abstract: The Mar Menor is a hypersaline coastal lagoon with high environmental value and a characteristic example of a highly anthropized hydro-ecosystem located in the southeast of Spain. An unprecedented eutrophication crisis in 2016 and 2019 with abrupt changes in the quality of its waters caused a great social alarm. Understanding and modeling the level of a eutrophication indicator, such as chlorophyll-a (Chl-a), benefits the management of this complex system. In this study, we investigate the potential machine learning (ML) methods to predict the level of Chl-a. Particularly, Multilayer Neural Networks (MLNNs) and Support Vector Regressions (SVRs) are evaluated using as a target dataset information of up to nine different water quality parameters. The most relevant input combinations were extracted using wrapper feature selection methods which simplified the structure of the model, resulting in a more accurate and efficient procedure. Although the performance in the validation phase showed that SVR models obtained better results than MLNNs, experimental results indicated that both ML algorithms provide satisfactory results in the prediction of Chl-a concentration, reaching up to 0.7 R^2_{CV} (cross-validated coefficient of determination) for the best-fit models.

Keywords: multilayer neural network (MLNN); support vector regression (SVR); water quality; eutrophication; chlorophyll-a; Mar Menor coastal lagoon

1. Introduction

Coastal lagoons are natural systems with significant environmental and socioeconomic value which occupy approximately 13% of the world's coastline, representing 5.3% of Europe [1]. They are shallow coastal water bodies isolated from the sea by a barrier, land spit or other similar land feature, but connected to the sea by one or more inlets, through which there is a more or less restricted exchange of water and organisms with the open sea [2,3]. They often exhibit high rates of primary production stimulated by the considerable amounts of nutrients received from surrounding basins [4]. Coastal lagoons are among the most productive ecosystems on the planet, being valuable ecosystems for fishing and aquaculture [5]. In addition, they are generally interesting environments for the development of other human activities, with ideal conditions for nautical sports and swimming, health and entertainment activities, saltworks and the retention and purification of pollutants, among others. However, these natural systems are especially vulnerable to human impacts and the entry of runoff materials [6]. Due to the great variety of transformations and anthropogenic pressures that alter the balance of the coastal lagoon ecosystem, the management of these territories is complex and can result

in environmental catastrophes [7]. Among other threats, eutrophication problems associated to human activities have been identified as one of the main causes of water quality impairment of inland and marine waters [8], and is a serious problem worldwide [9]. Eutrophication is a process derived from an increase in the ratio of organic matter supply to an ecosystem, where nutrient enrichment is the most common factor increasing this supply in coastal systems [10]. This nutrient load contributes to an accelerated algal bloom and higher forms of plant life that produce an undesirable disturbance of the equilibrium of the organisms present in the water. The biomass of phytoplankton, represented by chlorophyll-a (Chl-a), is an important indicator to evaluate the state of eutrophication of water bodies [11], and has been studied for decades [12,13].

In this context, this study focuses on the Mar Menor lagoon located in the south-eastern Spain as a representative case, due to the serious environmental problems it has suffered in recent decades [14]. Indeed, it is one of the most representative examples of environmental resilience, and has one of the most varied catalogs of anthropic effects on a coastal lagoon in the Mediterranean [15]. In 2016, an unprecedented eutrophication crisis led to serious social and economic problems for this region [15–17]. Subsequently, after the “Santa Maria” flood in September 2019, attributed to the meteorological phenomenon known as “gota fría” (cold drop), the ecological degradation of the Mar Menor was aggravated by the massive input of nitrogen, phosphorus and organic matter [18]. As a result of these incidents, the local government approved by decree-law some urgent measures to ensure the environmental sustainability of the Mar Menor area [19,20], and the decree-law on the integral protection of the Mar Menor [21]. These measures include controlling the application of fertilizers and implementing best management practices in the surrounding areas of the coastal lagoon to mitigate water quality problems.

The automatic water quality monitoring is a useful tool to control water quality, especially in critical areas where (1) potential episodes of pollution are expected and/or (2) relevant socioeconomic activities, which require preventive actions are performed. However, to the best of our knowledge, there is no automatic device that accurately measures Chl-a in real time, so Chl-a measurements have to be done in laboratories, which means high latency and high cost. Therefore, it is important to minimize the number of parameters to be measured [22], and it would be very interesting to estimate the water quality parameters values with sufficient precision from the other measured parameters, and to implement and adopt these water quality prediction models that can provide a powerful tool to improve the management of the coastal lagoons. The prediction of surface water quality is a basic task in studies of water resource management, to establish the reasons for the deterioration in water quality and to keep pollution within permissible limits [23,24]. The objective of this study is to obtain a predictive model to calculate the concentration of chlorophyll-a (Chl-a) values from other measured parameters in the Mar Menor Lagoon.

The variables involved in estimating Chl-a in water bodies are complex. In the existing literature, different statistical approaches have been used to determining the Chl-a based on regression analyses. However, these traditional data processing methods generally apply a linear relationship to simplify complex problems, leading to unsatisfactory results because they are not efficient enough to cope with the complicated non-linear relationships between the variables involved [25]. Machine learning (ML) algorithms have demonstrated to be more effective than traditional approaches in determining the water quality [26] as they are very well-suited for predicting nonlinear and complex functions. Previous studies have confirmed the superiority or comparability of ML over traditional approaches in modelling water quality parameters [27–29]. ML provides the advantage of performing regressions without the need for a greater knowledge of the water body or the water quality parameters investigated [30]. In particular, Li et al. [31] and Yi et al. [32] applied different types of artificial neural networks (ANNs) to estimate the concentration of Chl-a in 27 lakes in China and in one Korean river, respectively. Another example was presented by Su et al. [25] which developed a structurally simplified hybrid model of the genetic algorithm (GA) and the support vector machine (SVM) for the prediction of monthly concentration of Chl-a in a reservoir of northern China. Nazeer et al. [33] suggested using ML methods,

such as ANNs, for more accurate and efficient routine monitoring of coastal water quality parameters, particularly Chl-a, in a coastal area of Hong Kong. Keller et al. [30] concluded that regression models, such as ANN and SVM were very valuable in estimating five water quality parameters, including Chl-a on the river Elbe in Germany. Considering that the SVM and ANN achieved the best result for different water quality parameters in several studies [26,28–30], it can be expected that these models will obtain satisfactory Chl-a estimation results in this study.

According to superiority of the ML algorithms, this study has been designed to fulfill the following objectives: (1) To select the specific variables that most related to the Chl-a production using wrapper feature selection algorithms in the Mar Menor lagoon; (2) to develop a predictive model to estimate the Chl-a concentration based on multilayer neural network (MLNN) and support vector regression (SVR) models; (3) to validate the performance of predictive models using different evaluation metrics and identify the best method in estimation of the Chl-a concentration for the Mar Menor lagoon. Several studies have been carried out on the eutrophication process and the water quality parameters of the Mar Menor lagoon [17,34–36]. However, to the best of our knowledge, there is no previous research using machine learning models to predict water quality parameters in this lagoon, specifically Chl-a concentration. The rest of the article is organized as follows. Section 2 describes the study area, data collection and the methodology of the study. Section 3 presents the analysis of the data and the results. The discussion of the results is presented in Section 4. Finally, Section 5 summarizes the conclusions.

