Knowledge-Based Systems 257 (2022) 109895

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Forecasting biotoxin contamination in mussels across production areas of the Portuguese coast with Artificial Neural Networks

Rafaela C. Cruz [a], Pedro R. Costa [b,c,d], Ludwig Krippahl [a,e], Marta B. Lopes [a,e,f,*]

[a] NOVA School of Science and Technology (FCT NOVA), Caparica, Portugal
[b] Instituto Português do Mar e da Atmosfera (IPMA), Lisboa, Portugal
[c] Centro de Ciências do Mar, Universidade do Algarve (CCMAR), Faro, Portugal
[d] S2AQUA - Collaborative Laboratory, Association for a Sustainable and Smart Aquaculture, Olhão, Portugal
[e] NOVA Laboratory for Computer Science and Informatics (NOVA LINCS), FCT NOVA, Caparica, Portugal
[f] Center for Mathematics and Applications (NOVA MATH), FCT NOVA, Caparica, Portugal

ARTICLE INFO

Article history:
Received 13 April 2022
Received in revised form 10 September 2022
Accepted 12 September 2022
Available online 17 September 2022

Keywords:
Time series
Forecasting
Artificial Neural Networks
Biotoxins
Shellfish contamination
Harmful algal blooms

ABSTRACT

Harmful algal blooms (HABs) and the consequent contamination of shellfish are complex processes depending on several biotic and abiotic variables, turning prediction of shellfish contamination into a challenging task. Not only the information of interest is dispersed among multiple sources, but also the complex temporal relationships between the time-series variables require advanced machine methods to model such relationships. In this study, multiple time-series variables measured in Portuguese shellfish production areas were used to forecast shellfish contamination by diarrhetic shellfish poisoning (DSP) toxins one to four weeks in advance. These time series included DSP concentration in mussels (*Mytilus galloprovincialis*), toxic phytoplankton cell counts, meteorological, and remotely sensed oceanographic variables. Several data pre-processing and feature engineering methods were tested, as well as multiple autoregressive and artificial neural network (ANN) models. The best results regarding the mean absolute error of prediction were obtained for a bivariate long short-term memory (LSTM) neural network based on biotoxin and toxic phytoplankton measurements, with higher accuracy for short-term forecasting horizons. When evaluating all ANNs model ability to predict the contamination state (below or above the regulatory limit for contamination) and changes to this state, multilayer perceptrons (MLP) and convolutional neural networks (CNN) yielded improved predictive performance on a case-by-case basis. These results show the possibility of extracting relevant information from time-series data from multiple sources which are predictive of DSP contamination in mussels, therefore placing ANNs as good candidate models to assist the production sector in anticipating harvesting interdictions and mitigating economic losses.

1. Introduction

Shellfish farming is a sustainable solution to the growing demand for seafood products and for nutritious and healthy food, being a good protein source to feed an increasing world population. With a small ecological footprint and requiring neither fresh water nor artificial feed, shellfish farming also contributes to eutrophication management by active nutrient extraction [1,2], since shellfish are filter-feeding organisms taking their energy to grow from small particles, such as microalgae, available in the seawater column. However, of the thousands of microalgae species present in the oceans, estuaries, and coastal lagoons, some produce toxins that can accumulate in shellfish and become a threat to human health [3]. Symptoms of acute intoxication due to consumption of contaminated shellfish can be serious, including diarrhea, amnesia or paralysis [3], making it necessary to monitor contamination in shellfish-producing areas. To promote seafood safety, most coastal countries conduct a monitoring program of their shellfish-producing areas. A survey of toxic algae abundance in seawater and determination of the algae toxins, also known as biotoxins, in representative shellfish samples is carried out on a weekly basis. When biotoxin concentration exceeds the legal safety limit, harvesting and commercialization of shellfish are prohibited. This prohibition may last weeks or months since algal bloom intensity and shellfish toxin kinetics and metabolism affect how long contamination remains high [4]. This reactive approach of mandatory closures in case of contamination, established in the European Union (EU) directives [5,6], makes farmers susceptible to severe economic losses. Hence the interest in proactive strategies for predicting shellfish contamination from environmental changes associated with harmful algal

* Corresponding author.
E-mail address: marta.lopes@fct.unl.pt (M.B. Lopes).

https://doi.org/10.1016/j.knosys.2022.109895

blooms (HABs) formation and the subsequent contamination of shellfish (e.g., [7–10]).

HABs formation is a natural process induced and regulated by atmospheric (light, wind, precipitation), oceanographic (seawater temperature, salinity, currents), chemical (nutrients, metal elements) and biological drivers (grazers, prey for mixotrophic species) [11–15], which stand, alongside biological parameters, as natural candidates for predictors of HABs and subsequent biotoxin shellfish contamination. As reviewed in [16], HABs forecasting based on available time-series data has been attempted with autoregressive models, such as the autoregressive integrated moving average (ARIMA) model and the vector autoregressive (VAR) model (e.g., [17,18]). However, autoregressive models are not able to capture non-linear relationships among variables, which is important because relations between biological and environmental variables in organisms and ecosystems are often complex and highly non-linear [19]. Hence, artificial neural networks (ANNs), namely, multilayer perceptrons (MLP), which are fully connected feed forward networks, have been used for these purposes (e.g., [20–22]). However, due to limitations such as difficulty in tuning the parameters, slow convergence speed and requiring a significant amount of training data to obtain a stable forecasting result [23], these networks have been increasingly replaced by convolutional neural networks (CNNs) and recurrent neural networks (RNNs), which are better able to model time series.

Hill et al. (2020) [24] used a CNN to extract feature vectors from several remotely sensed variables, including reflectance bands, chlorophyll-*a* (chl-*a*) concentrations, and sea surface temperature (SST). These vectors were then input to other machine learning models to detect and predict HAB events in the Gulf of Mexico. Recent works have also used RNNs to forecast chl-*a* concentration. Lee and Lee (2018) [25] proposed an RNN and a long short-term memory (LSTM), a variant of standard RNNs, to predict one-week ahead chl-*a* concentration on four major rivers in South Korea. Cho et al. (2018) [26] and Cho and Park (2019) [27] used data from multiple sources and a deep recurrent network composed of LSTM units to forecast chl-*a* concentration. These models achieved high predictive performance, outperforming MLP with different numbers of hidden layers. Also, Yussof et al. (2021) [28] trained an LSTM and a CNN with remotely sensed chl-*a* time-series data to forecast HABs eight days in advance in the West Coast of Sabah, Malaysia. The results showed that the CNN was outperformed by the LSTM, which had the ability to handle long-term dependency data.

ANNs have extensively been used in a wider scope beyond HABs to model and forecast water quality (e.g., [29–32]). Recent efforts encompassing the development of deep neural networks, namely deep LSTM models for time-series forecasting, have also been made for water quality prediction [33–35]. However, while several studies have attempted to forecast water quality and HABs, very few works dedicated to forecasting shellfish contamination, a natural consequence of HABs. Grasso et al. (2019) [36] predicted closures to shellfish harvesting areas due to paralytic shellfish poisoning (PSP) toxins in blue mussels one to ten weeks in advance. The authors used four years of weekly toxin data and created four classification categories based on toxin concentrations, and used an MLP with one hidden layer to perform the classification task. The model was able to predict closure-level toxic events at a two-week advance notice with high accuracy.

Similarly, Harley et al. (2020) [37] used several environmental time series, such as salinity, air temperature, and SST, to predict the concentration of PSP toxins in blue mussels. The authors created a random forest (RF) model to classify shellfish above and below a toxicity threshold one week in advance. However, the obtained forecast accuracies were very low (below 50%).

This study addressed the challenge of integrating several biological and environmental time-series variables involved in HABs' formation and shellfish contamination, available from multiple sources, and modeling the complex dependencies between them. The goal was to predict when the concentration of biotoxins in bivalve molluscs exceeds the safety limits, leading to the closure of affected production areas. For that, multiple autoregressive and ANNs forecasting models based on multivariate time-series data from shellfish-producing areas along the Portuguese coast were comprehensively explored to predict biotoxin concentration in mussels up to four weeks in advance. The ability to forecast interdictions to shellfish harvesting and production represents an invaluable contribution to decision-making, in particular, in the definition of proper actions to be taken upon closures regarding production management, stock distribution and storage, thus reducing waste and economic losses.

2. Methods

In this work, multiple in-situ, remotely sensed and climate time series were used to forecast biotoxins concentration in shellfish in various producing areas of Portugal. The overall methodology, from sample acquisition to time-series analysis and forecasting can be found in Fig. 1, with a detailed description of each step provided next.

2.1. Study area and data acquisition

This study was focused on predicting the concentrations of diarrhetic shellfish poison (DSP) toxins, which is the type of toxins that leads to most contamination cases in several European countries, including Portugal [38,39]. Eight shellfish production areas were chosen, corresponding to known areas regarding the frequency and extension of the contamination events, showing distinct biotic and abiotic characteristics, and at the same time containing the most complete time-series data. These include Ria de Aveiro (RIAV1), Litoral Cabo Raso – Lagoa de Albufeira (L5b), Estuário do Tejo (ETJ1), Lagoa de Albufeira (LAL), Litoral São Vicente – Lagos (L7c1), Ria de Alvor (LAG, POR2), and Ria Formosa (FAR1).

The model organism chosen for DSP forecasting was the shellfish species *Mytilus galloprovincialis*, a mussel species that is usually used as an indicator species for shellfish contamination. The weekly concentration of DSP toxins in *M. galloprovincialis* from 2015 to 2020 was obtained from the Portuguese Institute for Sea and Atmosphere (IPMA) website,[1] which runs a monitoring system for shellfish contamination that collects shellfish samples from 41 shellfish production areas along the Portuguese coast. For each shellfish production area, IPMA also determines toxic phytoplankton cell counts in seawater samples collected along with mussels specimens, and these data were also used in this study.

