sustainability

MDPI

Article

A Predictive Model for the Bioaccumulation of Okadaic Acid in *Mytilus galloprovincialis* Farmed in the Northern Adriatic Sea: A Tool to Reduce Product Losses and Improve Mussel Farming Sustainability

Fabrizio Capoccioni [1], Laura Bille [2], Federica Colombo [3], Lidia Contiero [2], Arianna Martini [1], Carmine Mattia [3], Riccardo Napolitano [1], Nicolò Tonachella [1], Marica Toson [2] and Domitilla Pulcini [1,*]

[1] Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Centro di Ricerca Zootecnia e Acquacoltura, Via Salaria 31, 00015 Monterotondo, Italy; fabrizio.capoccioni@crea.gov.it (F.C.); riccardo.napolitano@crea.gov.it (R.N.); nicolo.tonachella@crea.gov.it (N.T.)
[2] Istituto Zooprofilattico Sperimentale Delle Venezie, Viale dell'Università 10, 35020 Legnaro, Italy; lbille@izsvenezie.it (L.B.); mtoson@izsvenezie.it (M.T.)
[3] ColomboSky S.r.l., Via Giacomo Peroni 442/444, 00131 Roma, Italy
* Correspondence: domitilla.pulcini@crea.gov.it; Tel.: +39-0690090263

Abstract: Over the last decades, harmful dinoflagellate (*Dinophysis* spp.) blooms have increased in frequency, duration, and severity in the Mediterranean Sea. Farmed bivalves, by ingesting large amounts of phytoplankton, can become unsafe for human consumption due to the bioaccumulation of okadaic acid (OA), causing Diarrhetic Shellfish Poisoning (DSP). Whenever the OA concentration in shellfish farmed in a specific area exceeds the established legal limit (160 μg·kg^{-1} of OA equivalents), harvesting activities are compulsorily suspended. This study aimed at developing a machine learning (ML) predictive model for OA bioaccumulation in Mediterranean mussels (*Mytilus galloprovincialis*) farmed in the coastal area off the Po River Delta (Veneto, Italy), based on oceanographic data measured through remote sensing and data deriving from the monitoring activities performed by official veterinarian authorities to verify the bioaccumulation of OA in the shellfish production sites. LightGBM was used as an ML algorithm. The results of the classification algorithm on the test set showed an accuracy of 82%. Further analyses showed that false negatives were mainly associated with relatively low levels of toxins (<100 μg·kg^{-1}), since the algorithm tended to classify low concentrations of OA as negative samples, while true positives had higher mean values of toxins (139 μg·kg^{-1}). The results of the model could be used to build up an online early warning system made available to shellfish farmers of the study area, aimed at increasing the economic and environmental sustainability of these production activities and reducing the risk of massive product losses.

Keywords: harmful algal blooms; LightGBM algorithm; machine learning; okadaic acid; remote sensing; shellfish

check for updates

Citation: Capoccioni, F.; Bille, L.; Colombo, F.; Contiero, L.; Martini, A.; Mattia, C.; Napolitano, R.; Tonachella, N.; Toson, M.; Pulcini, D. A Predictive Model for the Bioaccumulation of Okadaic Acid in *Mytilus galloprovincialis* Farmed in the Northern Adriatic Sea: A Tool to Reduce Product Losses and Improve Mussel Farming Sustainability. *Sustainability* **2023**, *15*, 8608. https://doi.org/10.3390/su15118608

Academic Editor: George P. Kraemer

Received: 30 March 2023
Revised: 20 May 2023
Accepted: 22 May 2023
Published: 25 May 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Over the last two decades, Harmful Algal Blooms (HABs) have increased in frequency, duration, and severity worldwide due to coastal eutrophication caused by the growing human population and the consequent rise in rates of coastal urbanisation, industrialisation, and intensification of agriculture. This phenomenon is expected to intensify in the coming years due to the effects of climate change [1–4]. Some species of harmful algae, such as dinoflagellates, which benefit from water column stratification and from the increased nutrient concentration due to land runoff, could become even more successful and their seasonal window of growth could expand, with earlier timing of peaks of abundance [3]. The definition of HAB is not strictly scientific, but social: blooms fit the HAB criterion

whether they have an adverse effect on human health, socio-economic interests, or on aquatic ecosystems.

Among phytoplanktonic groups, dinoflagellates stand out for their morphological diversity, adaptive radiation in colonising the diverse habitats found in the sea, and species richness, with approximately 2000 species recognised. About 10% of dinoflagellate species are associated with toxic blooms or benign red tides. Of these, approximately 3% (ca. 60 species) are reported to be harmful due to the production of okadaic acid (OA) and other toxic compounds, whereas about 130 species can lead to the onset of massive red tides, with related and serious anoxic events in the upper layers of the sea [5].

Bivalve molluscs, by ingesting large amounts of phytoplankton during HABs, could become unsafe for human consumption [6,7]. Phytoplanktonic dinoflagellates of the cosmopolitan genus *Dinophysis* occur in various coastal waters [8], including those of the Adriatic Sea [9]. Several *Dinophysis* species produce diarrhoetic toxins (OA and dinophysistoxins) and pectenotoxins, causing a gastrointestinal illness known as Diarrhetic Shellfish Poisoning (DSP), even at low cell densities (<100 cells·L^{-1}) [10]. Although the exact function of OA in some *Dinophysis* species is not entirely clear, it is likely produced as a chemical defence mechanism to protect these dinoflagellates from predation by zooplankton and other organisms [11].

The Mediterranean mussel (*Mytilus galloprovincialis* L. 1819) farming areas along the Italian coasts of the northern Adriatic Sea have experienced recurrent DSP events since 1989 (Slovenian coast of the Gulf of Trieste) [9,12], particularly frequently in the marine area facing the Po River Delta (Veneto, Italy) [13]. This coastal area is characterised by the highest production of bivalve molluscs (mussels and clams) of the eastern Mediterranean, with more than 56,000 tons in 2021 (60% mussel, 40% clam—Italian Aquaculture Data Collection, Reg. EC 762/2008) thanks to the optimal environmental conditions favouring the growth of bivalves.

In these areas, several *Dinophysis* species were found to be involved in the occurrence of DSP, with higher abundances registered from May to October and peaks in early summer and autumn [14,15]. Several studies showed a marked species succession during this period [16,17]: in general, *Dinophysis sacculus* is a late-spring species, peaking in June, followed by *Dinophysis caudata* in early autumn and *Dinophysis fortii* in October–November. Blooms of toxic *Dinophysis* species (such as *Dinophysis acuminata*, *D. sacculus*, *D. fortii*, *Dinophysis rotundata*, and *D. caudata*) are of particular concern for shellfish farmers because bivalves may accumulate enough toxins to become acutely harmful even at low densities (from 10 to 200 cells·L^{-1}) [18–21].

According to the observed trends, the impact of *Dinophysis*' blooms is expected to increase in the future, together with the even more increasing use of marine coastal waters for multiple human activities (i.e., tourism, fisheries and aquaculture, gas extraction, navigation, etc.).