2. Materials and Methods

2.1. Study Area and Data Collection

The Mar Menor is the largest hypersaline coastal lagoon in Europe located in the Region of Murcia, a semi-arid area of southeastern Spain (Figure 1). It has an area of 135 km² with 73 km of coastline and houses five islands of volcanic origin in its interior that increase the environmental and landscape value of the area. This lagoon is relatively shallow, with a mean depth of 3.6 m and a maximum depth of 7 m, and is isolated from the sea by a 22 km sand coastal barrier (called La Manga) that is crossed by five channels, causing exchanging its waters with the Mediterranean Sea.

Tourism and agriculture along the shoreline of the Mar Menor lagoon are a very important activity for the local economy. The drainage area of the lagoon, known as Campo de Cartagena, is a long plain of more than 1600 km² with non-permanent, but abundant surface watercourses that collect the sparse, but intense rainfall [37]. Campo de Cartagena is characterized by intensive agriculture, and its southern zone was a very active mining region for hundreds of years, although this area is currently abandoned [38].

Traditionally, the Mar Menor has been characterized by oligotrophic waters and by its great resistance and resilience to the eutrophication process. Its main distinctive feature has been the transparency of its waters, but in the last decade, the lagoon has developed eutrophic characteristics [39]. Changes in agricultural practices in the drainage basin, with the introduction of intensively irrigated crops, have increased inputs with high amounts of nutrient to lagoon during the last decades. The results of several studies [40] delineated the flow transfer from the Campo de Cartagena aquifer to the Mar Menor lagoon with the adverse effect of entry of nitrates and other agrochemical elements from fertilization. These inputs caused a quick increase in the pollution in the lagoon [41], and induced a eutrophication process, leading to a loss of water quality [42,43]. An unprecedented eutrophication crisis in 2016, caused by an abrupt increase in the average concentration of nutrients and chlorophyll, generated an evident change in the quality of the waters, with important increase in the turbidity, change of colour of its waters and loss of transparency with a decrease in the depth of visibility of the Secchi disc to less than 1 m. These caused a great alarm both in environmental circles and in the tourism sector with important socioeconomic consequences [15–17]. After considerably reducing the supply of nutrients from agricultural sources, the system began a rapid recovery, which was evident in the spring and summer of 2018 [17]. In September 2019, surface runoff water from the drainage

basin that flowed into the Mar Menor, due to a meteorological phenomenon of the cold drop, caused a strong increase in chlorophyll levels, even above the maximum reached in 2016. The huge amounts of nitrogen, phosphorus, organic matter and sediments washed away by runoff were the main factors for the primary productivity [18].

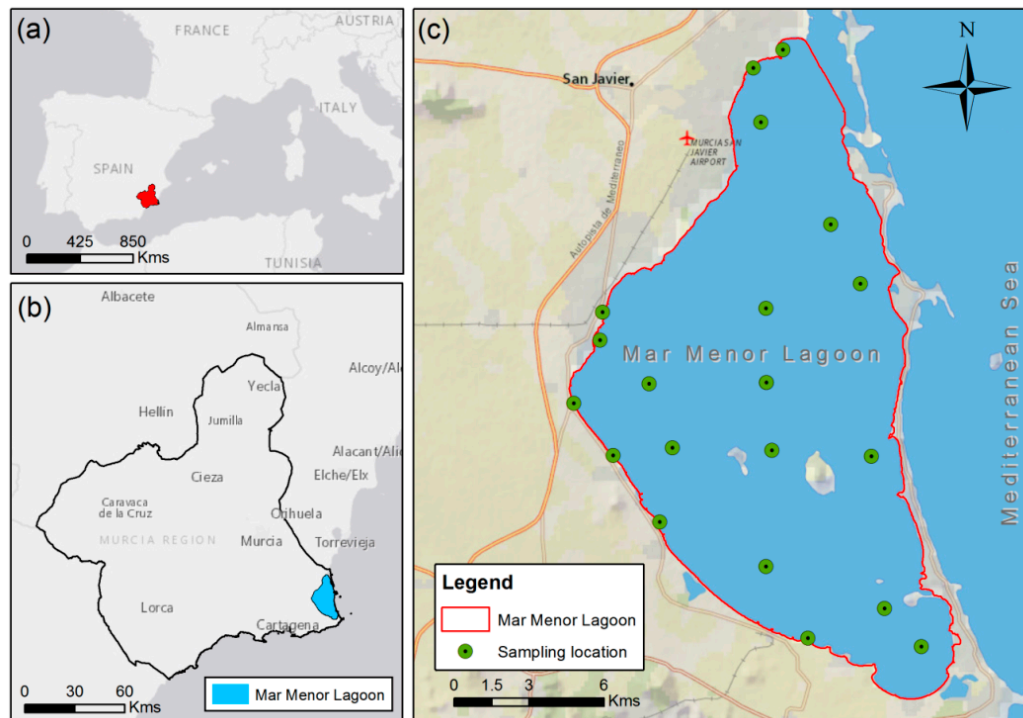


Figure 1. Study site location: (a) Location of the Region of Murcia in Spain; (b) Location of the Mar Menor lagoon in the Region of Murcia; (c) Mar Menor Lagoon and the network of sampling points.

In this context, urgent action is needed to reduce the entry of nutrients and other pollutants into the lagoon. This need is further reinforced by the declaration of the Mar Menor basin as a nitrate pollution vulnerable zone (91/676/EEC), the declaration of the lagoon as a sensitive area subject to eutrophication in application of the Urban Wastewater Directive (91/271/ECC) and the application of the Water Framework Directive (2000/60 EC), which obliges to achieve and maintain the Good Status of all bodies of water and requires the monitoring and management of the ecological status of surface waters by all Member States, including coastal and transitional ones. In addition, the Mar Menor and associated wetlands has been protected by a series of regional national and international rules and resolutions [9,42]: (1) The Ramsar List of Wetlands of International Importance, (2) Special Protected Areas of Mediterranean Interest, (3) Specially Protected Area under the EU Wild Birds Directive, and (4) the Nature 2000 Network as a Site of Community Importance. For these reasons, the conservation of the Mar Menor requires integrated and sustainable planning and management of its basin.

In this study, the daily data on water quality in the Mar Menor were obtained from oceanographic campaigns carried out by the local government, Comunidad Autónoma de la Región de Murcia (CARM), and the information is available on the Mar Menor information service website (<http://www.canalmarmenor.es/web/canalmarmenor/parametros>). The data sets consisted of 10 physical and chemical parameters measured at 20 sampling points throughout the lagoon (Figure 1) from September 2017 to December 2018. Only 126 daily data were available to date. These parameters were Chl-a, water temperature (T), pH, suspended solids (SS), turbidity (TU), Secchi Disk depth (SD), salinity (S), dissolve oxygen (DO), total nitrogen (TN) and total phosphorus (TP).

2.2. Modeling Approaches and Feature Selection (FS)

The different steps followed in this study are as follows: (1) Selection of the input parameters by FS algorithms; (2) building a predictive model (MLNN and SVR) using the selected input scenarios for model learning; and (3) evaluating the predictive models using several metrics to validate the performance.