During preprocessing, non-numerical codes (i.e., "non detected" and "non quantifiable") were replaced by the detection limit of the methods used to determine biotoxin concentration and phytoplankton cell counts, and errors resulting from the incorrect insertion of the samples were corrected.

Besides biotoxins and phytoplankton *in-situ* measurements, relevant meteorological and oceanographic variables, as previously identified as HAB drivers for the Portuguese coast [40,41], were also collected and used in this study. The meteorological data were obtained from IPMA, including mean, maximum and minimum air temperature, mean wind direction, mean wind intensity, and rainfall from several meteorological stations along

[1] https://www.ipma.pt/

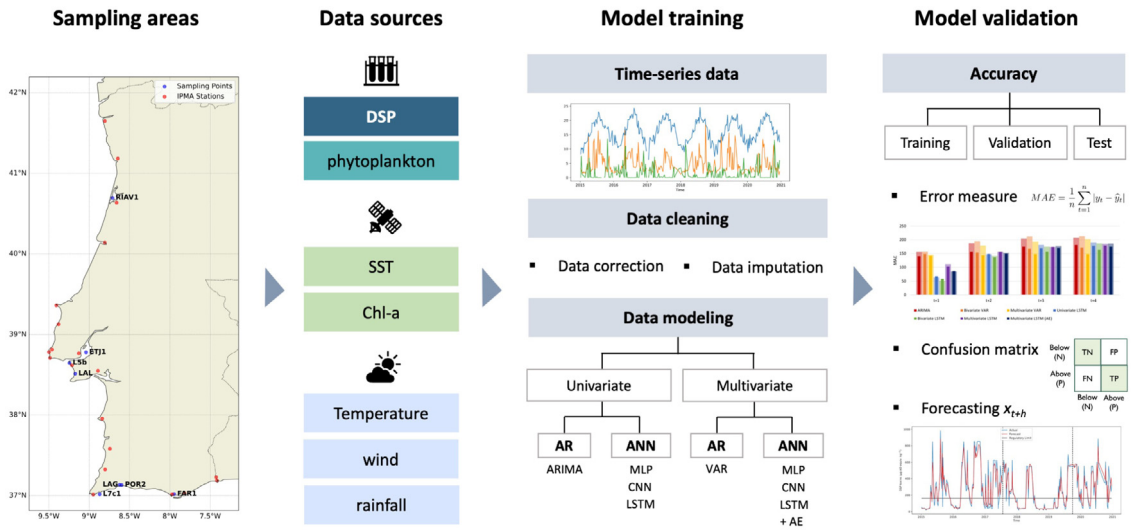

Fig. 1. Overall methodology for predicting mussel contamination by DSP toxins across Portuguese shellfish production areas (AR, autoregressive; ARIMA, autoregressive integrated moving average; VAR, vector autoregressive; ANN, artificial neural network; MLP, multilayer perceptron; CNN, convolutional neural network; LSTM, long short-term memory; AE, autoencoder).

the Portuguese coast (Fig. 1). For each production area, the online meteorological station closest to the sampling point in that area was chosen and the weekly means of the meteorological data for each week of the year was computed.

Finally, remotely sensed SST and chl-*a* time-series data for the considered production zones were collected from the Copernicus Marine Environment Monitoring Service (CMEMS) website,[2] which is part of the European Union's Earth Observation Programme [42]. The data were provided on a network common data form (netCDF) format, which contained, for each time step and latitude/longitude grid point, SST and chl-*a* concentration values. The coordinates for the sampling point in each production area were then used to choose the closest grid point and daily time series for SST and chl-*a* concentration were obtained for that point. The weekly means of these variables were used for each corresponding week of the year.

After merging all the variables, a dataset containing 11 weekly time series was obtained for each production area. The description of these variables can be found in Tables S1 and S2 for the RIAV1 and L5b production areas, respectively.

2.2. Time-series forecasting

A time series is a set of values measured sequentially through time. Time series forecasting aims at predicting one or more future values, $\{x_{T+h}, x_{T+h+1}, \ldots\}$, from an observed time series $\{x_1, x_2, \ldots, x_T\}$, where h is an integer called the lead time or the forecasting horizon [43]. Forecasting can be univariate, if predictions depend only on the present and past values of the variable being forecast, or multivariate if forecasts of a given variable depend not only on its own past values but also on past values of one or more additional variables [43].

2.2.1. Autoregressive models

The first forecasting models used to predict shellfish contamination were autoregressive models, namely ARIMA and VAR. ARIMA is a univariate model that combines an autoregressive model (AR) with a moving average model (MA). An AR(p) model is a linear model in which the current value of a variable is obtained through a weighted sum of its past p values plus a random shock, ε [43,44]. While an AR(p) model captures the autocorrelations of the series values up to lag p, an MA(q) model captures the autocorrelations of the random shocks up to lag q [45]. Thus, in an ARIMA(p, d, q) model, the current value of a time series is given by

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (1)$$

Since ARIMA presumes the series to be stationary, it includes a differencing step to remove trends. This consists in converting the series of values into the differences between consecutive values and can be done several times. The order of differencing in an ARIMA(p, d, q) model, d, indicates how many rounds of differencing are performed.

The VAR model is an extension of the AR model to the multivariate setting. A VAR model is an n-equation, n-variable linear model in which each variable is explained by its own lagged values, plus current and past values of the remaining $n - 1$ variables [44]. In a VAR model of order p, the current values of a vector $\mathbf{X}_t = [X_{1,t}, X_{2,t}, \ldots, X_{n,t}]$ are given by

$$\boldsymbol{X}_t = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} (\boldsymbol{\phi}_i \boldsymbol{X}_{t-i}) + \boldsymbol{a}_t, \quad (2)$$

where $\boldsymbol{\phi}_0$ is a n-dimensional constant vector, $\boldsymbol{\phi}_i$ are $n \times n$ matrices for $i > 0$, and $\boldsymbol{a}_t$ is a sequence of independent and identically distributed random vectors with mean zero and time invariant covariance matrix Σ_a [46]. Both ARIMA and VAR are very simple and popular models that can be used as a baseline to which more complex models, namely ANNs, can be compared.

ARIMA and VAR models were built for RIAV1 and L5b, two distinct (estuarine and coastal, respectively) and problematic production areas, with the highest number of contamination cases registered from 2015 to 2020. The model building process was similar for both production areas and will only be described for RIAV1.

As a univariate method, ARIMA was only applied to the DSP toxins variable. This variable contained a small percentage of missing values (Table S1), which were replaced using linear interpolation, a simple imputation method that creates a straight line between the gaps created by the missing values.

The data was then split into a training set, which was used to train the ARIMA model, a validation set for model tuning

[2] https://marine.copernicus.eu/

and selection, and a test set for final evaluation. The first 130 weeks of the time series, corresponding to two and a half years of data, were used as the training set, and the last 52 weeks, corresponding to one year of data, were used as the test set. The remaining data, corresponding to the period between the training and test sets, was used as the validation set.

The order (p, d, q) of the model were then determined. Since the augmented Dickey–Fuller (ADF) test showed the time series to be stationary (with a p-value of approximately 0) there was no need for a differencing step, and thus $d = 0$. The parameters (p, q) were determined by finding the optimal order for an ARIMA model based on the training set using the Akaike information criterion (AIC). Several combinations of p and q, in which each parameter was varied between 0 and 5, were tested. The combination that allowed to obtain the lowest AIC value was $p = 1$ and $q = 1$. The selected model for the RIAV1 production area was an ARIMA(1, 0, 1). The model for the L5b area was selected in a similar way, resulting in an ARIMA(2, 0, 0).

As a multivariate method, VAR can be applied to more than one variable. A first bivariate VAR was trained with the DSP toxins concentrations and phytoplankton cell count variables, since the DSP-producing phytoplankton are the direct cause of shellfish contamination by DSP toxins. The procedure for imputing missing values, testing for stationarity and splitting data was the same as described for the ARIMA model. The order p was determined based on the AIC, Schwarz–Bayesian (BIC), and Hannan–Quinn (HQ) selection criteria, which are commonly used information criteria across many application domains [46–48], with $p = 1$ showing as the best order for both RIAV1 and L5b production areas.

A VAR model was also trained on all available variables. Missing values were imputed as described above for most variables. However, wind variables contained many missing values, with large time gaps without measurements (Table S1). For these, variables containing more than 25% missing values were not considered when building the VAR model, reserving the linear imputation method for those with less than 25% missing values. Thus, in RIAV1, wind intensity and direction were discarded, and in L5b wind direction was discarded.

To determine which variables to include in the VAR model based on their predictive value to the DSP toxins variable, the Granger causality test [49] was used. This statistical method tests the null hypothesis that the past values of a time series variable $\{X_{1,t}\}$ do not provide significant information about the future values of another time series variable $\{X_{2,t}\}$ [50]. For the L5b production area the phytoplankton cell counts and the wind intensity variables were not found to be predictive of the DSP toxins variable and were discarded. In all other cases the p-values for the Granger causality test were lower than 0.05, so all other variables were retained. The data was split into training, validation and test sets and the p was selected based on the AIC, BIC, and HQ selection criteria as described for the bivariate VAR model. The values obtained were $p = 1$, for both RIAV1 and L5b production areas.