To protect human health, the European Union implemented mandatory monitoring of classified shellfish production areas on a weekly basis (EU Commission, 2019). Whenever the OA concentration in shellfish exceeds the established legal limit (equal to 160 µg·kg^{-1} of OA equivalents, EU Council, 2004), shellfish harvesting activities are compulsorily suspended and the product already collected is seized and dumped.

Therefore, mitigation of such a phenomenon should prioritise applied research on the development of early warning systems designed to avoid economic losses for farmers and to optimise the use of resources for monitoring programs.

To the best of our knowledge, the available studies that tried to predict *Dinophysis* blooms by correlating environmental (e.g., temperature, salinity, density, tidal velocity, wind speed and direction, dissolved oxygen, and nutrient content) [22] or biological parameters (e.g., availability of *Dinophysis* spp. preys, such as *Mesodinium* spp.) [23] to *Dinophysis* spp. abundance failed in finding linear relationships and significant temporal correlations, due to the unique ecological behaviour (i.e., mixotrophy) of this dinoflagellate [24,25]. On

the contrary, Hattenrath-Lehmann et al. [18] found a positive correlation for *D. acuminata* between its abundance and toxin content with ammonium.

Since these ecological relationships are very complex, Velo-Suárez and Gutiérrez-Estrada [19] successfully applied an Artificial Neural Network (ANN) modelling approach to forecast *D. acuminata* blooms in western Andalucía (Spain) using cell concentrations data. Recently, in the coastal waters of the NW Adriatic Sea, a mathematical model obtained promising results in predicting the presence of a different species of toxic dinoflagellate (*Alexandrium minutum*) using a Random Forest trained with molecular data of *A. minutum* occurrence obtained using molecular PCR assay [20].

The present study aimed at developing a machine learning (ML) predictive model for OA bioaccumulation in farmed Mediterranean mussel (*M. galloprovincialis*) in the coastal area off the Po River Delta (Veneto, Italy) based on oceanographic data measured through remote sensing and data derived from the official monitoring activities to predict the bioaccumulation of OA in farmed shellfish. While a precise threshold level does exist for OA bioaccumulation in bivalve mussels (Reg. (EC) 627/19 and updates), there is not a scientifically established abundance level of harmful cells during algal blooms, since *Dinophysis* species may be toxic even at very low concentrations [17–20]. For this reason, to the best of our knowledge, this is the first study focused on data analysis of OA bioaccumulation in mussels, rather than on those of abundance of OA-producing species.

Although the performance of such a model can be improved by adding new and more detailed data, the promising results of this study could be used to develop an online early warning system able to promptly alert shellfish farmers to the chance of OA bioaccumulation events in the farming areas. The potential of this tool would give farmers the possibility to better manage farming activities, not only reducing losses in terms of harvested bivalve biomass and economic shortages but also improving the environmental sustainability of the mussel farming sector. In the last year, in fact, due to the increase in the frequency and severity of HAB events, mussel farms in the northern Adriatic Sea have been forced to close and keep the product alive in the farms, without being able to sell it. This has often resulted in increased running costs and greenhouse gas emissions.

2. Materials and Methods

Data on OA bioaccumulation in Mediterranean mussels' *(M. galloprovincialis*) flesh derived from the official monitoring activities carried out by the local veterinary services according to the Reg. (EC) 627/19 from 2016 to 2020 in the offshore production sites located in the marine environment off the Po River Delta. This is one of the most important areas for the production of *M. galloprovincialis* in the north-eastern Mediterranean, with a mean production of about 8050 tons from 2018 to 2021 (Aquaculture Data Collection, Reg. (EC) 762/2008). Offshore sampling stations (N = 167) were selected as those with the highest satellite coverage (Figure 1). The dataset consisted of 1486 observations, distributed per year as reported in Table 1. The number of observations available in 2020 was noticeably lower than that in previous years since sampling activities were performed based on the production areas instead of the single farm (a classified production area may include several farms), as required by the legislation in force (Reg. (EC) 627/19 and updates). The reference analytical method for the detection of OA and all the lipophilic toxins for the purposes of official controls at any stage in the food chain is the EU-RL (European Union Reference Laboratory on marine biotoxins) LC-MS/MS (liquid chromatography-mass spectrometry) (Reg. (EU) No 15/2011 and updates). According to the legal limit set by Reg. (EC) No 853/2004, 74 positive samples were detected in the dataset (>160 $\mu g \cdot kg^{-1}$ of OA equivalents). Since the goal of the study was to predict OA bioaccumulation in mussels, and due to the limited number of positive samples, data were classified according to a binary system as follows: class "1" for samples with a toxin content above the limit of quantification (LOQ) of the analytical method (LOQ = 40 $\mu g \cdot kg^{-1}$ of OA equivalents), class "0" for samples with a toxin value below the LOQ (Table 1). Data were divided into (1) a training set, consisting of 1057 samples, collected before 2019 (of which 155 classified as

class "1"); (2) a test set, consisting of 429 samples, collected after 2019 (of which 89 classified as class "1").

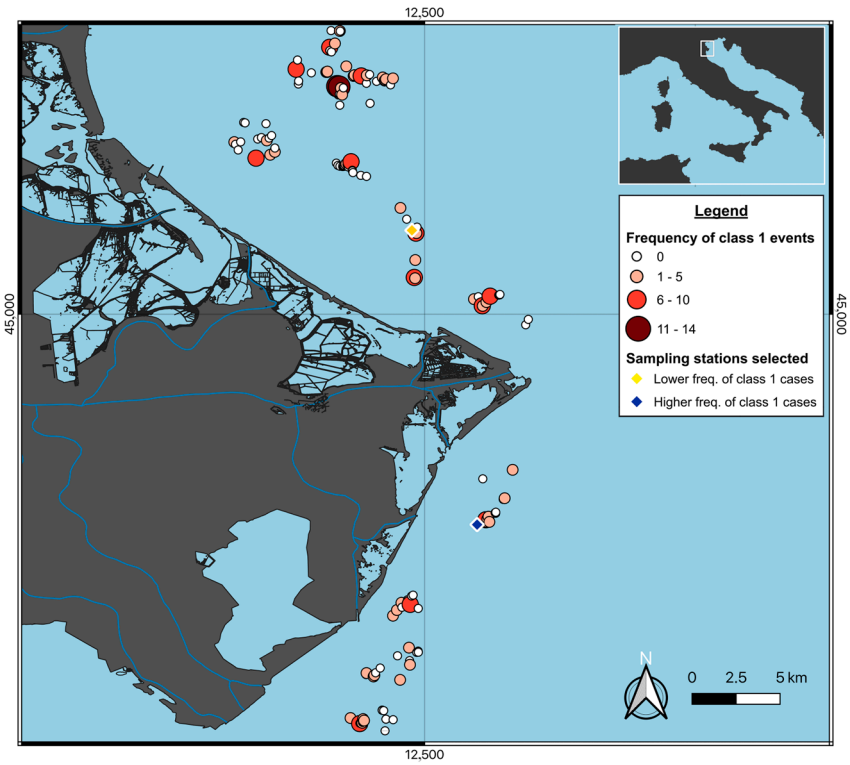


Figure 1. Map of offshore mussel sampling stations (2016–2020) in the marine area in front of the Po River Delta. Sampling stations are displayed according to the frequency of bioaccumulation events (≥40·µg kg^{-1} of OA equivalents). Two sampling stations have been also selected for being characterised by the lowest (=0; yellow diamond) and the highest (=244; blue diamond) frequency of positive events during the five years of data collection.