2.2.1. Multilayer Neural Network (MLNN)

The ANN is a massively parallel-distributed information-processing system that attempts to simulate the functioning of brain neurons using a network of artificial neurons organized into layers [44]. The network receives a stimulus and transforms this input into an output signal through a transfer function. The ANN model is an appropriate technique for modelling because of its capability to assign significance to input parameters and to map the inputs to the outputs when the relationships between the variables of the underlying physical processes are complex or unknown [45]. These neural networks are a non-linear modeling tool that can manage a large number of inputs (independent variables) to determine one or more outputs (dependent variables) [46].

There are many types of ANNs for different applications. In a feedforward network, the direction of information flow between nodes or neurons is from the input to the output layer, and each node in a layer is connected to each of the nodes in the next layer, but not to those in the same layer [47]. The nodes are connected to other nodes by links which have an associated weight that represents its connection strength and stores the knowledge of the network [48]. The mathematical operation of a node can be summarized according to the following Equation (1):

$$y_j = f \left(\sum_{i=1}^n x_i \cdot w_i - b_j \right) \quad (1)$$

where y is the output of a neuron j , f is an activation function, x_i is an input of the vector of inputs ($i = 1, 2, \dots, n$), w_i is the weight associated with the connection link through which the input x_i arrives to current neuron j from a neuron in the preceding layer and b_j is a bias associated with neuron j . Therefore, the connection weights, biases, and transfer functions parameterize the mathematical relationship between inputs and outputs of the network [49]. These weights and biases need to be adjusted in the training process of the networks to minimize the model error.

To estimate the Chl-a level in the Mar Menor, MLNNs were developed based on the feedforward back-propagation method. These networks consist of a number of nodes organized in an input layer, one or more hidden layers and an output layer. In the hidden layer, which is the most important part of the ANN, the nodes receive the signals only from neurons in the previous layer and process data. This processed data was fed to the output layer where the output is calculated.

The script of the MLNN models was implemented in MATLAB® software (version 8.2.0.701 (R2013b), The Mathworks, MA, USA). In this study, one or two hidden layers were considered, with a sigmoid function in the hidden layers and one output layer with a linear function. Specifically, the logistic sigmoid (logsig) and hyperbolic tangent sigmoid (tansig) functions were tested to obtain better results with respect to non-linearity of this process. Figure 2 describes an example of an MLNN used in this study.

The input layer of the neural networks contained as many nodes as there were input parameters and the output layer contained only one node. Numbers of layers (1 or 2) and a number of nodes (between 5 and 40) in the hidden layers were tested and determined using trial and error. All these adjustable parameters were tested to yield a good performance of the network. The mean squared error (MSE) was used to define the network error and minimized during the training process. Four training algorithms, which are the fastest and most commonly adopted in MLNN training [50,51], were tested: Levenberg-Marquardt (LM) backpropagation, BFGS quasi-Newton backpropagation (BFG), resilient backpropagation (RP) and conjugate gradient backpropagation with Fletcher-Reeves updates (CGF).

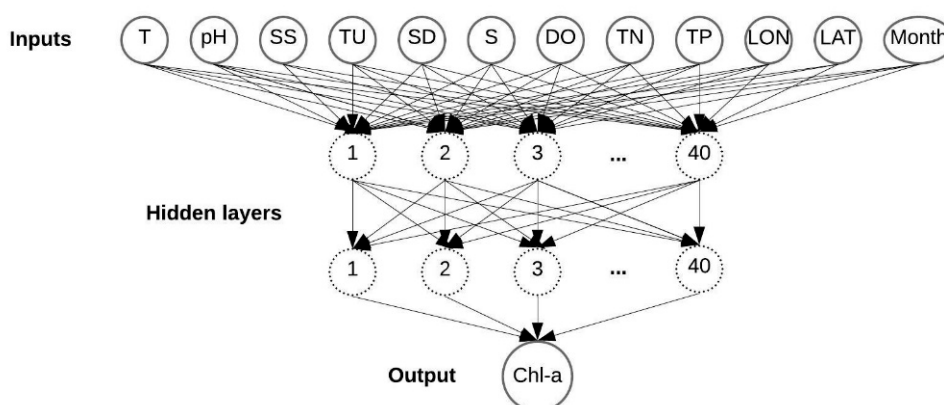


Figure 2. An multilayer neural network (MLNN) structure developed for predicting the concentration of chlorophyll-a (Chl-a) with two hidden layers and 12 variables as inputs (water temperature (T), pH, suspended solids (SS), turbidity (TU), Secchi Disk depth (SD), salinity (S), dissolve oxygen (DO), total nitrogen (TN), total phosphorus (TP), latitude (LAT), longitude (LON) and month).

2.2.2. Support Vector Regression (SVR)

SVM is a supervised machine learning technique that is executed following the structural risk minimization principle and statistical learning theory. The SVM algorithm transforms the original input space into a higher dimensional feature space to find an optimum hyper plane of separation [52]. The SVM model is commonly used in classification problems, but can be easily adopted in regression problems. In fact, the originally developed method was extended into SVR by introducing a ϵ -insensitive loss function for application in regression case studies. The theory of SVR development is available in Vapnik et al. [53]. In this study, the SVR models were developed in the R software and a radial basis function implemented in the “caret” package [54,55] was selected as kernel function and used to estimate the concentration of Chl-a. There are two tuning parameters in this model, the scale function (σ) in the radial basis function (see Equation (2)) and the cost value (C) used to control the complexity of the decision boundary. The application of an adaptive cross-validation resampling technique [56] provides a computationally efficient way to identify these parameters for each specific model of SVR.

$$K(x, x_i) = \exp(-\sigma \|x - x_i\|^2) \quad (2)$$

where x_i is the input vector with $x \in R^n$.

2.2.3. Assessing Model Performance

In an effort to check any overfitting, the 5-fold cross-validation was performed. The data set was divided into five subsets—four subsets were used for training, and the remaining one for validation. The holdout method was repeated five times, and the regression results of 5-fold cross-validation were averaged and presented as overall testing results of the models. The performance of the models was evaluated using three indicators calculated from predicted and measured data. Overall performance was analyzed with cross-validated coefficient of determination (R^2_{CV}), the proportion of systematic error in the overall with cross-validated root mean squared error ($RMSE_{CV}$) and overall errors with cross-validated mean absolute error (MAE_{CV}). These statistics are defined in Table 1.

Table 1. Performance metrics.

Performance Metric	Equation	Range
Cross-validated coefficient of determination (R^2_{CV})	$\frac{[\sum_{i=1}^n (O_i - \bar{O}) \cdot (E_i - \bar{E})]^2}{\left[\left[\sum_{i=1}^n (O_i - \bar{O})^2 \right]^{0.5} \cdot \left[\sum_{i=1}^n (E_i - \bar{E})^2 \right]^{0.5} \right]^2}$	[0, 1]
Cross-validated root mean squared error ($RMSE_{CV}$)	$\sqrt{\frac{\sum_{i=1}^n (O_i - E_i)^2}{n}}$	[0, ∞]
Cross-validated mean absolute error (MAE_{CV})	$\frac{\sum_{i=1}^n O_i - E_i }{n}$	[0, ∞]

O_i is the i th observed data, \bar{O} is the mean of the observed data, E_i is the i th estimated data, \bar{E} is the mean of the estimated data and n is the total number of observations.