2.2.2. Artificial Neural Networks

An Artificial Neural Network (ANN) is a model composed of many simple non-linear computing units, called neurons, and can generalize from examples to produce useful solutions. ANNs are good candidates for HABs and shellfish contamination forecasting because they are universal approximators, meaning that, in theory, they can approximate any relation between independent and dependent variables [51]. With enough training data, ANNs can generalize and transfer the learned relations to new data [51].

The multi-layer perceptron (MLP) is the simplest example of an ANN architecture. In a MLP, the neurons are organized in layers, with weighted links from each neuron in the kth layer being directed to each neuron in the $(k + 1)$th layer [52]. A MLP is usually composed of one input layer, which receives inputs from the environment, one output layer that yields the network's response, and one or more intermediate hidden layers [53].

A convolutional neural network (CNN) is a type of neural network specialized for processing data with a grid-like topology, such as images and time series [54]. In each hidden layer of a CNN, a weight matrix (also called kernel or filter) slides over the input and computes the dot product between the input and the weight matrix to create a feature map. The intuition behind a CNN is thus to learn in each layer a weight matrix that will be able to extract useful features from the input [55,56].

Recurrent neural networks (RNNs) are specialized for processing sequential data, such as time series [54]. This processing is achieved by maintaining an internal memory state which acts as a compact summary of past information and is recursively updated with new observations at each time step [57]. This is made possible by the existence of feedback loops in an RNN unit, which connect its previous state to the current state, allowing for information about the past to be transmitted to the present.

One disadvantage of using RNNs is their weakness in managing long-term dependencies due to the vanishing gradient problem. To solve this problem, a variant of standard RNNs, long short-term memory (LSTM) [58], has been proposed. An LSTM unit has two components to its state: the hidden state and the cell state. At each time step t, the LSTM unit is presented with the corresponding data element in the input sequence and, in addition, the hidden state and cell state at the previous time step $t-1$. As in a generic RNN, the hidden state results from non-linear transformations of the LSTM unit input. However, since the cell state is linear function of previous states, it preserves gradient information better and can learn longer patterns in the data, making it useful when prediction requires considering longer time windows [51].

Several ANN architectures were applied to DSP toxins concentration forecasting in mussels, namely, MLP, CNN, and LSTM, three ANN architectures which are applicable to the problem under study. A first pre-processing task was undertaken, which involved multiple steps, as described next.

Data preparation. Since ANNs need to have a large training set in order to learn a desired function from the data, the data from all the acquired production areas (RIAV1, L5b, ETJ1, LAL, L7c1, LAG, POR2, and FAR1) were first merged. To guarantee that all variables, each showing different units and ranges (Tables S1 and S2), have the same weight in the final result, the variables were normalized to a scale of [0, 1].

Two different approaches were then considered to deal with the missing values. For the variables where the percentage of missing values was less than 25%, these values were replaced using a linear interpolation method, as for the autoregressive models. For the cases where the percentage of missing values was superior to 25%, these values were replaced by -1, since the data had previously been scaled to [0, 1], and, therefore, the networks should be able to identify these values and not use them to make predictions.

To predict one step in the future, the observations at previous time steps (i.e., lag observations) are used as input X, and the output y is the observation at the current time step [59]. For a MLP, each batch of examples consists of a two-dimensional matrix of [samples, time steps $\times$ features] as input into the network, to predict the future value, since all feature values in the time window are presented as a flattened vector for each example in the batch. In the case of CNNs and LSTMs, the input must be three-dimensional, with a [samples, time steps, features] data shape since the time will be a different dimension for these

networks. The features represent the number of variables used to train the model, which can be 1, in the case of univariate forecasting, or more, in the case of multivariate forecasting.

In this study, the goal was to predict the future values of the DSP toxins variable, using the past values of the DSP toxins variable and the remaining ten explanatory variables described in Tables S1 and S2. In different cases the number of features used could be one or two, for univariate and bivariate models, 11 for models using the complete data, or 5 for models using a reduced representation of the meteorological data obtained with an autoencoder (see below). Since the objective was to predict the concentrations of the toxins up to four weeks ahead, the shape of y had to be [samples, 4, 1]. In the case of X, the time steps refer to the number of past weeks used as input to the models. Multiple time steps were tested, with 12 being the value that worked best for the different ANN models. For the case of the univariate models, instead of 11 features only 1 was used, the past DSP values.

After transforming the data into X and y, these sets were split into training, validation and test sets. Like for the autoregressive models, for each production area and variable, the first two and a half years of data were used as the training set, the following one and a half years as the validation set, and the remaining one year as the test set.

Network building. Three types of ANNs were implemented, namely, MLP, CNN, and LSTM. The networks were built using Keras,[3] a deep learning API written in Python, running on top of the machine learning platform TensorFlow. The MLP models were built using one to three hidden layers with ReLU as the activation function, and a final layer with four neurons and linear activation. This final layer outputs a vector with four elements that can be interpreted as the forecasts up to four weeks in advance.

The CNN models were implemented with two to four convolutional hidden layers. The kernel size was set to 2 for all the convolutional layers. Furthermore, the padding was set to "causal" to create a causal convolution network, as described in Borovykh et al. (2018) [55]. Each hidden layer used the LeakyReLU activation function. A `Flatten` layer was then used to flatten the input in order to be used by a `Dense` layer with four neurons and a linear activation function. As for the MLP models, this final layer outputs a vector with four elements that can be interpreted as the forecasts up to four weeks in advance.

Lastly, several LSTM models were built, with one to two hidden layers, using the Keras default activation functions. In the end, a `Dense` layer with four neurons and a linear activation function was used to output the forecasts up to four weeks in advance.

MLP and CNN experiments were performed using the Adam optimiser, whereas LSTM experiments used the RMSProp optimiser. All experiments considered a batch size of 8, 75 training epochs, and the Huber function as the loss function.

In order to determine the best hyper-parameters for each architecture, several experiments were conducted by testing several ANN models with increasing complexity with respect to the number of layers and the number of neurons in each layer. Furthermore, for some experiments, L1 and L2 regularization were applied to the networks. For each experiment, the training and the validation MAE values were monitored in order to identify which models were underfitting or overfitting. The experiment with the lowest validation error was chosen for each architecture. For more detail regarding these experiments and the parameter setting strategy, refer to the Supporting Information.

[3] https://keras.io/

Network training and validation. As with ARIMA, univariate ANN models were trained only with past values of the DSP toxins and bivariate ANNs were trained with the DSP toxins and the phytoplankton cell counts variables. Multivariate ANNs were trained with all the variables described in Tables S1 and S2. In addition, an autoencoder was used to decrease the number of features. An autoencoder is a neural network trained to reconstruct its input in its output from a constrained inside the network. It can be considered in two parts: an encoder, which maps the input to a latent space, $h = f(x)$, and a decoder, that maps the representation in the latent space back to the input, $r = g(h) = g(f(x))$. If the latent space has fewer dimensions than the input, the encoder learns to map the data into a lower dimensional representation [54]. An autoencoder was built to reduce the dimension of the seven meteorological variables provided by IPMA (mean, maximum and minimum air temperature, wind intensity, wind direction, and rainfall) to one feature. Daily data from twenty meteorological stations along the Portuguese coast, from January 2015 to December 2020, were used. As with the data used by the prediction models, these data were normalized to a scale of [0, 1] and split into training and validation sets.

The encoder part, composed of three dense layers with a ReLU activation function, reduced the dimension of the features from seven to one. Then, the decoder part, composed of two dense layers with a ReLU activation function and a final layer with a linear activation function, reconstructed the input back to its original shape. The autoencoder was trained during 20 epochs, using a batch size of 8 and Adam optimiser with a learning rate of 0.001. The mean squared error was used as the loss function, and this measure was also used to monitor the error on training and validation sets during the training process.

After training, the encoder part of the model, which had been trained to map the seven original variables to a useful representation, was stored. This encoder was then applied to the meteorological variables used as input to the prediction models, reducing their dimension from seven to one and allowing the ANN models to be trained using five variables (DSP toxins, phytoplankton cell counts, SST, chl-*a*, and the meteorological variable obtained by the encoder) instead of eleven.

For all the considered approaches, several ANN models with different complexities with respect to the number of layers and the number of neurons in each layer were tested. Then, models with the lowest validation mean absolute error (MAE) were chosen for each approach and architecture. The MAE is a commonly used metric to evaluate forecast accuracy [60], for its ease of interpretation. This error is given by

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |y_t - \widehat{y}_t| \,, \tag{3}$$

where y_t and $\widehat{y}_t$ are the actual and predicted values for time step t, respectively, and n is the number of predicted values.

Besides being used to select the best experiments, the MAE measure was also used to analyze the results obtained by each model (Section 3).

For practical purposes, it is also important to correctly predict whether toxins concentration will be above (positive) or below (negative) the regulatory limit, which, for the case of DSP toxins, corresponds to 160 μg OA equiv. kg^{-1} [5]. The ability of a model to predict this can be evaluated by creating a confusion matrix comparing the predicted classes to the actual classes and computing the true positive rate (TPR) and true negative rate (TNR) values, given, respectively, by

$$TPR = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \tag{4}$$

Table 1
MAE values obtained for each forecasting horizon in the training, validation and test sets of the RIAV1 production area.