Table 1. Data classification. Class "1": ≥40 µg·kg^{-1} of OA equivalents; class "0": <40 µg·kg^{-1} of OA equivalents.

Year	Observations	Class "1"	Class "0"
2016	312	17	295
2017	391	130	261
2018	354	8	346
2019	306	54	252
2020	123	35	88
Total	1486	244	1242

A total of 15 physical and chemical parameters measured during the 45 days prior to each mussel sampling were associated with each OA quantification essay. The time window was chosen to be as wide as possible to cover the entire temporal development of a *Dinophysis* spp. bloom and the consequent bioaccumulation of OA in the mussel flesh, according to the available literature [22,23]. Parameters were selected according to their reasonable causal correlation with *Dinophysis* spp. blooms, and were: chlorophyll concentration ([Chl], mg·m^{-3}); Dinophyta contribution to total [Chl] (mg·m^{-3}), calculated according to [26]; wind speed, split into eastern and northern components of wind direction (m·s^{-1}); thickness of water column mixed layer (m); NH_4 concentration ([NH_4], mol·L^{-1}); NO_3 concentration ([NO_3], mmol·m^{-3}); dissolved O_2 concentration ([O_2], mol·L^{-1}); PO_4 concentration ([PO_4], mmol·m^{-3}); NO_3/PO_4 ratio; daily precipitations (mm·day^{-1}); salinity (PSU); surface water temperature (°C); eastern and northern components of sea current

($m \cdot s^{-1}$); wind pressure on the sea surface (Pa); and solar radiation ($Wh \cdot m^{-2}$). Data were extracted from satellite observations (https://data.marine.copernicus.eu/products, accessed on 29 March 2023) and, where necessary to increase resolution or estimate missing data, were combined with a custom oceanographic model (AQUAX, ColomboSky, Rome, Italy). Daily precipitations and solar radiation were obtained from the Copernicus atmospheric model (https://atmosphere.copernicus.eu, accessed on 29 March 2023). To reduce the background noise and, at the same time, emphasise the most useful information to describe the phenomenon and improve the performance of the classification algorithm, a series of transformations of the environmental parameters were applied. For each parameter, the weekly minimum, maximum, mean, variance, skewness, and kurtosis were calculated and used in the model.

ML algorithms can provide inadequate performance in the presence of highly unbalanced datasets (i.e., when the data of one class are very numerous or scarce compared with those of the other class) [27]. Being that the considered dataset was highly unbalanced (dominance of class "0", 83.6%), the class "1" category for the training step was oversampled using Synthetic Minority Oversampling Technique (SMOTE) [28]. Different ML algorithms were evaluated with a model selection procedure (3 stratified k-fold cross-validation) to identify the best performing one. In this context, the selected optimisation metric was the Mathews correlation coefficient (MCC). The model selection procedure led to defining the LightGBM algorithm as the most suitable [29]. The LightGBM algorithm is a recent algorithm known for combining a tree structure (which allows a good interpretation of the results) with a very high performance [30].

The trends from 2016 to 2020 of the environmental parameters of two sites, selected to be characterised by the highest and the lowest frequency of "class 1" events (represented in Figure 1 as blue and yellow diamonds, respectively), were compared throughout the non-parametric Kolmogorov–Smirnov test.

3. Results and Discussion

Dinophyceae blooms pose a serious threat to the economic and environmental sustainability of mussel farming in the northern Adriatic Sea. The loss of the product caused by a high-intensity HAB phenomenon can amount to up to 30% [31], with consequent economic shortfalls. Furthermore, growing mussels up to commercial size, but not being able to harvest and market them, represents an environmental cost. For example, consider the case study documented in Martini et al. [32], relative to a mussel farm in the northern Adriatic Sea which produces about 400 tons per year, the overall CO_2 equivalent was approximately 49,200 tons per year (0.12 kg CO_2 eq$\cdot kg^{-1}$ mussels). Assuming a product loss of 30%, with the same energy and material consumption, the carbon footprint would increase by about 50% (0.18 kg CO_2 eq$\cdot kg^{-1}$ mussels).

For these reasons, the importance of improving and developing tools to predict such events is paramount. The present study aimed at developing an ML predictive model for OA bioaccumulation in Mediterranean mussels.

3.1. Frequency and Distribution of OA Bioaccumulation Events from 2016 to 2020

The highest number of class "1" events was observed in 2017 (N = 130), followed by 2019 (N = 54) (Figure 2A), whereas 2018 and 2016 were characterised by the lowest occurrence, with OA concentrations always <160 $g \cdot kg^{-1}$ of OA equivalents (Figure 2B). From 2016 to 2019, all (100% in 2016 and 2018) or most (98% in 2019) of the class "1" events occurred in autumn. In 2017, however, 44 class "1" events (34% of the total) were registered in the summer. In 2020, class "1" events were almost equally distributed throughout the year, with a slight prevalence in summer. The most severe events of OA bioaccumulation were registered in 2019 (Figure 2B), with a concentration higher than 600 $\mu g \cdot kg^{-1}$ of OA equivalents.