2.2.4. Feature Selection (FS)

Machine learning algorithms (MLAs) are generally applied from a set of training instances in which each instance is described by a feature vector (input parameters), and target feature (output parameters) expressed as a continuous value in regression problems [57], where the main objective of predictive modeling is to maximize accuracy [58]. To estimate a parameter of water quality can use all available features, or select a smaller number of them. This can result in the inclusion of too few or too many inputs to the model, both of which are undesirable [59]. To address this issue, an FS stage had been considered in this study to eliminate redundant data. FS is a process that selects a subset of features from the original set, so that the feature space is optimally reduced according to a certain criterion. The goal of reducing the dimensionality of the feature space in ML is to speed-up the operation of the learning algorithm to improve predictive accuracy and enhance the comprehensibility of the learning results [58]. There are many parameters that influence the concentration of Chl-a. This study performed an FS to determine the appropriate input vectors of a prediction model with three hybrid methods: Wrapper algorithms (recursive feature elimination (RFE), GA and simulated annealing (SA)) combined with random forest (RF) and SVR. The wrapper algorithms evaluate multiple models using procedures that add and/or remove predictors to find the optimal combination that maximizes model performance. In essence, wrapper methods are search algorithms that treat predictors as inputs and use model performance as output to optimize. These procedures, as implemented in the “caret” package [55] of the R software. These hybrid methods contained two steps. In the first, RFE, GA or SA is used to select a subset of features from all features. Moreover, in the second step, RF or SVR is used to assign a weight of importance to each feature included in the selected subset.

RFE uses a backwards elimination approach, starting with all features and eliminating one at a time. At each step, the feature that is considered least useful for prediction is removed, and the overall performance of the predictor is reevaluated through cross-validation [60]. GA is based on Darwin’s “survival of the fittest” and was applied to search for the better subset of features. The subset of features was selected based on the current population (i.e., subset of features) through crossover, mutation, and selection according to the fitness function. SA algorithm is a global search method that makes small random changes to an initial candidate solution [55].

3. Results

3.1. Inputs Selection

The basic descriptive statistics of the input variables used in this study are presented in Table 2. The X_{\max} , X_{\min} , X_{mean} , St. Dev., C.V. and correlation coefficient (CC) denote the maximum, minimum, mean, standard deviation, coefficient of variation of the data and correlation coefficient with Chl-a, respectively.

The CC is used to explore the dependences between the variables. Each CC measures the degree of the linear relationship of the variable Chl-a with the other parameters. The CCs marked as bold were significant at p -level < 0.05.

In this study, eight scenarios with different input combinations of the variables are tested for estimating of Chl-a concentration values in the Mar Menor lagoon. As water quality can change spatially and seasonally the variables month, latitude (LAT) and longitude (LON) are added to the set of variables. The first input scenario (M1) considered all parameters as inputs without feature selection. The second scenario (M2) included only the most highly correlated parameters. The other input scenarios (M3–M8) are extracted through the wrappers algorithms. Summarized results of feature selection ranked by importance using wrappers are shown in Table 3.

Table 2. Daily statistics of the water quality parameters between September 2017 and December 2018 in the Mar Menor Lagoon.

Parameters (Units) ¹	X_{\max}	X_{\min}	X_{mean}	St. Dev. ²	C.V. ³	CC ⁴
Chl-a (mg/m ³)	7.50	0.13	2.02	1.43	0.71	1.00
T (°C)	27.70	11.00	20.09	6.41	0.32	−0.08
pH	8.46	7.83	8.17	0.12	0.01	−0.02
SS (mg/l)	35.35	5.00	8.64	4.94	0.57	0.22
TU(NTU)	24.00	0.50	2.82	3.37	1.19	0.002
SD (m)	6.50	0.30	2.39	1.63	0.68	−0.55
S (PSU)	46.38	41.86	44.21	0.95	0.02	−0.35
DO (mg/l)	8.12	4.25	6.55	0.80	0.12	−0.17
TN (mg N/l)	8.84	0.16	0.59	0.80	1.35	0.09
TP (mg P/l)	0.07	0.01	0.01	0.01	0.70	0.25

¹ T is water temperature, SS is suspended solids, TU is turbidity, SD is Secchi Disk depth, S is salinity, DO is dissolve oxygen, TN is total nitrogen and TP is total phosphorus; ² St. Dev.: Standard deviation; ³ C.V.: Coefficient of variation (St. Dev./Xmean); ⁴ CC: Correlation coefficient with Chl-a.

Table 3. Summarized results of feature selection using wrapper algorithms.

Algorithm ¹	N. of Features Selected	Features Selected ²	Input Scenario
Without feature selection	12	T, pH, SS, SD, S, DO, TU, TN, TP, LON, LAT, month	M1
The most highly correlated features	4	SD,S,TP,SS	M2
RFE_RF	9	SD, S, SS,TN, T, pH, DO, TP, LON	M3
RFE_SVM	9	SD, T, SS, S, TP, pH, TN, LON, TU	M4
GA_RF	9	LON, T, pH, SS, TU, SD, S, DO, TP	M5
GA_SVM	10	month, LAT, LON, T, pH, TU, SD, DO, TN, TP	M6
SA_RF	8	LAT, LON, pH, TU, SD, S, DO, TN	M7
SA_SVM	7	LON, T, SS, TU, S, TN, TP	M8

¹ Wrapper algorithms: recursive feature elimination (RFE), genetic algorithm (GA) and simulated annealing (SA) combined with random forest (RF) or support vector regression (SVR). ² LAT is latitude and LON is longitude.

3.2. Model Comparison and Prediction Accuracy

The MLNN and SVR models with eight input scenarios described in Table 3 were developed to simulate the Chl-a concentration. The eight versions of each model represent eight substantially different chlorophyll models, due to the different combinations of variables used as predictors. Different ML modes are compared based on the three statistical indices obtained in 5-fold cross validation: R^2_{CV} , $RMSE_{CV}$ and MAE_{CV} . These performance measurements in the training and testing phase are summarized in Tables 4 and 5 for MLNNs and SVRs, respectively. Different parameters are tried for each MLNN and SVR model and the best ones; i.e., with the minimum $RMSE_{CV}$ in the testing phase, are selected for each input scenario.

The best results are obtained with a network structure of two hidden layers with logsig transfer functions for the eight versions of the MLNN model. The BFG algorithm is the most efficient in all versions except for the MLNN-M5, MLNN-M7 and MLNN-M8 where the RP algorithm defeats them by a slight margin. The comparative results between the eight versions of the MLNN model reveal that the MLNN-M5 with nine inputs selected by GA-RF wrapper algorithm yielded the best accuracy among all the developed MLNN models in term of higher R_{CV} and lower $RMSE_{CV}$ and MAE_{CV} values

for the training ($R_{CV} = 0.74$, $RMSE_{CV} = 0.73 \text{ mg/m}^3$ and $MAE_{CV} = 0.49 \text{ mg/m}^3$) and testing phase ($R_{CV} = 0.62$, $RMSE_{CV} = 0.89 \text{ mg/m}^3$ and $MAE_{CV} = 0.66 \text{ mg/m}^3$). MLNN-M5 is only outperformed by the MLNN-M4 model in terms of lower MAE_{CV} for the testing phase. With also nine inputs, the MLNN-M4 model is the second most accurate model with a performance close to the best.

Table 4. Performance of Chl-a estimation from MLNN models based on eight different input scenarios obtained in 5-fold cross validation.