Models		Training set				Validation set				Test set			
		$t+1$	$t+2$	$t+3$	$t+4$	$t+1$	$t+2$	$t+3$	$t+4$	$t+1$	$t+2$	$t+3$	$t+4$
Univariate	ARIMA	155.6	187.6	204.2	207.5	100.8	131.1	150.0	165.1	140.2	155.9	175.1	181.4
	MLP	83.8	175.4	198.0	235.6	41.9	90.1	113.8	122.5	75.8	169.7	183.0	208.2
	CNN	72.6	153.0	185.7	207.2	40.9	96.2	111.8	128.5	66.9	157.5	171.6	177.1
	LSTM	62.8	146.8	182.0	189.9	32.8	84.9	96.5	114.9	67.7	148.0	168.8	177.4
Bivariate	VAR	156.5	194.7	211.9	212.6	115.9	158.2	181.7	191.1	148.0	153.8	167.5	170.8
	MLP	60.8	156.1	180.1	200.6	49.9	99.2	114.0	127.2	77.5	158.1	176.2	179.9
	CNN	65.2	150.1	183.6	195.4	43.7	98.9	113.5	124.4	69.8	151.9	164.6	174.6
	LSTM	50.9	141.4	175.2	186.9	24.8	74.5	96.3	107.1	58.8	135.8	156.2	163.1
Multivariate	VAR	145.2	178.8	192.8	202.5	120.3	156.7	169.1	179.6	141.3	143.2	147.4	147.6
	MLP	101.4	157.8	187.6	198.9	84.3	111.5	126.7	131.8	137.9	162.0	172.2	180.8
	CNN	113.6	158.5	162.9	170.8	83.9	110.0	125.6	130.5	126.9	164.1	164.6	173.1
	LSTM	111.3	156.1	173.9	184.9	69.1	98.9	114.2	123.4	102.6	155.2	174.0	178.4
Multivariate (Autoencoder)	MLP	91.6	148.1	175.4	189.2	63.9	91.2	109.7	117.2	107.5	157.1	168.9	181.1
	CNN	90.9	156.3	181.8	192.2	41.9	85.9	104.9	120.6	87.0	149.8	168.1	177.3
	LSTM	85.4	152.8	176.9	186.1	45.0	84.3	104.0	116.1	86.8	149.1	169.6	175.2

Table 2
MAE values obtained for each forecasting horizon in the training, validation and test sets of the L5b production area.

Models		Training set				Validation set				Test set			
		$t+1$	$t+2$	$t+3$	$t+4$	$t+1$	$t+2$	$t+3$	$t+4$	$t+1$	$t+2$	$t+3$	$t+4$
Univariate	ARIMA	134.4	169.9	200.3	214.5	96.0	124.3	142.0	156.4	98.4	120.7	132.5	139.2
	MLP	86.0	138.7	175.0	184.9	39.0	87.1	106.7	110.6	36.4	96.5	114.2	121.6
	CNN	71.9	137.5	162.0	190.7	44.2	99.4	116.9	122.5	37.2	99.9	116.1	121.7
	LSTM	58.0	111.2	134.1	146.1	34.6	66.3	80.6	87.7	33.6	89.3	106.1	115.1
Bivariate	VAR	138.5	181.5	206.8	220.8	164.2	171.3	180.6	186.1	242.3	226.2	222.3	201.6
	MLP	62.3	133.7	156.2	171.9	36.3	89.7	109.2	109.4	33.2	95.6	112.7	119.6
	CNN	65.3	138.7	168.1	186.9	39.7	92.5	109.5	117.3	36.6	96.6	113.6	115.5
	LSTM	44.4	118.7	151.5	170.0	33.3	75.9	93.4	100.0	32.1	84.4	98.4	101.4
Multivariate	VAR	137.5	170.4	194.1	208.3	98.0	128.9	146.4	164.9	99.4	114.8	124.5	137.2
	MLP	101.1	140.9	172.1	183.6	79.5	98.7	112.7	119.2	76.4	103.0	123.0	119.2
	CNN	109.3	139.3	146.4	155.5	67.3	90.6	101.2	107.6	80.8	102.7	111.0	106.3
	LSTM	92.4	128.6	148.6	163.5	61.9	78.2	91.0	97.9	73.7	94.1	105.1	108.7
Multivariate (Autoencoder)	MLP	90.6	129.2	153.4	166.1	67.8	85.4	102.2	106.9	72.6	93.0	110.2	108.1
	CNN	98.1	143.1	153.7	168.9	56.1	84.6	94.2	105.7	57.0	95.8	102.4	107.4
	LSTM	79.9	131.3	151.6	165.5	49.7	78.1	93.4	100.5	52.4	89.9	99.7	102.1

$$TNR = \frac{\text{True negatives}}{\text{True negatives} + \text{False positives}} \quad (5)$$

Lastly, it is useful to evaluate whether the models can predict changes to the contamination state, i.e., if toxins concentration below the regulatory limit in a certain week will rise above the limit in the following week, or vice-versa. This can also be evaluated by creating a confusion matrix and analyzing the TPR and TNR values as above. More than evaluating the overall error of prediction, confusion matrices represent an easy and more intuitive way for model selection regarding the ability to predict the contamination state (above or below the regulatory limit) and transitions between contamination states, leading to opening or closure of the production area. A model that better predicts contamination state and state transitions is more useful from the shellfish production sector point of view. This is expected to increase knowledge transferability and better support decision.

2.2.3. Model comparison

To evaluate the significance of the differences between the 15 models generated, the Friedman non-parametric test was used on the model residuals, paired by prediction, to reject the hypothesis that all models behave in a statistically similar manner. Then the Nemenyi test was used for pairwise comparisons between the models, focusing on the models with the best performance. The Benjamini–Hochberg correction was applied due to multiple significance tests (each model is compared to the other 14 models) and the statistical significance of the differences between the models was obtained.

3. Results and discussion

This section shows the results obtained for the forecast of DSP toxins in mussels in the RIAV1 and L5b production areas using different prediction models: ARIMA, VAR, MLP, CNN and LSTM models. At first these were trained with data from each production area in separate but, due to the small amount of data collected for each area, this led to overfitting (results not shown). Thus, to increase the size of the training set, the data from eight production areas were merged to train the ANN models, as described in Section 2.2.2. Table S3 shows the combinations of ANN models and hyper-parameters used.

The MAE values obtained for the univariate models on the training, validation and test sets up to four weeks in the future are shown in Tables 1 and 2, for the RIAV1 and L5b areas, respectively. The results show a large increase in the error from the forecasting horizon $t+1$ to $t+2$ for the ANN models in all the datasets, suggesting that the models had much more difficulty predicting beyond one week into the future.

Comparing the classical ARIMA model with the univariate ANN models in the RIAV1 and L5b production areas (Tables 1 and 2), the ANNs outperformed ARIMA in the training and validation sets at nearly all forecasting horizons. These results indicate that the ANNs could better capture the patterns in the data when compared to the simpler autoregressive models. According to the validation MAE, among the ANNs, the LSTM outperformed all other univariate models at all forecasting horizons in both RIAV1 and L5b areas.

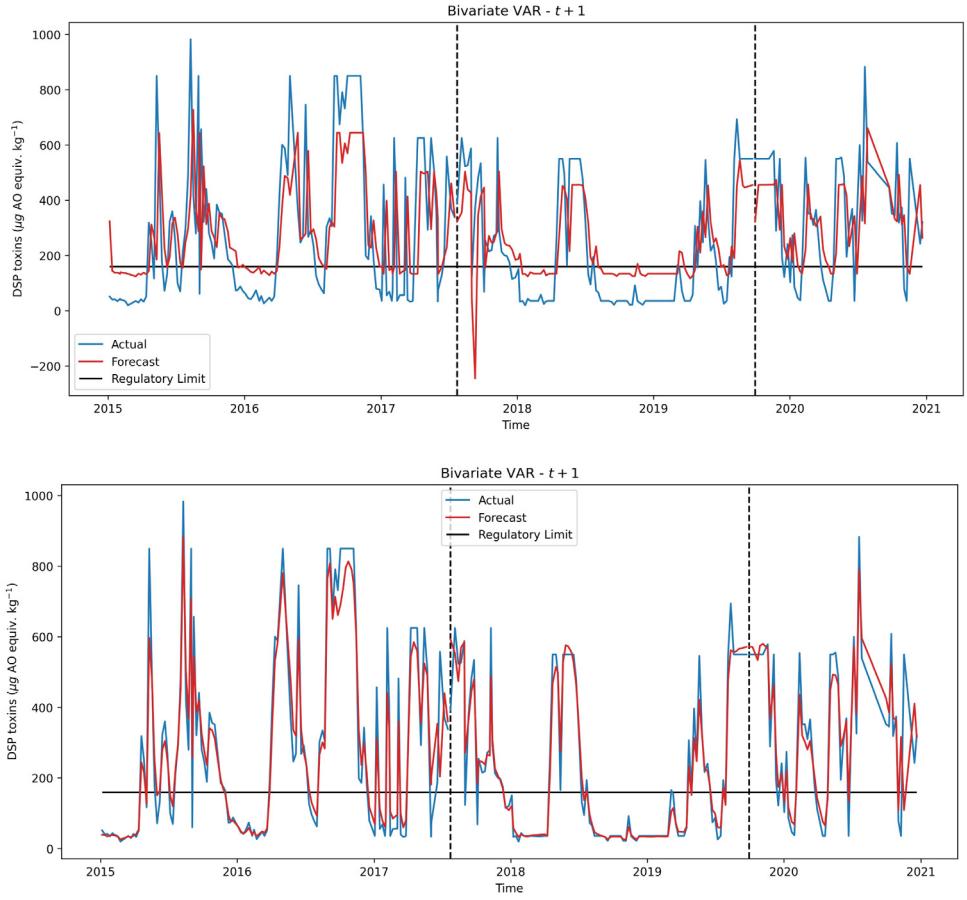

Fig. 2. Observed values for the DSP concentration in mussels (in blue) in the RIAV1 production area and one-week ahead forecasts (in red) obtained by the bivariate VAR (upper figure) and LSTM (bottom figure) models. The dashed lines separate the training, validation and test sets.