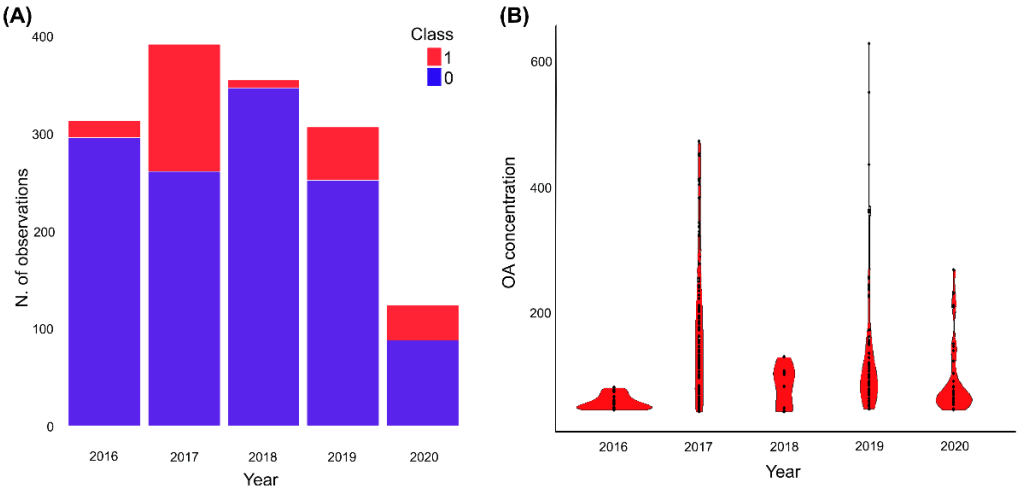
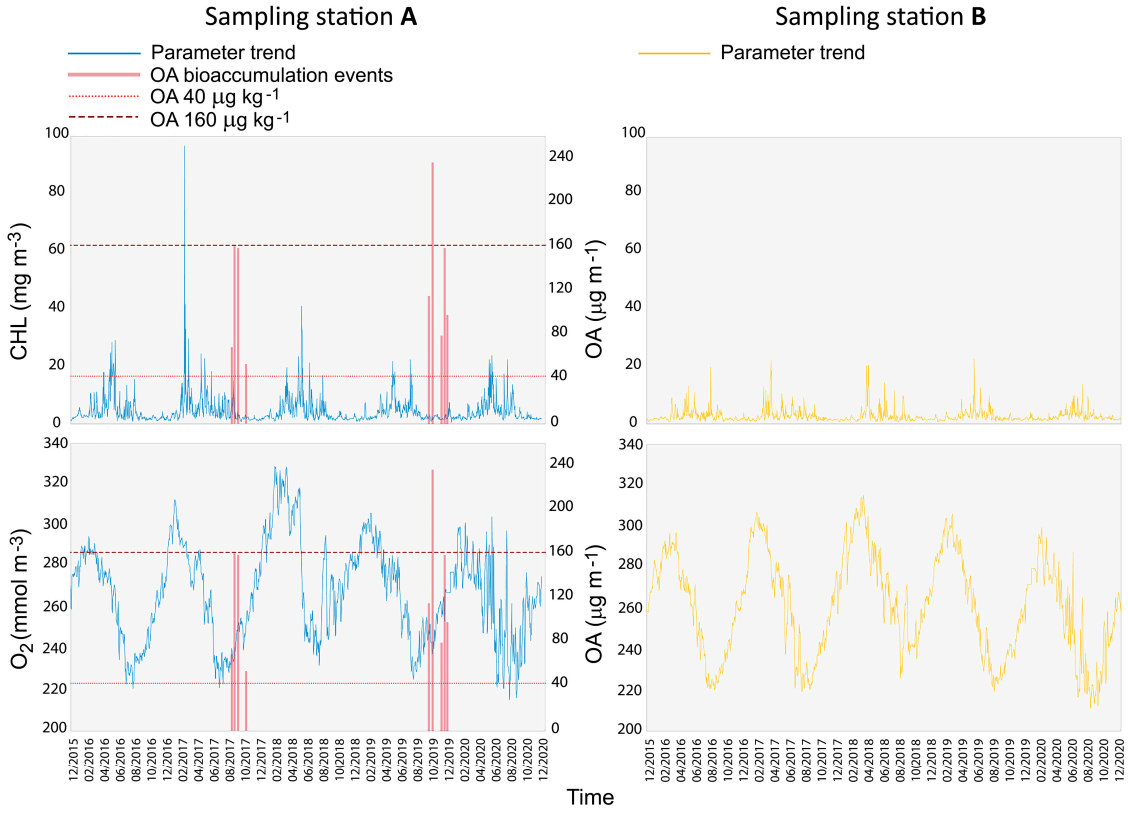


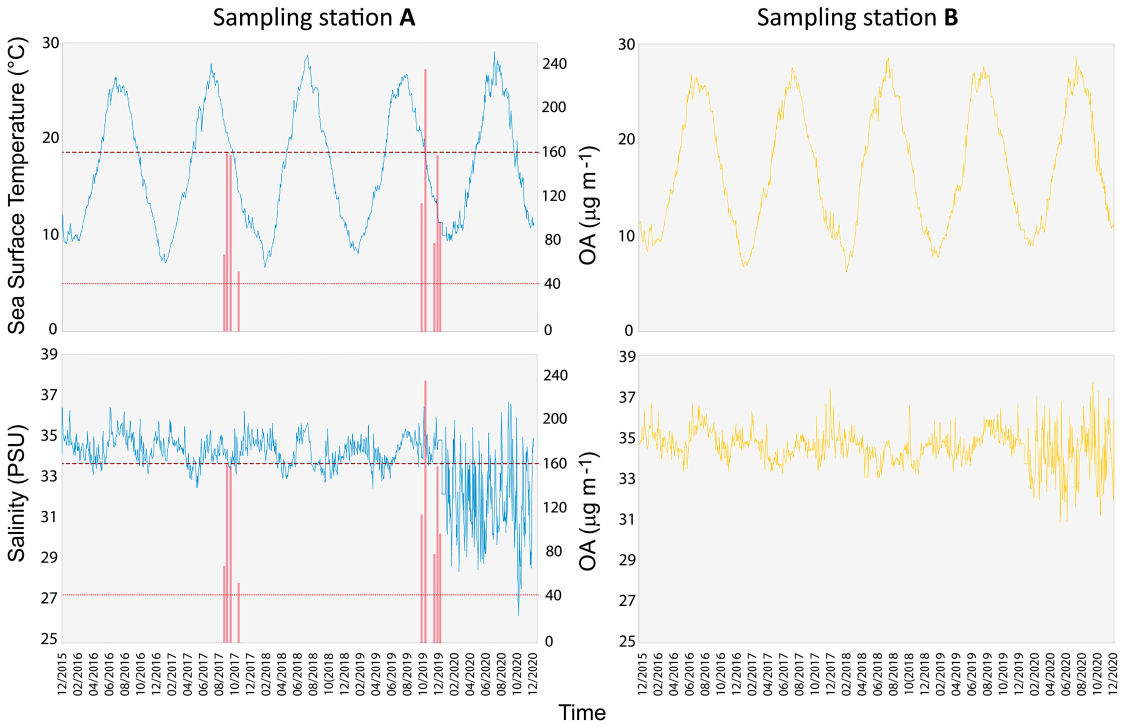
Figure 2. (**A**) Class "0" and class "1" observations for each year. (**B**) Violin plot of okadaic acid concentration (OA > 40 µg·kg^{-1}) in mussels for each year.

When two sites (geolocalised in Figure 1), characterised by the highest and the lowest frequency of class "1" events, were compared (Figure 3), the trends from 2016 to 2020 of some of the selected parameters were notably and significantly different: chlorophyll concentration (highest: 4.47 ± 4.78, lowest: 2.74 ± 2.29 (mg·m^{-3}, mean ± S.D.); Kolmogorov–Smirnov $p < 0.001$), ocean mixed-layer thickness (highest: 10.99 ± 0.27, lowest: 12.17 ± 0.48 (m, mean ± S.D.); Kolmogorov–Smirnov $p < 0.001$), and nutrient concentration (highest: NO_3 = 29.96 ± 13.14, PO_4 = 0.34 ± 0.15; lowest: NO_3 = 17.61 ± 5.39, PO_4 = 0.19 ± 0.07 (mmol·m^{-3}, mean ± S.D.); Kolmogorov–Smirnov $p_{NO_3} < 0.001$, $p_{PO_4} < 0.001$).