Model-Input Scenario	Architecture [I-H1-H2-O] ¹	Training Phase			Testing Phase		
		R^2_{CV}	$RMSE_{CV}$ (mg/m ³)	MAE_{CV} (mg/m ³)	R^2_{CV}	$RMSE_{CV}$ (mg/m ³)	MAE_{CV} (mg/m ³)
MLNN-M1	[12-16-27-1]	0.63 ± 0.14	0.85 ± 0.21	0.59 ± 0.21	0.53 ± 0.16	0.98 ± 0.31	0.70 ± 0.22
MLNN-M2	[4-12-17-1]	0.62 ± 0.11	0.88 ± 0.16	0.61 ± 0.09	0.52 ± 0.17	0.95 ± 0.32	0.71 ± 0.17
MLNN-M3	[9-31-23-1]	0.67 ± 0.17	0.80 ± 0.25	0.54 ± 0.15	0.60 ± 0.23	0.89 ± 0.41	0.66 ± 0.24
MLNN-M4	[9-32-39-1]	0.72 ± 0.07	0.76 ± 0.14	0.50 ± 0.06	0.61 ± 0.16	0.89 ± 0.35	0.63 ± 0.18
MLNN-M5	[9-40-39-1]	0.74 ± 0.09	0.73 ± 0.17	0.49 ± 0.13	0.62 ± 0.13	0.89 ± 0.26	0.66 ± 0.14
MLNN-M6	[9-40-33-1]	0.72 ± 0.13	0.74 ± 0.20	0.51 ± 0.20	0.55 ± 0.21	0.99 ± 0.34	0.76 ± 0.19
MLNN-M7	[8-39-16-1]	0.72 ± 0.08	0.78 ± 0.15	0.57 ± 0.13	0.54 ± 0.16	0.96 ± 0.32	0.71 ± 0.18
MLNN-M8	[7-23-26-1]	0.72 ± 0.08	0.74 ± 0.14	0.47 ± 0.10	0.53 ± 0.11	0.98 ± 0.24	0.68 ± 0.13

¹ I is the number of neurons in input layer; H1 and H2 is the number of neurons in hidden layer 1 and hidden layer 2; O is the number of neurons in output layer.

Table 5. Performance of Chl-a estimation from SVR models based on eight different input scenarios obtained in 5-fold cross validation.

Model-Input Scenario	Model Parameters	Training Phase			Testing Phase		
		R^2_{CV}	$RMSE_{CV}$ (mg/m ³)	MAE_{CV} (mg/m ³)	R^2_{CV}	$RMSE_{CV}$ (mg/m ³)	MAE_{CV} (mg/m ³)
SVR-M1	$\sigma = 0.07$ $C = 3.31$	0.53 ± 0.05	0.99 ± 0.07	0.70 ± 0.03	0.56 ± 0.09	0.85 ± 0.26	0.62 ± 0.14
SVR-M2	$\sigma = 0.24$ $C = 0.58$	0.45 ± 0.06	1.12 ± 0.08	0.80 ± 0.03	0.49 ± 0.19	0.98 ± 0.34	0.71 ± 0.16
SVR-M3	$\sigma = 0.13$ $C = 2.54$	0.56 ± 0.05	0.96 ± 0.12	0.66 ± 0.05	0.65 ± 0.11	0.82 ± 0.30	0.58 ± 0.16
SVR-M4	$\sigma = 0.10$ $C = 2.59$	0.58 ± 0.06	0.94 ± 0.10	0.66 ± 0.04	0.68 ± 0.10	0.81 ± 0.32	0.56 ± 0.17
SVR-M5	$\sigma = 0.10$ $C = 3.01$	0.58 ± 0.05	0.92 ± 0.09	0.64 ± 0.04	0.67 ± 0.09	0.82 ± 0.30	0.57 ± 0.18
SVR-M6	$\sigma = 0.10$ $C = 4.20$	0.52 ± 0.09	0.98 ± 0.11	0.69 ± 0.06	0.61 ± 0.16	0.86 ± 0.31	0.61 ± 0.15
SVR-M7	$\sigma = 0.13$ $C = 3.38$	0.51 ± 0.10	1.03 ± 0.15	0.72 ± 0.08	0.57 ± 0.20	0.90 ± 0.37	0.64 ± 0.21
SVR-M8	$\sigma = 0.30$ $C = 2.63$	0.47 ± 0.05	1.06 ± 0.07	0.74 ± 0.04	0.54 ± 0.13	0.94 ± 0.21	0.66 ± 0.13

Regarding the SVR models, the results of parameter optimization of SVR models indicated that the optimal C values are between 0.58 and 4.20, and the optimal σ values in radial basis kernel function are between 0.07 and 0.30. The SVR-M5 model obtained the best results in the training phase ($R_{CV} = 0.58$, $RMSE_{CV} = 0.92 \text{ mg/m}^3$ and $MAE_{CV} = 0.64 \text{ mg/m}^3$) and the SVR-M4 model in the test phase ($R_{CV} = 0.68$, $RMSE_{CV} = 0.81 \text{ mg/m}^3$ and $MAE_{CV} = 0.56 \text{ mg/m}^3$).

Models that used M4 scenario with nine input variables (SD, T, SS, S, TP, pH, TN, LON and TU) and models that used M5 input scenario with also nine inputs (LON, T, pH, SS, TU, SD, S, DO and TP), obtain an accuracy very similar for both models (MLNN and SVR). The difference between these models is that M5 uses the DO variable instead of the TN variable.

4. Discussion

The Chl-a concentrations had a significant correlation with water quality variables as SS, SD, S and TP in the study area, but the weakness of these correlations indicates that the use of traditional regression methods in modeling such a complex process is irrelevant, so there is a great need to use more powerful techniques [61].

The results of the feature selection (see Table 3) are similar to those obtained in other studies. Although the influence variables of Chl-a were different in several research works, the TP and TN concentration generally were among the main variables. For example, Palani et al. [62] applied the ANN model with location variables (LON and LAT), PO₄, DO and T as the explanatory variables to predict Chl-a concentration. Li et al. [31] selected the concentration of PT and TN, T, SD, and DO among the most influential input variables for Chl-a using a genetic algorithm optimized back-propagation neural network. Furthermore, Kuo et al. [63] defined the Chl-a model by the input of month, T, pH, SD, SS, PO₄ and NO₃. Moreover, GA_RF was the algorithm that selected more features with ten of the twelve features selected. In contrast, the algorithm that selected the least characteristics is SA_SVM by selecting only seven. LON is selected by all the algorithms, being the only variable common to all the scenarios. The month factor which shows that the Chl-a changes seasonally is only selected as input in one case.

According to the average performance of the models (Tables 4 and 5), MLNN and SVR models performed reasonably. Considering that the length of the data and its quality have a significant influence on the modelling process [64], the results can be considered acceptable and promising, considering that only 126 daily data are available.