The DSP-producing phytoplankton are the direct cause of shellfish contamination by DSP toxins. Thus, in order to try to improve the results for the univariate models, multiple bivariate ANNs (Table S3) and a bivariate VAR model were built, which were trained using both the DSP toxins and the DSP-producing phytoplankton cell counts variables as input.

The bivariate VAR was outperformed by the univariate ARIMA in both training and validation sets from both production areas (Tables 1 and 2). For MLP and CNN, the bivariate models yielded a lower training error for RIAV1 but a slightly higher validation error than the univariate models. The bivariate LSTM model yielded better results than the univariate model in both training and validation sets. For the L5b area (Table 2), the use of the phytoplankton data as an explanatory variable in ANNs allowed to improve the results at the forecasting horizon $t + 1$ in the validation set, with the best results among the bivariate models being obtained by the LSTM.

Including the phytoplankton improved the results of some forecasting models. To try to improve these results, the meteorological and oceanographic variables described in Table S1 were also included. With this addition, multivariate VAR yielded better results than bivariate VAR (Tables 1 and 2). However, for the ANNs the multivariate models were generally outperformed by the univariate and bivariate models. The increased in the MAE, observed mainly in the validation set, suggested overfitting, i.e., the increase in the number of variables lead to the use of more complex models, which caused them to adapt too much to the details in the training data.

Thus, in order to improve the results of the multivariate models, an autoencoder was used to decrease the dimension of the seven meteorological variables into one feature. Networks trained on the data obtained by the autoencoder generally performed better than networks trained with all variables both for RIAV1 and L5b production areas (Tables 1 and 2), obtaining a lower validation error, indicating less overfitting. These results suggest that encoding meteorological data might be more suitable than using all the variables without transformation. However, univariate and bivariate ANNs perform even better at the first forecasting horizons for the MLP and CNN models, and at all forecasting horizons for the LSTM. Thus for the MAE, the univariate and bivariate models achieved a better performance than the multivariate models, especially at the first forecasting horizons. These results indicate that the additional variables used in this study are not as informative as expected, with the most useful variables for predicting toxins concentration being previous toxin concentration values and the phytoplankton that produces those toxins.

To summarize the results in terms of MAE, of all the models examined, for the RIAV1 production area the best results were obtained by the bivariate LSTM, followed by the univariate LSTM. In the test set, the best results were also obtained by the bivariate LSTM at the first forecasting horizons. However, the networks were outperformed by VAR at long forecasting horizons. For the L5b production area, the best models in both validation and test sets were also the univariate and bivariate LSTMs.

Besides evaluating the MAE values obtained by each model, it is also very useful to compare the predicted toxin values with the actual values using time plots. Taking the RIAV1 area at the forecasting horizon $t + 1$ as an illustrative example, the best results were obtained by the bivariate LSTM. The time plots

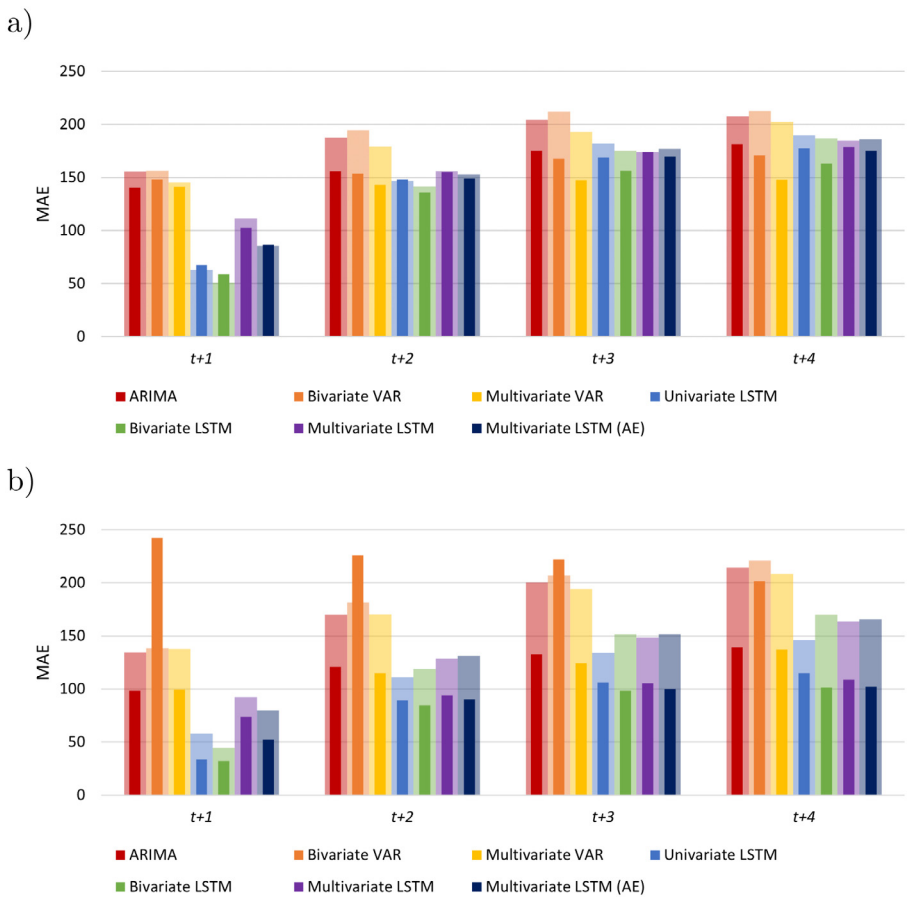

Fig. 3. MAE values obtained by autoregressive and LSTM models in (a) RIAV1 and (b) L5b production areas, for the training (wider bars) and test (narrower bars) sets.

for this model and its classical autoregressive counterpart for comparison, the bivariate VAR model, are shown in Fig. 2.

The VAR model seems to obtain predictions with a delay of about one week, i.e., the forecasts for the following week seem to be identical to the values of the current week, which despite leading to a relatively low MAE, is not useful. The bivariate LSTM gives better results, with very accurate predictions obtained for all data sets.

For a closer inspection of the predictive ability of the LSTM models, Fig. 3 compares their training and test MAE values with the corresponding values for the classic autoregressive models.

For both production areas, LSTM outperformed autoregressive models, in particular at $t + 1$, though the advantage was not as evident for longer forecast horizons in the RIAV1 area (Fig. 3). Among the autoregressive models, ARIMA and multivariate VAR performed similarly in both areas, with the multivariate model outperforming the univariate model at long forecasting horizons in RIAV1. Among the LSTMs, the best performance was with the univariate and bivariate LSTMs, followed by the LSTM combined with the autoencoder. Using all variables with the LSTM seems to result in overfitting.

The ability to predict whether toxins concentration will be above or below the regulatory limit was evaluated computing the TPR and TNR (Eqs. (4) and (5)) values for each model and forecasting horizon, shown in Fig. 4 for the autoregressive and ANN models for the RIAV1 and L5b areas. The autoregressive models yielded high TPR values at all forecasting horizons, particularly in RIAV1, but TNR values were low, meaning a high rate of false positives, with the model mistakenly predicting DSP values above the regulatory limit. This error can lead to unnecessary mitigation efforts in anticipation of interdictions that do not occur, leading to economic losses.

With the ANN models these errors were not as common. The TPR values in RIAV1 were similar among the models evaluated, with the CNN and LSTM models outperforming the MLP models and bivariate models outperforming all others for long forecasting horizons. The CNN and LSTM models also have a high TNR at $t + 1$, but not at longer forecasting horizons. For L5b, the models yielded lower TPR values in comparison to the RIAV1. In contrast, TNR values were significantly higher for L5b than for RIAV1. This is probably due to L5b having fewer high concentration events (Figures S1 and S2)). In fact, about 62% of the DSP time series values for this area are below the regulatory limit. On the other hand, for the RIAV1 area, only about 45% of the DSP concentration values are below the regulatory limits. This might have enabled forecasting models to better learn the dependencies between the variables under study for non-toxic events in L5b. For this region, multivariate models, namely the LSTM, had the best results with these metrics.