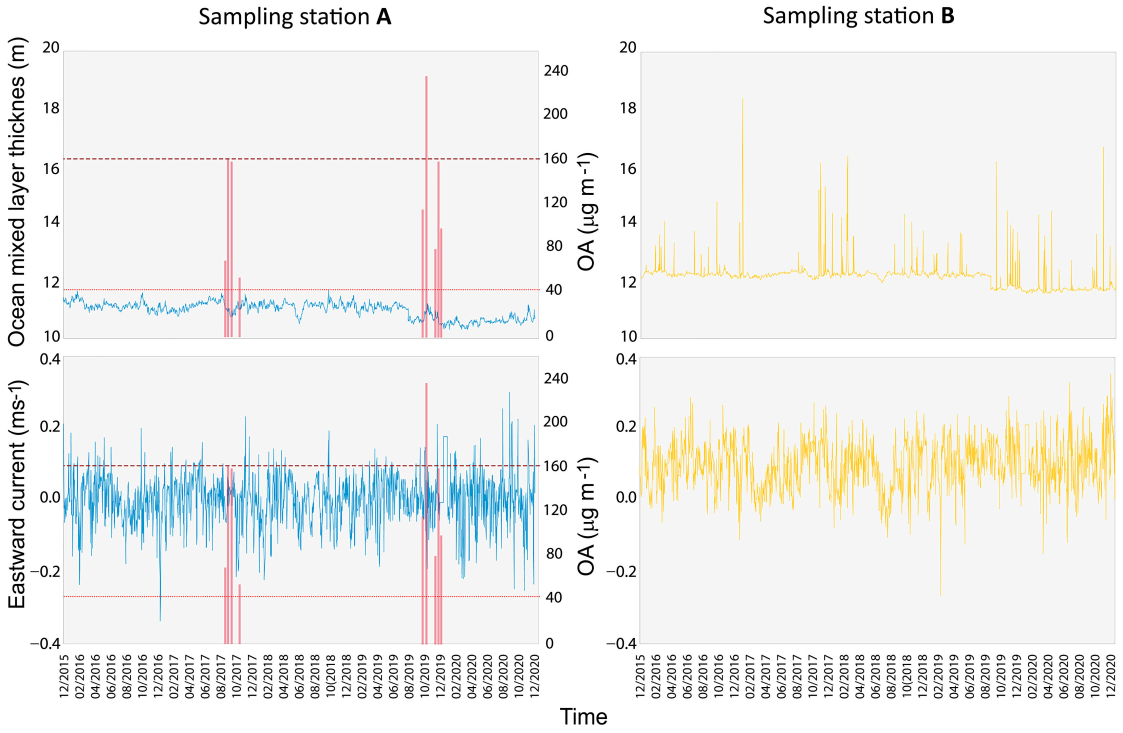


Panel 1: Chlorophyll, Dissolved oxygen.

Figure 3. *Cont.*

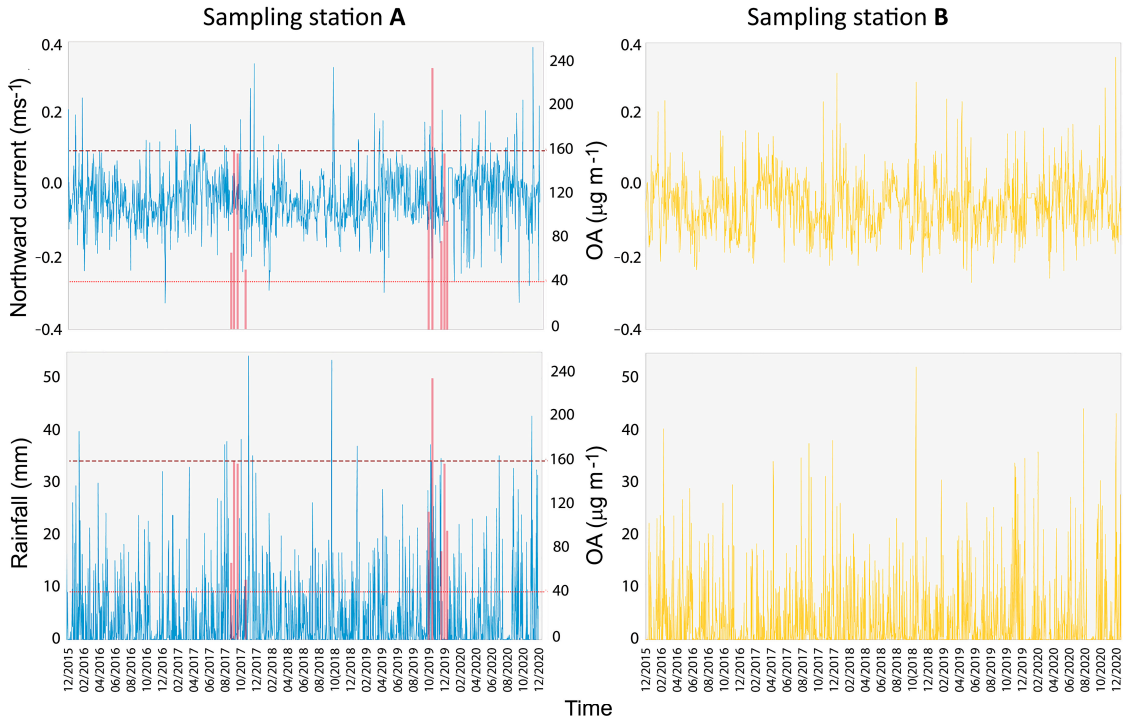


Panel 2: Sea Surface Temperature, Salinity.