To conclude, the SVR models perform better than the MLNN models for all input scenarios in the testing phase except for the M2 scenario, and the best-fit model is the SVR-M4. In addition, MLNNs have the disadvantage of having to be trained for each problem to obtain the optimal architecture, and this requires great computational resources, greater than for SVR models [65]. The worst results are obtained by the models with the M2, M6, M7 and M8 scenarios. M6, M7 and M8 scenarios did not include as input the variables SS, S, SD or TP that are significantly correlated with Chl-a or other variables, such as T or pH. However, when only the most correlated variables are included (M2 scenario), the models also perform poorly. This fact could support the conclusion drawn by Maier et al. [59] that using a linear approach to identify which of the potential input variables have a significant relationship with the model output is not appropriate for the development of ANN models. Consequently, there is a need to adopt other input selection approaches as FS using algorithms. On the other hand, the FS managed to improve the results with respect to scenario M1 (without the selection of feature) in three scenarios (M3, M4 and M5). The wrapper algorithm that provides the lowest results was the SA. However, FS contribute to reducing the modeling time by reducing the number of input parameters in all cases.

5. Conclusions

The current significant ecological deterioration of the Mar Menor lagoon makes it necessary to improve the understanding of the interaction between the water quality parameters of the coastal lagoon to adopt integrated and sustainable management strategies. This coastal lagoon has recently suffered a eutrophication crisis. Chl-a concentration is one of the most important indicators of the existence and degree of eutrophication in water bodies.

The results confirmed the importance and usefulness of intelligent modeling as a rapid, easy to operate and not expensive tool. The MLNN and SVR models have great potential in modelling complex and heterogeneous systems, such as coastal lagoons. The capabilities of these models in estimation of the concentration of Chl-a have been investigated using water quality inputs, such as T, pH, SS, SD, S, DO, TU, TN and TP monitored at different stations in the Mar Menor lagoon. Both algorithms with eight input combinations and the cross validation method are employed in this study. Comparison between simulated and observed data showed the effectiveness of the models, where most

of the values of $RMSE_{CV}$ and MAE_{CV} are small, and most of R^2_{CV} were close to 0.7 for the best models. Finally, the better model is selected based on the performance measurements. Average performances of the models indicated that the SVR with nine input variables (SD, T, SS, S, TP, pH, TN, LON and TU) is more capable than the alternative models in estimating the values of Chl-a concentration.

The main finding in this study revealed that the MLNN and SVR models could be used as a successful tool to estimate Chl-a concentration based on other measured parameters and these estimation results are useful for the coastal lagoon quality management. This can be especially useful in estimating other parameters, which are difficult to measure, as an alternative way for monitoring and assessment water quality in real time based on other water quality monitoring data from the studied area. Furthermore, MLNN and SVR models can also be used to simulate the future state of the study area using several scenarios for water quality parameters. The limitations of this study include its limited data set. The results obtained in this work could be improved in future work if more on-site measured data were available. Therefore, it is strongly recommended that the sectors related to water quality in the Mar Menor lagoon focus on improving monitorization of environmental variables and sharing the available data. Additionally, other algorithms should be tested in future work.

Author Contributions: Conceptualization, P.J.-S. and J.S.-A.; Formal analysis, P.J.-S.; Investigation, P.J.-S.; Methodology, P.J.-S. and J.S.-A.; Project administration, J.M.C.; Software, P.J.-S.; Supervision, J.M.C. and J.P.-S.; Writing—review and editing, P.J.-S., J.S.-A., J.M.C. and J.P.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Fundación Séneca del Centro de Coordinación de la Investigación de la Región de Murcia under Project 20813/PI/18, and by Spanish Ministry of Science, Innovation and Universities under grants RTI2018-096384-B-I00 and RTC-2017-6389-5.

Acknowledgments: The authors acknowledge the Comunidad Autonoma de la Region de Murcia (CARM) for the data provided through the Mar Menor information service website (Canal Mar Menor).

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

References

1. Barnes, R.S.K. *Coastal Lagoons: The Natural History of a Neglected Habitat*; Cambridge University Press: Cambridge, UK, 1980.
2. Kjerfve, B. Coastal Lagoons. In *Coastal Lagoons Processes*; Kjerfve, B., Ed.; Elsevier Oceanography Series; Elsevier Science Publishers: Amsterdam, The Netherlands, 1994; Volume 60, pp. 1–7.
3. Pérez-Ruzafa, A.; Pérez-Ruzafa, I.M.; Newton, A.; Marcos, C. Coastal lagoons: Environmental variability, ecosystem complexity and goods and services uniformity. In *Coasts and Estuaries, the Future*; Wolanski, E., Day, J., Elliott, M., Ramesh, R., Eds.; Elsevier: New York, NY, USA, 2019; pp. 253–276. [[CrossRef](#)]
4. Kennish, M.J. Coastal lagoons. In *Encyclopedia of Estuaries*; Kennish, M.J., Ed.; Springer: Dordrecht, The Netherlands, 2016. [[CrossRef](#)]
5. Nixon, S.W. Nutrient dynamics, primary production and fisheries yields of lagoons. *Oceanol. Acta* **1982**, *5*, 357–371.
6. Pérez-Ruzafa, A.; Marcos, C. El Mar Menor como motor del cambio de paradigmas en el estudio de las lagunas costeras. In *Mar Menor: Una Laguna Singular y Sensible. Evaluación Científica de su Estado*; Leon, V.M., Bellido, J.M., Eds.; Temas de Oceanografía; Instituto Español de Oceanografía, Ministerio de Economía y Competitividad: Madrid, Spain, 2016; Volume 9, pp. 31–57.
7. García-Ayllón, S. New Strategies to Improve Co-Management in Enclosed Coastal Seas and Wetlands Subjected to Complex Environments: Socio-Economic Analysis Applied to an International Recovery Success Case Study after an Environmental Crisis. *Sustainability* **2019**, *11*, 1039. [[CrossRef](#)]
8. Le Moal, M.; Gascuel-Oudou, C.; Ménesguen, A.; Souchon, Y.; Étrillard, C.; Levain, A.; Moatar, F.; Pannard, A.; Souchu, P.; Lefebvre, A.; et al. Eutrophication: A new wine in an old bottle? *Sci. Total Environ.* **2019**, *651*, 1–11. [[CrossRef](#)] [[PubMed](#)]
9. Alcolea, A.; Contreras, S.; Hunink, J.E.; García-Aróstegui, J.L.; Jiménez-Martínez, J. Hydrogeological modelling for the watershed management of the Mar Menor coastal lagoon (Spain). *Sci. Total Environ.* **2019**, *663*, 901–914. [[CrossRef](#)] [[PubMed](#)]