Aside from predicting when toxins are above or below the regulatory limit, it is particularly important that models predict when contamination changes from one state to the other, since the exact timing is crucial for the producer to take appropriate measures. This can also be evaluated by analyzing TPR and TNR values obtained for each model (Fig. 5), though in this case a positive event corresponds to a change of state, and a negative event to remaining in the same state as the previous week. In both production areas this seems to be more difficult than predicting contamination events. Nevertheless, TPR for ANN models are clearly superior to those for autoregressive models at all

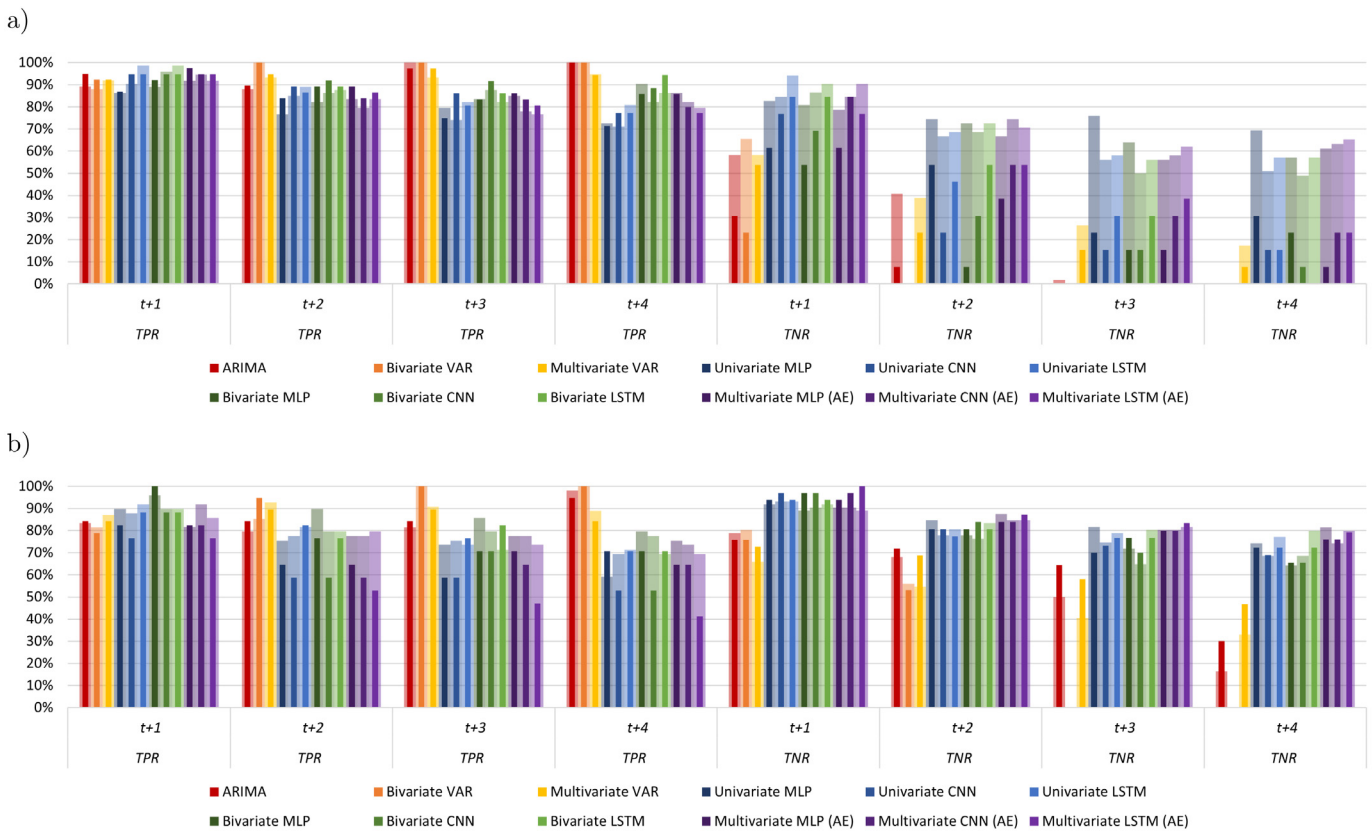

Fig. 4. TPR and TNR values for the toxins concentration above or below the regulatory limit obtained by the autoregressive and ANN models in (a) RIAV1 and (b) L5b production areas, for the training (wider bars) and test (narrower bars) sets.

forecasting horizons, although this metric is lower for forecasting horizons higher than $t + 1$. Among the ANN models differences are not consistent, with univariate and multivariate CNN and the LSTM giving higher TPR values in RIAV1 but bivariate MLP giving better results than the other networks in the L5b, particularly at $t + 1$.

Model comparison. The p value obtained with the Friedman test was very low (10^{-184}), indicating the models are not all behaving in the same way. the Nemenyi posthoc test was then used for the pairwise comparisons. Fig. 6 shows the results focusing on those models that seem to perform the best under the different criteria we used. These seem to be those using the features extracted with the autoencoder and the univariate and bivariate LSTM models.

The best performing models are nearly all significantly different from each other, with the exception of the CNN and MLP models using the features extracted with the autoencoder. Comparing with the remaining models, the LSTM models are not significantly different from some of these, especially the univariate LSTM, which seems to be indistinguishable from most other models.

4. Conclusions

Shellfish production can be affected by contamination by marine biotoxins, which causes sudden closures to the Portuguese production areas and leads to economic losses. Thus, it becomes necessary to forecast biotoxins concentration in shellfish in order to assist the productive sector in anticipating and mitigating these negative impacts.

In this study, several prediction models were explored with the goal of understanding which are the most useful and accurate in making these forecasts. We focused on predicting DSP toxins concentrations in mussels one to four weeks in advance, using univariate, bivariate, and multivariate time series forecasting methods.

The results showed that the models were considerably more accurate for one-week ahead predictions than for longer forecasting horizons. Thus, even though the forecasts beyond one week into the future are still useful to warn the production sector about possible closures to the shellfish production areas, they cannot be used with the same confidence as the one-week ahead forecasts.

Furthermore, in the areas under study, ANN models generally outperformed the autoregressive models, VAR and ARIMA, showing the potential for ANN models for these predictions. LSTM models seem to provide the lowest MAE, a global measure of performance to compare the models, but for the more useful metrics of contamination state and state change predictions, different combinations of ANN architectures (MLP, CNN and LSTM) and inputs (univariate, bivariate and multivariate) produced better results on a case-by-case basis, depending on the specificities of the time series under study. Interestingly, bivariate models yielded comparable and for some cases superior results compared to the multivariate models. Indeed, the oceanographic and meteorological variables used for model building showed not be as informative as expected. For instance, remotely sensed chl-*a* is used as a proxy for phytoplankton, but contamination only arises from toxic phytoplankton, which might turn this variable not as relevant as foreseen. Moreover, the meteorological variables were obtained by IPMA's meteorological stations, not exactly at the same location as the sampling points, which might lower the contribution of these variables to the model outcome. Further modeling efforts to predict shellfish contamination by marine biotoxins might encompass the incorporation of other time-series

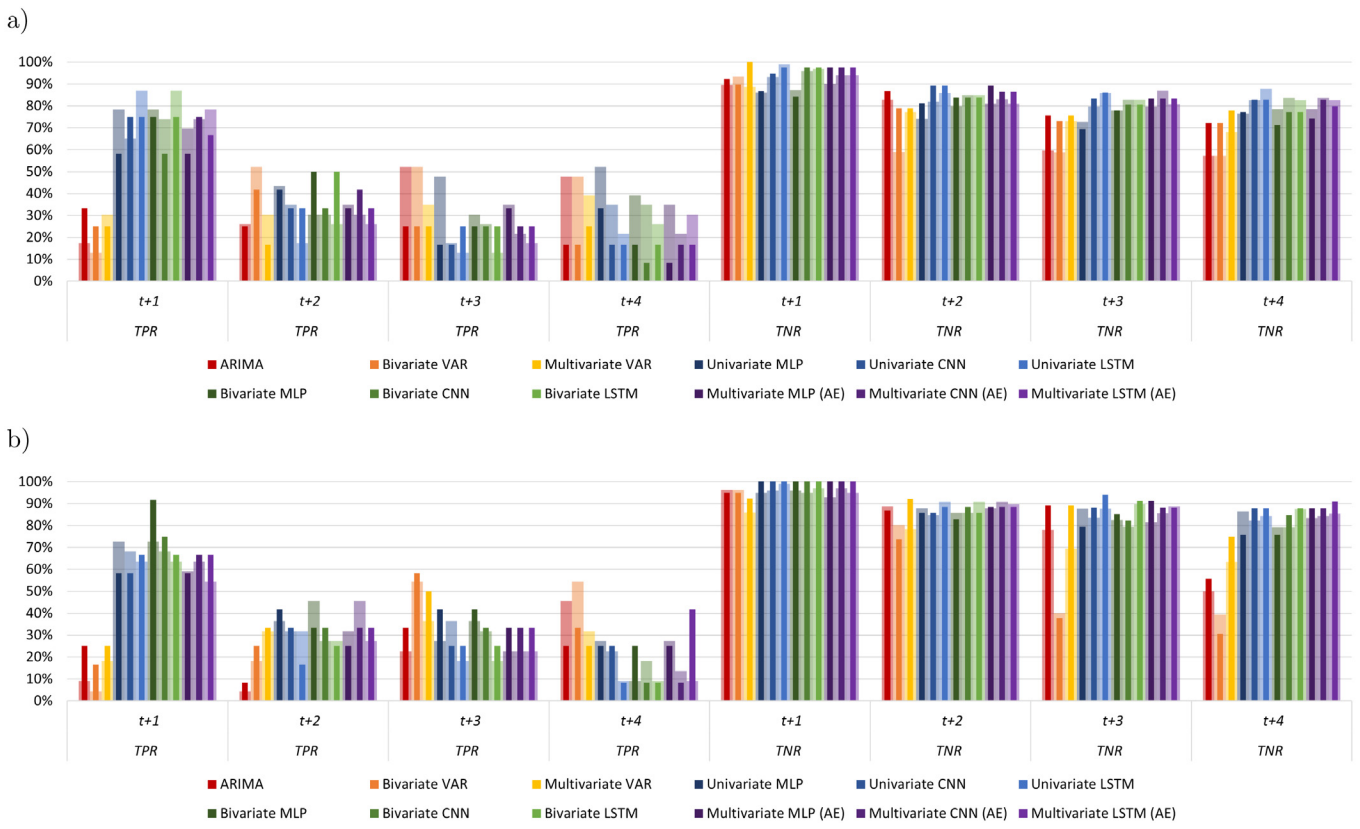

Fig. 5. TPR and TNR values for the changes to the contamination state obtained by the autoregressive and ANN models in (a) RIAV1 and (b) L5b production areas, for the training (wider bars) and test (narrower bars) sets.