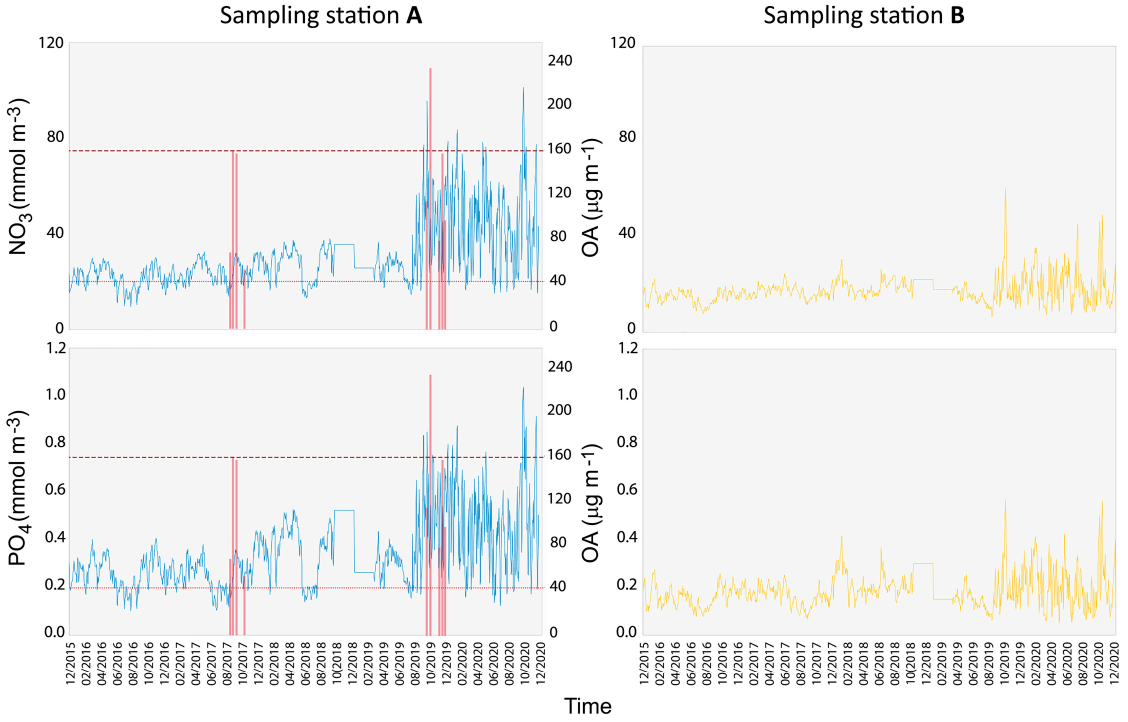


Panel 3: Ocean mixed-layer thickness, Eastward current.

Figure 3. *Cont.*



Panel 4: Northward current, Rainfall.



Panel 5: Nitrates (NO_3), Phosphates (PO_4).

Figure 3. Time distribution and intensity of class "1" events from 2016 to 2020 in sampling stations A and B, selected for being respectively characterised by the highest frequency of positive and the lowest (=0) one (see Figure 1 for the geolocalisation of sampling stations). Investigated parameters are: Panel 1, Chlorophyll concentration (CHL) and Dissolved oxygen (O_2); Panel 2, Sea Surface Temperature and Salinity; Panel 3, Ocean mixed-layer thickness and Eastward current; Panel 4, Northward current and Rainfall; Panel 5, Nitrates (NO_3) and Phosphates (PO_4).

The entire dataset of the values of each environmental parameter associated with the OA bioaccumulation events (2016–2020) used to feed the classification algorithm is fully reported in Table S1.

3.2. Results of the Classification Algorithm

A Confusion Matrix was used to summarise the results of the classification algorithm, broken down into correct (TRUE) and incorrect (FALSE) predictions (Figure 4).

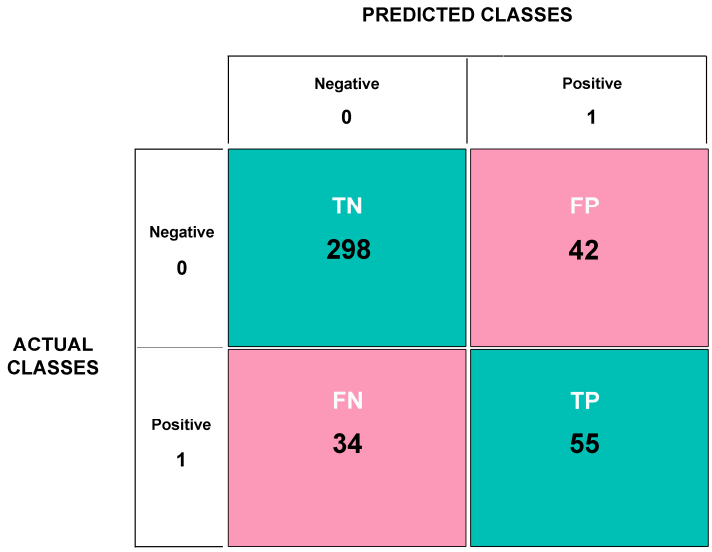


Figure 4. Confusion Matrix obtained from the application of LightGBM to the test set. TN: True Negative (predicted values correctly predicted as negative); TP: True Positive (predicted values correctly predicted as positive); FN: False Negative (positive values predicted as negative); FP: False Positive (negative values predicted as positive).

The accuracy of the Confusion Matrix was calculated as:

$$\text{Accuracy (A)} = (TP + TN)/(TP + TN + FP + FN)$$

and was found to be equal to 0.82 (82%). False negatives were mainly associated with levels of OA < 100 μg·kg^{-1}, while true positives had higher mean toxin values (139 μg·kg^{-1}); thus, the algorithm tended to classify low concentrations of OA as class "0" events. Indeed, though all samples were processed according to the same analytical procedure, concentration values <100 μg·kg^{-1} were affected by less accuracy than concentrations >100 μg·kg^{-1}, due to the different quality-control strategy applied to the screening and confirmatory test. Toxin concentrations for class "0" samples (<40 μg·kg^{-1}) were not available, as the LOQ of the analytical method was 40 μg·kg^{-1}. Thus, it was not possible to examine in-depth false positives (e.g., concentrations just below the LOQ value classified as positive by the model).

The parameters' (features) importance plot evidenced the ten most significant variables for the prediction of OA bioaccumulation events with LightGBM, according to split importance (how many times the nodes split on the feature) (Figure 5). Since none of the features measured during the two weeks before the event improved the accuracy of the classification algorithm, these were removed from the analysis and not used for the training. The most relevant feature analysed was the minimum value in wind stress, occurring three weeks prior to the event, followed by the minimum value in dissolved oxygen concentration six weeks prior to the event. Apart from the minimum in wind stress, in the third week prior to the event the minimum in the eastward wind, the variance and kurtosis of sea surface temperature, and the maximum in chlorophyll concentration and nitrate to phosphate ratio were the variables that most influenced the accuracy.

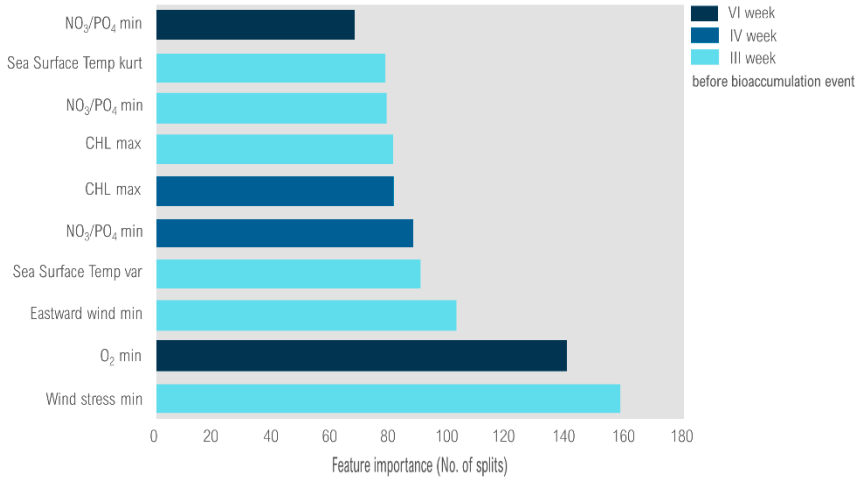


Figure 5. Feature importance plot. min: minimum; max: maximum; kurt: kurtosis; var: variance.