10. Nixon, S.W. Coastal marine eutrophication: A definition, social causes, and future concerns. *Ophelia* **1995**, *41*, 199–219. [[CrossRef](#)]
11. Huang, J.; Gao, J.; Zhang, Y. Combination of artificial neural network and clustering techniques for predicting phytoplankton biomass of Lake Poyang, China. *Limnology* **2015**, *16*, 179–191. [[CrossRef](#)]
12. Canfield, D.E., Jr. Prediction of chlorophyll a concentrations in Florida lakes: The importance of phosphorus and nitrogen. *J. Am. Water Resour. Assoc.* **1983**, *19*, 255–262. [[CrossRef](#)]
13. Phillips, G.; Pietiläinen, O.P.; Carvalho, L.; Solimini, A.; Lyche Solheim, A.; Cardoso, A.C. Chlorophyll-nutrient relationships of different lake types using a large European dataset. *Aquat. Ecol.* **2008**, *42*, 213–226. [[CrossRef](#)]
14. EL PAÍS. Available online: https://elpais.com/elpais/2019/10/22/inenglish/1571743580_215496.html (accessed on 16 January 2020).
15. García-Ayllón, S. Integrated management in coastal lagoons of highly complexity environments: Resilience comparative analysis for three case-studies. *Ocean Coast. Manag.* **2017**, *143*, 16–25. [[CrossRef](#)]
16. García-Ayllón, S. The Integrated Territorial Investment (ITI) of the Mar Menor as a model for the future in the comprehensive management of enclosed coastal seas. *Ocean Coast. Manag.* **2018**, *166*, 82–97. [[CrossRef](#)]
17. Pérez-Ruzafa, A.; Campillo, S.; Fernández-Palacios, J.M.; García Lacunza, A.; García-Oliva, M.; Ibañez, H.; Navarro-Martínez, P.C.; Pérez-Marcos, M.; Pérez-Ruzafa, I.M.; Quispe-Becerra, J.I. Long term dynamic in nutrients, chlorophyll a and water quality parameters in a coastal lagoon during a process of eutrophication for decades, a sudden break and a relatively rapid recovery. *Front. Mar. Sci.* **2019**, *6*, 1–23. [[CrossRef](#)]
18. Ruiz-Fernandez, J.M.; León, V.M.; Marín-Guirao, L.; Giménez-Casaldueiro, F.; Álvarez-Rogel, J.; Esteve-Selma, M.A.; Gómez-Cerezo, R.; Robledano-Aymerich, F.; González-Barberá, G.; Martínez Fernández, J. Informe de síntesis sobre el estado actual del Mar Menor y sus causas en relación a los contenidos de nutrientes. In *Projects of Sustainability and Conservation of Mar Menor Lagoon and Its Basin*; Universidad de Alicante: Alicante, Spain, 2019; Available online: <https://dcmba.ua.es/es/documentos/carteles-seminarios-doctorado/informe-estado-mar-menor.pdf> (accessed on 25 November 2019).
19. Comunidad Autónoma de la Región de Murcia. *Decreto-Ley n° 1/2017, de 4 de abril, de Medidas Urgentes Para Garantizar la Sostenibilidad Ambiental en el Entorno Del Mar Menor*; Boletín Oficial de la Región de Murcia: Murcia, Spain, 2017. (In Spanish)
20. Comunidad Autónoma de la Región de Murcia. *Ley 1/2018, de 7 de Febrero, de Medidas Urgentes Para Garantizar la Sostenibilidad Ambiental en el Entorno Del Mar Menor*; Boletín Oficial de la Región de Murcia: Murcia, Spain, 2018. (In Spanish)
21. Comunidad Autónoma de la Región de Murcia. *Decreto-Ley n° 2/2019, de 26 de diciembre, de Protección Integral del Mar Menor*; Boletín Oficial de la Región de Murcia: Murcia, Spain, 2019. (In Spanish)
22. Iglesias, C.; Martínez Torres, J.; García Nieto, P.J.; Alonso Fernández, J.R.; Díaz Muñoz, C.; Piñeiro, J.I.; Taboada, J. Turbidity Prediction in a River Basin by Using Artificial Neural Networks: A Case Study in Northern Spain. *Water Resour. Manag.* **2014**, *28*, 319–331. [[CrossRef](#)]
23. Najah, A.; El-Shafie, A.; Kari, O.A.; El-Shafie, A.H. Application of artificial neural networks for water quality prediction. *Neural Comput. Appl.* **2013**, *22* (Suppl. 1), S187–S201. [[CrossRef](#)]
24. Li, X.; Cheng, Z.; Yu, Q.; Bai, Y.; Li, C. Water-quality prediction using multimodal support vector regression: Case study of Jialing River, China. *J. Environ. Eng.* **2017**, *143*, 04017070. [[CrossRef](#)]
25. Su, J.; Wang, X.; Zhao, S.; Chen, B.; Li, C.; Yang, Z. A Structurally Simplified Hybrid Model of Genetic Algorithm and Support Vector Machine for Prediction of Chlorophyll a in Reservoirs. *Water* **2015**, *7*, 1610–1627. [[CrossRef](#)]
26. Abba, S.I.; Hadi, S.J.; Abdullahi, J. River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. *Procedia Comput. Sci.* **2017**, *120*, 75–82. [[CrossRef](#)]
27. Juntunen, P.; Liukkonen, M.; Pelu, M.; Lehtola, M.; Hiltunen, Y. Modelling of Water Quality: An Application to a Water Treatment Process. *Appl. Comput. Intell. Soft Comput.* **2012**. [[CrossRef](#)]
28. Li, X.; Sha, J.; Wang, Z.-L. A comparative study of multiple linear regression, artificial neural network and support vector machine for the prediction of dissolved oxygen. *Hydrol. Res.* **2016**. [[CrossRef](#)]
29. Charulatha, G.; Srinivasalu, S.; Maheswari, O.U. Evaluation of ground water quality contaminants using linear regression and artificial neural network models. *Arab. J. Geosci.* **2017**, *10*, 128. [[CrossRef](#)]