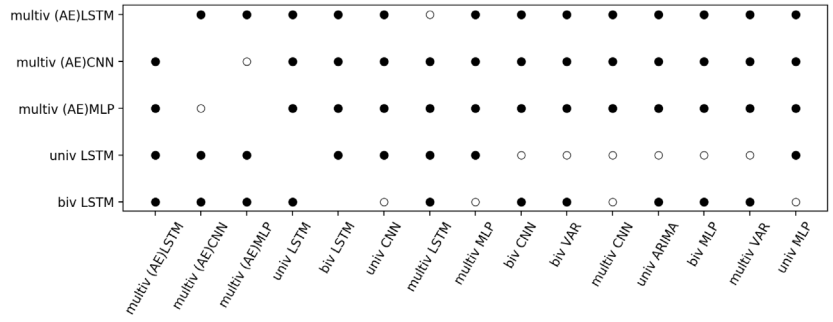

Fig. 6. Pairwise comparison of models using a significance value of 0.05. Black circles indicate statistically significant differences, using the corrected Nemenyi test. White circles indicate p values greater than 0.05.

variables which are expected to have a role in HABs formation and subsequent shellfish contamination, namely, salinity, pH, oxygen concentration, dissolved organic matter, and hydrodynamic variables. In which concerns model development, future research might consider the use of deep neural networks for time-series forecasting, as recently shown promising for water quality prediction scenarios.

In conclusion, forecasting shellfish contamination by marine biotoxins is a complex and challenging problem. Nonetheless, using, for example, a bivariate LSTM trained on the biotoxins and the phytoplankton variables, which are variables routinely measured within shellfish monitoring systems worldwide, it is possible to obtain accurate predictions one week into the future. This represents an invaluable contribution to the shellfish production sector, by enabling anticipating and mitigating the negative impacts that closures impose on all players involved in the shellfish production chain, and a great step forward towards the development of a forecasting system for shellfish contamination by marine biotoxins. Further research efforts might be employed to extend the present methodology to less frequent toxins and other shellfish species.

CRediT authorship contribution statement

Rafaela C. Cruz: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Pedro R. Costa:** Conceptualization, Writing – review & editing, Funding acquisition. **Ludwig Krippahl:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Marta B. Lopes:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by the project "MATISSE: A machine learning-based forecasting system for shellfish safety" (DSAIPA/DS/0026/2019). The work was also supported by national funds through Fundação para a Ciência e a Tecnologia (FCT), Portugal with references CEECINST/00102/2018, UIDB/04516/2020 (NOVA LINCS), UIDB/00297/2020 and UIDP/00297/2020 (NOVA MATH), and UID/Multi/04326/2020 (CCMAR).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.knosys.2022.109895.

References

[1] Daniel Taylor, Camille Saurel, Pernille Nielsen, Jens Kjerulf Petersen, Production characteristics and optimization of mitigation mussel culture, Front. Mar. Sci. 6 (2019) 698, http://dx.doi.org/10.3389/fmars.2019.00698.

[2] Felipe Matarazzo Suplicy, A review of the multiple benefits of mussel farming, Rev. Aquacult. 12 (1) (2020) 204–223, http://dx.doi.org/10.1111/raq.12313.

[3] Hallegraeff G.M., Anderson D.M., Cembella A.D., Enevoldsen H.O., Manual on Harmful Marine Microalgae (Monographs on Oceanographic Methodology, Vol. 11), second ed., UNESCO, ISBN: 92-3-103948-2, 2004.

[4] Patrice Guillotreau, Véronique Le Bihan, Baptiste Morineau, Sophie Pardo, The vulnerability of shellfish farmers to HAB events: An optimal matching analysis of closure decrees, Harmful Algae (ISSN: 1568-9883) 101 (2021) 101968, http://dx.doi.org/10.1016/j.hal.2020.101968.

[5] European Commission, Commission Regulation (EC) No 853/2004 of the European Parliament and of the Council of 29 April 2004 laying down specific hygiene rules for on the hygiene of foodstuffs, Official J. Eur. Union L 139 (2004) 55–205.

[6] European Commission, Commission Regulation (EC) No 854/2004 of the European Parliament and of the Council of 29 April 2004 laying down specific rules for the organisation of official controls on products of animal origin intended for human consumption, Official J. Eur. Union L 139 (2004) 206–320.

[7] Joachim W. Dippner, Lam Nguyen-Ngoc, Hai Doan-Nhu, Ajit Subramaniam, A model for the prediction of harmful algae blooms in the Vietnamese upwelling area, Harmful Algae (ISSN: 1568-9883) 10 (6) (2011) 606–611, http://dx.doi.org/10.1016/j.hal.2011.04.012.

[8] Caroline Cusack, Tomasz Dabrowski, Kieran Lyons, Alan Berry, Guy Westbrook, Rafael Salas, Conor Duffy, Glenn Nolan, Joe Silke, Harmful algal bloom forecast system for SW Ireland. Part II: Are operational oceanographic models useful in a HAB warning system, Harmful Algae (ISSN: 1568-9883) 53 (2016) 86–101, http://dx.doi.org/10.1016/j.hal.2015.11.013, Applied Simulations and Integrated Modelling for the Understanding of Toxic and Harmful Algal Blooms (ASIMUTH).

[9] Julie Maguire, Caroline Cusack, Manuel Ruiz-Villarreal, Joe Silke, Deirdre McElligott, Keith Davidson, Applied simulations and integrated modelling for the understanding of toxic and harmful algal blooms (ASIMUTH): Integrated HAB forecast systems for Europe's Atlantic Arc, Harmful Algae (ISSN: 1568-9883) 53 (2016) 160–166, http://dx.doi.org/10.1016/j.hal.2015.11.006, Applied Simulations and Integrated Modelling for the Understanding of Toxic and Harmful Algal Blooms (ASIMUTH).

[10] Marvin F. Li, Patricia M. Glibert, Vyacheslav Lyubchich, Machine learning classification algorithms for predicting karenia brevis blooms on the west florida shelf, J. Mar. Sci. Eng. (ISSN: 2077-1312) 9 (9) (2021) http://dx.doi.org/10.3390/jmse9090999.

[11] Theodore J. Smayda, Turbulence, watermass stratification and harmful algal blooms: an alternative view and frontal zones as "pelagic seed banks", Harmful Algae (ISSN: 1568-9883) 1 (1) (2002) 95–112, http://dx.doi.org/10.1016/S1568-9883(02)00010-0.

[12] Mark L. Wells, Vera L. Trainer, Theodore J. Smayda, Bengt S.O. Karlson, Charles G. Trick, Raphael M. Kudela, Akira Ishikawa, Stewart Bernard, Angela Wulff, Donald M. Anderson, William P. Cochlan, Harmful algal blooms and climate change: Learning from the past and present to forecast the future, Harmful Algae 49 (2015) 68–93, http://dx.doi.org/10.1016/j.hal.2015.07.009.

[13] Scott A. Condie, Eric C.J. Oliver, Gustaaf M. Hallegraeff, Environmental drivers of unprecedented Alexandrium catenella dinoflagellate blooms off eastern Tasmania, 2012–2018, Harmful Algae (ISSN: 1568-9883) 87 (2019) 101628, http://dx.doi.org/10.1016/j.hal.2019.101628.

[14] E. Zohdi, M. Abbaspour, Harmful algal blooms (red tide): a review of causes, impacts and approaches to monitoring and prediction, Int. J. Environ. Sci. Technol. 16 (2019) 1789–1806, http://dx.doi.org/10.1007/s13762-018-2108-x.

[15] J. Wen, J. Yang, Y. Li, L. Gao, Harmful algal bloom warning based on machine learning in maritime site monitoring, Knowl.-Based Syst. 245 (2022) 108569, http://dx.doi.org/10.1016/j.knosys.2022.108569.

[16] Rafaela C. Cruz, Pedro Reis Costa, Susana Vinga, Ludwig Krippahl, Marta B. Lopes, A review of recent machine learning advances for forecasting harmful algal blooms and shellfish contamination, J. Mar. Sci. Eng. (ISSN: 2077-1312) 9 (3) (2021) http://dx.doi.org/10.3390/jmse9030283.

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

[18] Gilbert C.S. Lui, W.K. Li, Kenneth M.Y. Leung, Joseph H.W. Lee, A.W. Jayawardena, Modelling algal blooms using vector autoregressive model with exogenous variables and long memory filter, Ecol. Model. 200 (2007) 130–138, http://dx.doi.org/10.1016/j.ecolmodel.2006.06.017.

[19] Muriel Gevrey, Ioannis Dimopoulos, Sovan Lek, Review and comparison of methods to study the contribution of variables in artificial neural network models, Ecol. Model. 160 (2003) 249–264, http://dx.doi.org/10.1016/S0304-3800(02)00257-0.

[20] F. Recknagel, M. French, P. Harkonen, K.-I. Yabunaka, Artificial neural network approach for modelling and prediction of algal blooms, Ecol. Model. 96 (1997) 11–28, http://dx.doi.org/10.1016/S0304-3800(96)00049-X.

[21] Joseph H.W. Lee, Yan Huang, Mike Dickman, A.W. Jayawardena, Neural network modelling of coastal algal blooms, Ecol. Model. 159 (2–3) (2003) 179–201, http://dx.doi.org/10.1016/S0304-3800(02)00281-8.

[22] L. Velo-Suárez, J.C. Gutiérrez-Estrada, Artificial neural network approaches to one-step weekly prediction of *Dinophysis acuminata* blooms in Huelva (Western Andalucía, Spain), Harmful Algae (ISSN: 1568-9883) 6 (3) (2007) 361–371, http://dx.doi.org/10.1016/j.hal.2006.11.002.