These 10 features are specific to the case study analysed, also considering the specific subdivision between test and training samples implemented.

The environmental factors that drive increases in *Dinophysis* spp. abundance are poorly understood. Researchers often fail in the attempt to describe the relationship between the increase in *Dinophysis* abundances and single environmental variables, because the response is often non-linear and/or exhibits abrupt changes. Notwithstanding this, it was found that temperature and salinity are among the main factors affecting dinoflagellate community structure [33], and, using a Logistic Generalised Linear Model, Singh et al. [34] found sea surface temperature, sea surface salinity, pH, total suspended solids, and N:P ratio as prominent drivers of *Dinophysis* abundance in the south-eastern Arabian Sea. It should also be taken into consideration that favourable environmental parameters could indeed trigger the blooms of non-toxic *Dinophysis* species. Even if OA accumulation in mussels would not occur in such cases, the initiation of red tides could occur, which could lead to severe anoxic (even lethal) conditions for local communities, including fish or shellfish.

Hou et al. [35] and Luz et al. [36] pointed out that, even though some factors could be considered reliable predictors of bloom triggering, algal blooms are started by a complex interaction between several abiotic factors (e.g., sea water temperature; concentrations of nitrates, phosphates, and O_2; and salinity). Moreover, small-scale physical processes have been demonstrated to be very significant in the triggering of HABs. For example, off the French coast, a "thin layer" (ca. 10 cm) of *D. acuminata* is frequently observed at the thermocline [37]. The turbulent mixing of the water column may damage phytoplankton cells, often disrupting colonies and causing the end of the toxic bloom [38–40].

The increase in chlorophyll concentration and water column stratification plays a primary role in causing *Dinophysis* spp. blooms [35,41,42], even if several studies demonstrated that measurements of primary productivity are poorly informative when a single toxic species is of interest [6]. *Dinophysis* spp. growth occurs in a wide temperature range (16–30 °C). However, temperature peaks above 25 °C are reported as foregoing algal blooms' onset [43,44]. It has been previously observed that temperature rises above 25 °C about 30 days before the bloom onset and drops during the bloom itself [45].

Conversely, salinity does not seem to represent a primary factor influencing *Dinophysis* spp. blooms and the consequent OA accumulation in bivalves [13].

A critical factor in bloom development is nutrient supply (nitrogen, phosphorus, silicate, and micronutrients). Concentrations of inorganic nutrients, i.e., nitrates (NO_3) and phosphates (PO_4), have been frequently correlated with the volume of dinoflagellates at the peak of their bloom: a negative correlation was observed between the stability of the concentrations of these compounds and the cell counts in the month preceding the bloom [46]. However, *Dinophysis* spp. have a mixotrophic nature, obtaining organic nutrients by ingesting detritus, bacteria, and other phytoplankton [47–49].

The purpose of this study was to develop a machine learning (ML) predictive model for OA bioaccumulation in farmed Mediterranean mussels using inputs from satellite oceanographic data. The goal was to deliver shellfish farmers a tool for enhancing productivity and profitability by reducing product losses and protecting, at the same time, human health. Being able to predict toxin bioaccumulation in shellfish above the legal threshold would help to plan harvesting decisions and mitigation strategies in a timely manner.

Several statistical and ML forecasting tools have been developed to inform the shellfish industry and limit damages, improve mitigation measures, and reduce production losses. Since HABs are the source of the biotoxins responsible for shellfish contamination, research efforts have been dedicated to the development of early warning systems focusing on HAB modelling (e.g., [18–22]).

However, HAB predictions may fail to estimate shellfish contamination, as HAB-producing species vary greatly in toxin production and models often fail to find direct relationships between HABs and environmental parameters, also considering that the dynamics of accumulation/elimination in shellfish vary among species.

Compared with previous studies, the present work, using an ML tool, investigates the relationship between the probability of OA bioaccumulation in mussels and a large set of combined environmental parameters, recorded at different production sites during a wide time window (45 days) preceding the bioaccumulation events. This approach could represent a promising perspective for fine-tuning predictive models for OA bioaccumulation in farmed shellfish.

A few studies adopted a similar approach, applying ML to the concentration of Paralytic Shellfish Poisoning (PSP) toxins in shellfish. For instance, Harley et al. [50] used several environmental parameters, such as sea surface temperature, salinity, and air temperature to predict the concentration of PSP toxins in blue mussels. The authors created a Random Forest model to classify shellfish samples above and below a toxicity threshold one week in advance. Although the approach used was similar to the one presented in this study, the classification accuracy obtained was lower (<50% vs. 82%). However, a reliable comparison of the different performances between the two approaches is difficult due to the lack of a common benchmark.

Grasso et al. (2019) [51] forecasted shellfish biotoxin contamination by predicting the concentration of PSP toxins in blue mussels (*M. edulis*) from one to ten weeks in advance. The authors used four years of weekly toxin data and configured the model to mimic an image classification task. Four classification categories were created based on toxin concentrations, and a Feed Forward Neural Network (FFNN) was used to perform the classification task. The model was able to accurately predict closure-level toxic events at a two-week advanced notice. The authors proposed to forecast toxin concentration by using past trends of toxin concentrations. The present study, instead, used environmental parameters, accounting for large-scale abiotic drivers of toxin bioaccumulation.

This study shows that the ML method applied can predict toxin accumulation in mussels based on remotely sensed environmental parameters with good classification accuracy.

4. Conclusions

Machine learning presents an interesting potential for developing predictive tools for water quality parameter monitoring and aquaculture management systems. Here, this technology was applied to the development of a computational tool to support marine aquaculture activities. It exploited remote sensing data for the prediction of risk related to OA in shellfish to provide clear indicators for the aquaculture industry and the management bodies in charge. Moreover, combining toxin concentration in the shellfish forecasting model with accurate field testing would allow regulatory sampling to focus on critical periods when toxin concentrations in shellfish are indicated to increase above and decline below the regulatory threshold. This targeted approach could reduce the overall frequency of official sampling activities while providing both competent authorities and the industry with the tools to ensure consumers' protection.

Further steps to improve the performance of the model are foreseen. Firstly, the training dataset shall be expanded to include a higher number of positive observations (class 1, >40 $\mu g \cdot kg^{-1}$). Secondly, further developments should include data regarding the density of OA-producing species to investigate the contribution of these dinoflagellate populations to the predictive power of our early warning system. Improving the integration of monitoring activities and modelling approaches for applications in HAB early warning systems is necessary to enhance predictive power.