30. Keller, S.; Maier, P.M.; Riese, F.M.; Norra, S.; Holbach, A.; Börsig, N.; Wilhelms, A.; Moldaenke, C.; Zaake, A.; Hinz, S. Hyperspectral Data and Machine Learning for Estimating CDOM, Chlorophyll a, Diatoms, Green Algae and Turbidity. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1881. [[CrossRef](#)]
31. Li, X.; Sha, J.; Wang, Z.-L. Chlorophyll-A Prediction of Lakes with Different Water Quality Patterns in China Based on Hybrid Neural Networks. *Water* **2017**, *9*, 524. [[CrossRef](#)]
32. Yi, H.-S.; Park, S.; An, K.-G.; Kwak, K.-C. Algal Bloom Prediction Using Extreme Learning Machine Models at Artificial Weirs in the Nakdong River, Korea. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2078. [[CrossRef](#)] [[PubMed](#)]
33. Nazeer, M.; Bilal, M.; Alsahli, M.M.M.; Shahzad, M.I.; Waqas, A. Evaluation of Empirical and Machine Learning Algorithms for Estimation of Coastal Water Quality Parameters. *ISPRS Int. J. Geo Inf.* **2017**, *6*, 360. [[CrossRef](#)]
34. Erena, M.; Domínguez, J.A.; Aguado-Giménez, F.; Soria, J.; García-Galiano, S. Monitoring Coastal Lagoon Water Quality through Remote Sensing: The Mar Menor as a Case Study. *Water* **2019**, *11*, 1468. [[CrossRef](#)]
35. García-Oliva, M.; Marcos, C.; Umgiesser, G.; McKiver, W.; Ghezzi, M.; De Pascalis, F.; Pérez-Ruzafa, A. Modelling the impact of dredging inlets on the salinity and temperature regimes in coastal lagoons. *Ocean Coast. Manag.* **2019**, *180*, 104913. [[CrossRef](#)]
36. López-Ballesteros, A.; Senent-Aparicio, J.; Srinivasan, R.; Pérez-Sánchez, J. Assessing the Impact of Best Management Practices in a Highly Anthropogenic and Ungauged Watershed Using the SWAT Model: A Case Study in the El Beal Watershed (Southeast Spain). *Agronomy* **2019**, *9*, 576. [[CrossRef](#)]
37. Senent-Aparicio, J.; Pérez-Sánchez, J.; García-Aróstegui, J.L.; Bielsa-Artero, A.; Domingo-Pinillos, J.C. Evaluating Groundwater Management Sustainability under Limited Data Availability in Semiarid Zones. *Water* **2015**, *7*, 4305–4322. [[CrossRef](#)]
38. Navarro, M.C.; Pérez-Sirvent, C.; Martínez-Sánchez, M.J.; Vidal, J.; Tovar, P.J.; Bech, J. Abandoned mine sites as a source of contamination by heavy metals: A case study in a semi-arid zone. *J. Geochem. Explor.* **2008**, *96*, 183–193. [[CrossRef](#)]
39. Conesa, H.M.; Jiménez-Cárceles, F.J. The Mar Menor lagoon (SE Spain): A singular natural ecosystem threatened by human activities. *Mar. Pollut. Bull.* **2007**, *54*, 839–849. [[CrossRef](#)]
40. Domingo-Pinillos, J.C.; Senent-Aparicio, J.; García-Aróstegui, J.L.; Baudron, P. Long Term Hydrodynamic Effects in a Semi-Arid Mediterranean Multilayer Aquifer: Campo de Cartagena in South-Eastern Spain. *Water* **2018**, *10*, 1320. [[CrossRef](#)]
41. Stefanova, A.; Hesse, C.; Krysanova, V. Combined Impacts of Medium Term Socio-Economic Changes and Climate Change on Water Resources in a Managed Mediterranean Catchment. *Water* **2015**, *7*, 1538–1567. [[CrossRef](#)]
42. Velasco, J.; Lloret, J.; Millán, A.; Marín, A.; Barahona, J.; Abellán, P.; Sánchez-Fernández, D. Nutrient and particulate inputs into the Mar Menor lagoon (SE Spain) from an intensive agricultural watershed. *Water Air Soil Pollut.* **2006**, *176*, 37–56. [[CrossRef](#)]
43. García-Oliva, M.; Pérez-Ruzafa, Á.; Umgiesser, G.; McKiver, W.; Ghezzi, M.; De Pascalis, F.; Marcos, C. Assessing the Hydrodynamic Response of the Mar Menor Lagoon to Dredging Inlets Interventions through Numerical Modelling. *Water* **2018**, *10*, 959. [[CrossRef](#)]
44. Haykin, S. *Neural Networks, a Comprehensive Foundation*, 2nd ed.; Prentice Hall: Old Tappan, NJ, USA, 1999.
45. Wei, B.; Sugiura, N.; Maekawa, T. Use of artificial neural network in the prediction of algal blooms. *Water Res.* **2001**, *35*, 2022–2028. [[CrossRef](#)]
46. Fogelman, S.; Zhao, H.; Blumenstein, M.; Zhang, S. Estimation of oxygen demand levels using UV-Vis spectroscopy and artificial neural networks as an effective tool for real-time, wastewater treatment control. In Proceedings of the 1st Australian Young Water Professionals Conference, Sydney, Australia, 15–17 February 2006.
47. ASCE Task Committee. Artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Eng.* **2000**, *5*, 115–123. [[CrossRef](#)]
48. Jimeno-Sáez, P.; Senent-Aparicio, J.; Pérez-Sánchez, J.; Pulido-Velazquez, D. A Comparison of SWAT and ANN Models for Daily Runoff Simulation in Different Climatic Zones of Peninsular Spain. *Water* **2018**, *10*, 192. [[CrossRef](#)]
49. Nguyen, V.D.; Tan, R.R.; Brondial, Y.; Fuchino, T. Prediction of vapor-liquid equilibrium data for ternary systems using artificial neural networks. *Fluid Phase Equilibria* **2007**, *254*, 188–197. [[CrossRef](#)]

50. Bekkari, N.; Zeddouri, A. Using artificial neural network for predicting and controlling the effluent chemical oxygen demand in wastewater treatment plant. *Manag. Environ. Qual.* **2019**, *30*, 593–608. [CrossRef]
51. Zhang, Y.; Gao, X.; Smith, K.; Inial, G.; Liu, S.; Conil, L.B.; Pan, B. Integrating water quality and operation into prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network. *Water Res.* **2019**, *164*, 114888. [CrossRef]
52. Naghibi, S.A.; Ahmadi, K.; Daneshi, A. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resour. Manag.* **2017**, *31*, 2761–2775. [CrossRef]
53. Vapnik, V.; Golowich, S.; Smola, A. Support vector method for function approximation, regression estimation, and signal processing. In *Advances in Neural Information Processing Systems*; Mozer, M.C., Jordan, M.I., Petsche, T., Eds.; MIT Press: Cambridge, MA, USA, 1997; pp. 281–287.
54. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **2008**, *28*, 1–26. [CrossRef]
55. Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Benesty, M.; Lescarbeau, R.; et al. Caret: Classification and Regression Training, R Package Version 6.0-84. 2019. Available online: <https://CRAN.R-project.org/package=caret> (accessed on 12 December 2019).
56. Kuhn, M. Futility analysis in the cross-validation of machine learning models. *arXiv* **2014**, arXiv:1405.6974.
57. Kohavi, R.; John, G.H. The wrapper approach. In *Feature Extraction, Construction and Selection: A Data Mining Perspective*; Liu, H., Motoda, H., Eds.; Springer: Boston, MA, USA, 1998; Volume 453, pp. 33–50.
58. Motoda, H.; Liu, H. Feature selection, extraction and construction. Towards the Foundation of Data Mining Workshop. In Proceedings of the Sixth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'02), Taipei, Taiwan, 6–8 May 2002; pp. 67–72.
59. Maier, H.R.; Jain, A.; Dandy, G.C.; Sudheer, K. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Model. Softw.* **2010**, *25*, 891–909. [CrossRef]
60. Kumar, S.; Bucher, P. Predicting transcription factor site occupancy using DNA sequence intrinsic and cell-type specific chromatin features. *BMC Bioinform.* **2016**, *17* (Suppl. 1), S4. [CrossRef] [PubMed]
61. Mjalli, F.S.; Al-Asheh, S.; Alfadala, H.E. Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. *J. Environ. Manag.* **2007**, *83*, 329–338. [CrossRef]
62. Palani, S.; Liong, S.Y.; Tkalich, P. An ANN application for water quality forecasting. *Mar. Pollut. Bull.* **2008**, *56*, 1586–1597. [CrossRef] [PubMed]
63. Kuo, J.T.; Hsieh, M.H.; Lung, W.S.; She, N. Using Artificial Neural Network for reservoir eutrophication prediction. *Ecol. Model.* **2007**, *200*, 171–177. [CrossRef]
64. Khadr, M. Modeling of Water Quality Parameters in Manzala Lake Using Adaptive Neuro-Fuzzy Inference System and Stochastic Models. In *Egyptian Coastal Lakes and Wetlands: Part II—Climate Change and Biodiversity*; Negm, A., Bek, M., Abdel-Fattah, S., Eds.; The Handbook of Environmental Chemistry; Springer: Cham, Switzerland, 2017; p. 72.
65. Jimeno-Sáez, P.; Senent-Aparicio, J.; Pérez-Sánchez, J.; Pulido-Velazquez, D.; Cecilia, J.M. Estimation of Instantaneous Peak Flow Using Machine-Learning Models and Empirical Formula in Peninsular Spain. *Water* **2017**, *9*, 347. [CrossRef]