[23] Xiu Li, Jin Yu, Zhuo Jia, Jingdong Song, Harmful algal blooms prediction with machine learning models in tolo harbour, in: 2014 International Conference on Smart Computing, 2014, pp. 245–250, http://dx.doi.org/10.1109/SMARTCOMP.2014.7043865.

[24] Paul R. Hill, Anurag Kumar, Marouane Temimi, David R. Bull, HABNet: machine learning, remote sensing-based detection of harmful algal blooms, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13 (2020) 3229–3239, http://dx.doi.org/10.1109/JSTARS.2020.3001445.

[25] Sangmok Lee, Donghyun Lee, Improved prediction of harmful algal blooms in four major south korea's rivers using deep learning models, Int. J. Environ. Res. Public Health 15 (7) (2018) 1322, http://dx.doi.org/10.3390/ijerph15071322.

[26] Hyungmin Cho, U-Jin Choi, Heekyung Park, Deep learning application to time series prediction of daily chlorophyll-a concentration, WIT Trans. Ecol. Environ. 215 (2018) 157–163, http://dx.doi.org/10.2495/EID180141.

[27] H. Cho, H. Park, Merged-LSTM and multistep prediction of daily chlorophyll-a concentration for algal bloom forecast, IOP Conf. Ser.: Earth Environ. Sci. 351 (2019) 012020, http://dx.doi.org/10.1088/1755-1315/351/1/012020.

[28] Fatin Nadiah Yussof, Normah Maan, Mohd Nadzri Md Reba, LSTM networks to improve the prediction of harmful algal blooms in the West Coast of sabah, Int. J. Environ. Res. Public Health 18 (14) (2021) 7650, http://dx.doi.org/10.3390/ijerph18147650.

[29] J.I. Ubah, L.C. Orakwe, K.N. Ogbu, J.I. Awu, I.E. Ahaneku, E.C. Chukwuma, Forecasting water quality parameters using artificial neural network for irrigation purposes, Sci. Rep. 11 (2021) 24438, http://dx.doi.org/10.1038/s41598-021-04062-5.

[30] Saber Kouadri, Ahmed Elbeltagi, Abu Reza Md. Towfiqul Islam, Samir Kateb, Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast), Applied Water Science 11 (2021) 190, http://dx.doi.org/10.1007/s13201-021-01528-9.

[31] Shahaboddin Shamshirband, Ehsan Jafari Nodoushan, Jason E. Adolf, Azizah Abdul Manaf, Amir Mosavi, Kwok-wing Chau, Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters, Eng. Appl. Comput. Fluid Mech. 13 (1) (2019) 91–101, http://dx.doi.org/10.1080/19942060.2018.1553742.

[32] Senlin Zhu, Bahrudin Hrnjica, Mariusz Ptak, Adam Choiński, Bellie Sivakumar, Forecasting of water level in multiple temperate lakes using machine learning models, J. Hydrol. (ISSN: 0022-1694) 585 (2020) 124819, http://dx.doi.org/10.1016/j.jhydrol.2020.124819.

[33] Jing Bi, Yongze Lin, Quanxi Dong, Haitao Yuan, MengChu Zhou, Large-scale water quality prediction with integrated deep neural network, Inform. Sci. (ISSN: 0020-0255) 571 (2021) 191–205, http://dx.doi.org/10.1016/j.ins.2021.04.057.
[34] Nguyen Thai-Nghe, Nguyen Thanh-Hai, Nguyen Chi Ngon, Deep learning approach for forecasting water quality in IoT systems, Int. J. Adv. Comput. Sci. Appl. 11 (8) (2020) http://dx.doi.org/10.14569/IJACSA.2020.0110883.
[35] Ping Liu, Jin Wang, Arun Kumar Sangaiah, Yang Xie, Xinchun Yin, Analysis and prediction of water quality using LSTM deep neural networks in IoT environment, Sustainability 11 (7) (2019) 1050–2071, http://dx.doi.org/10.3390/su11072058.
[36] Isabella Grasso, Stephen D. Archer, Craig Burnell, Benjamin Tupper, Carlton Rauschenber, Kohl Kanwit, Nicholas R. Record, The hunt for red tides: Deep learning algorithm forecasts shellfish toxicity at site scales in coastal Maine, Ecosphere 10 (12) (2019) http://dx.doi.org/10.1002/ecs2.2960.
[37] John R. Harley, Kari Lanphier, Esther Kennedy, Chris Whitehead, Allison Bidlack, Random forest classification to determine environmental drivers and forecast paralytic shellfish toxins in Southeast Alaska with high temporal resolution, Harmful Algae 99 (2020) 101918, http://dx.doi.org/10.1016/j.hal.2020.101918.
[38] P. Vale, M.J. Botelho, S.M. Rodrigues, S.S. Gomes, M.A.D.M. Sampayo, Two decades of marine biotoxin monitoring in bivalves from Portugal (1986–2006): a review of exposure assessment, Harmful Algae 7 (2008) 11–25, http://dx.doi.org/10.1016/j.hal.2007.05.002.
[39] E. Bresnan, F. Arévalo, C. Belin, M. Branco, et al., Diversity and regional distribution of harmful algal events along the atlantic margin of europe, Harmful Algae 102 (2021) http://dx.doi.org/10.1016/j.hal.2021.101976.
[40] M.T. Moita, V. Pazos, C. Rocha, R. Nolasco, P.B. Oliveira, Toward predicting dinophysis blooms off NW iberia: a decade of events, Harmful Algae 5 (2016) 17–32, http://dx.doi.org/10.1016/j.hal.2015.12.002.
[41] L.A. Krug, T. Platt, S. Sathyendranath, A.B. Barbosa, Unravelling region-specific environmental drivers of phytoplankton across a complex marine domain (off SW Iberia), Remote Sens. Environ. 203 (2017) http://dx.doi.org/10.1016/j.rse.2017.05.029.
[42] About copernicus, URL: https://www.copernicus.eu/en/about-copernicus.
[43] Chris Chatfield, Time-Series Forecasting, first ed., Chapman & Hall/CRC, ISBN: 978-1584880639, 2001.
[44] James H. Stock, Mark W. Watson, Vector autoregressions, J. Econ. Perspect. 15 (4) (2001) 101–115, http://dx.doi.org/10.1257/jep.15.4.101.
[45] Galit Shmueli, Kenneth C. Lichtendahl, Practical Time Series Forecasting with R: A Hands-on Guide, second ed., Axelrod Schnall Publishers, 2016.
[46] Ruey S. Tsay, Multivariate Time Series Analysis: With R and Financial Applications, first ed., John Wiley & Sons, ISBN: 978-1-118-61790-8, 2014.
[47] R. Scott Hacker, Abdulnasser Hatemi-J., Optimal lag-length choice in stable and unstable VAR models under situations of homoscedasticity and ARCH, J. Appl. Stat. 35 (2008) 601–615, http://dx.doi.org/10.1080/02664760801920473, Applied Simulations and Integrated Modelling for the Understanding of Toxic and Harmful Algal Blooms (ASIMUTH).
[48] Alev Dilek Aydin, Seyma Caliskan Cavdar, Two different points of view through artificial intelligence and vector autoregressive models for ex post and ex ante forecasting, Computational Intelligence and Neuroscience 2015 (2015) 409361, http://dx.doi.org/10.1155/2015/409361.
[49] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37 (3) (1969) 424–438, http://dx.doi.org/10.2307/1912791.
[50] Spencer S. Jones, R. Scott Evans, Todd L. Allen, Alun Thomas, Peter J. Haug, Shari J. Welch, Gregory L. Snow, A multivariate time series approach to modeling and forecasting demand in the emergency department, J. Biomed. Inform. 42 (2009) 123–139, http://dx.doi.org/10.1016/j.jbi.2008.05.003.
[51] Hansika Hewamalage, Christoph Bergmeir, Kasun Bandara, Recurrent Neural Networks for Time Series Forecasting: Current status and future directions, Int. J. Forecast. 37 (1) (2021) 388–427, http://dx.doi.org/10.1016/j.ijforecast.2020.06.008.
[52] Kanad Chakraborty, Kishan Mehrotra, Chilukuri K. Mohan, Sanjay Ranka, Forecasting the behavior of multivariate time series using neural networks, Neural Netw. 5 (1992) 961–970, http://dx.doi.org/10.1016/S0893-6080(05)80092-9.
[53] Nitin Muttil, Kwok-Wing Chau, Neural network and genetic programming for modelling coastal algal blooms, Int. J. Environ. Pollut. 28 (2006) 223–238, http://dx.doi.org/10.1504/IJEP.2006.011208.
[54] Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016.
[55] Anastasia Borovykh, Sander Bohte, Cornelis W. Oosterlee, Conditional time series forecasting with convolutional neural networks, 2018, arXiv:1703.04691.
[56] Irena Koprinska, Dengsong Wu, Zheng Wang, Convolutional neural networks for energy time series forecasting, in: 2018 International Joint Conference on Neural Networks, IJCNN, ISBN: 978-1-5090-6014-6, 2018, pp. 1–8, http://dx.doi.org/10.1109/IJCNN.2018.8489399.
[57] Bryan Lim, Stefan Zohren, Time series forecasting with deep learning: a survey, 2020, arXiv:2004.13408.
[58] Sepp Hochreiter, Jürgen Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, http://dx.doi.org/10.1162/neco.1997.9.8.1735.
[59] J. Brownlee, Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in python, 2018.
[60] Rob J. Hyndman, Anne B. Koehler, Another look at measures of forecast accuracy, Int. J. Forecast. 22 (9) (2006) 679–688, http://dx.doi.org/10.1016/j.ijforecast.2006.03.001.