Moreover, further physical, biological, and atmospheric parameters should be considered and added to the classification model to improve model performances.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su15118608/s1, Table S1: Environmental parameter's values associated with each OA bioaccumulation event (2016–2020).

Author Contributions: Conceptualisation, F.C. (Fabrizio Capoccioni), L.B. and D.P.; methodology, F.C. (Federica Colombo) and C.M.; software, F.C. (Federica Colombo) and C.M.; validation, L.B., L.C. and M.T.; formal analysis, C.M., R.N. and D.P.; investigation, L.B., F.C. (Fabrizio Capoccioni) and D.P.; resources, D.P.; data curation, L.B., L.C., M.T. and D.P.; writing—original draft preparation, F.C. (Fabrizio Capoccioni) and D.P.; writing—review and editing, L.B., A.M., R.N., N.T., F.C. (Federica Colombo) and C.M.; supervision, F.C. (Fabrizio Capoccioni), L.B. and D.P.; project administration, F.C. (Fabrizio Capoccioni) and D.P.; funding acquisition, F.C. (Fabrizio Capoccioni) and D.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Italian Ministry of Agriculture and Forestry, projects VALUE-SHELL (n. J86B19001870007) and AQUADATA2 (n. J89C20000010001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, D.P., upon reasonable request.

Acknowledgments: The authors would thank the "Consorzio Cooperative Pescatori del Polesine O.P. S.C.Ar.L." for technical support and, in particular, Emanuele Rossetti for his precious help and his availability in sharing his deep professional knowledge. The authors would also thank Piergiorgio Fumelli (Azienda ULSS5 Polesana), who collected the samples used in this study, and Claudia Casarotto (Istituto Zooprofilattico Sperimentale delle Venezie), who validated the spatial data. We are grateful to Thomas Moranduzzo (ColomboSky) for his invaluable help during remote sensing data analysis and the evaluation phase.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Rabalais, N.N.; Turner, R.E.; Diaz, R.J.; Justic, D. Global change and eutrophication of coastal waters. *ICES J. Mar. Sci.* **2009**, *66*, 1528–1537. [CrossRef]
2. Grattan, L.M.; Holobaugh, S.; Morris, J.G. Harmful algal blooms and public health. *Harmful Algae* **2016**, *57*, 2–8. [CrossRef]
3. Hallegraeff, G.M. Ocean climate change, phytoplankton community responses, and harmful algal blooms: A formidable predictive challenge. *J. Phycol.* **2010**, *46*, 220–235. [CrossRef]
4. Kovats, R.S.; Valentini, R.; Bouwer, L.M.; Georgopoulou, E.; Jacob, D.; Martin, E.; Rounsevell, M.; Soussana, J.-F. Europe. Contribution of Working Group II to the Fifth Assessment Report of IPCC. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2014; pp. 1267–1326, ISBN 978-1-107-05816-3.
5. Smayda, T.J.; Reynolds, C.S. Strategies of marine dinoflagellate survival and some rules of assembly. *J. Sea Res.* **2003**, *49*, 95–106. [CrossRef]
6. Anderson, D.; Cembella, A.; Hallegraeff, G. Progress in understanding harmful algal blooms: Paradigm shifts and new technologies for research, monitoring, and management. *Annu. Rev. Mar. Sci.* **2012**, *4*, 143–176. [CrossRef]
7. Andersen, P. *Design and Implementation of Some Harmful Algal Monitoring Systems*; IOC Technical Series; UNESCO: Paris, France, 1996; Volume 44.

8. Sournia, A. Atlas du phytoplancton marin. In *Introduction, Cyanophycées, Dictyochophyées, Dinophycées et Raphidophycées*; Editions du CNRS: Paris, France, 1986; Volume I, p. 219.
9. France, J.; Mozetič, P. Ecological characterization of toxic phytoplankton species (*Dinophysis* spp., Dinophyceae) in Slovenian mariculture areas (Gulf of Trieste, Adriatic Sea) and the implications for monitoring. *Mar. Pollut. Bull.* **2006**, *52*, 1504–1516. [CrossRef]
10. Reguera, B.; Riobó, P.; Rodríguez, F.; Díaz, P.A.; Pizarro, G.; Paz, B.; Franco, J.M.; Blanco, J. *Dinophysis* toxins: Causative organisms, distribution and fate in shellfish. *Mar. Drugs* **2014**, *12*, 394–461. [CrossRef]
11. Nielsen, L.T.; Krock, B.; Hansen, P.J. Production and excretion of okadaic acid, pectenotoxin-2 and a novel dinophysistoxin from the DSP-causing marine dinoflagellate *Dinophysis acuta*—Effects of light, food availability and growth phase. *Harmful Algae* **2013**, *23*, 34–45. [CrossRef]
12. Sedmak, B.; Fanuko, N. Occurrence of *Dinophysis* spp. and toxic shellfish in the Northern Adriatic. *J. Appl. Phycol.* **1991**, *3*, 289–294. [CrossRef]
13. Pistocchi, R.; Guerrini, F.; Pezzolesi, L.; Riccardi, M.; Vanucci, S.; Ciminiello, P.; Dell'Aversano, C.; Forino, M.; Fattorusso, E.; Tartaglione, L.; et al. Toxin Levels and Profiles in Microalgae from the North-Western Adriatic Sea—15 Years of Studies on Cultured Species. *Mar. Drugs* **2021**, *10*, 140–162. [CrossRef]
14. Sidari, L.; Cok, S.; Cabrini, M.; Tubaro, A.; Honsell, G. Temporal distribution of toxic phytoplankton in the gulf of Trieste (Northern Adriatic Sea) in 1991 and 1992. In *Harmful Marine Algal Blooms*; Lassus, P., Arzul, G., Erard-Le Denn, E., Gentien, P., Marcaillou-Le Baut, C., Eds.; Lavoisier: Paris, France, 1995; pp. 231–236.
15. Poletti, R.; Cettul, K.; Bovo, F.; Milandri, A.; Pompei, M.; Frate, R. Distribution of toxic dinoflagellates and their impact on shellfish along Northwest Adriatic Coast. In *Harmful Algae*; Reguera, B., Blanco, J., Fernandez, M., Wyatt, T., Eds.; Xunta de Galicia: Santiago, Spain; IOC UNESCO: Santiago de Compostela, Spain, 1998; pp. 88–90.
16. Della Loggia, R.; Cabrini, M.; Del Negro, P.; Honsell, G.; Tubaro, A. Relationship between *Dinophysis* spp. in seawater and DSP toxin in mussels in the Northern Adriatic Sea. In *Toxic Phytoplankton Blooms in the Sea*; Smayda, T.J., Shimizu, Y., Eds.; Elsevier: Amsterdam, The Netherlands, 1993; pp. 483–488.
17. Bernardi Aubry, F.; Berton, A.; Bastianini, M.; Bertaggia, R.; Baroni, A.; Socal, G. Seasonal dynamics of *Dinophysis* in coastal waters of the NW Adriatic Sea. *Bot. Mar.* **2000**, *43*, 423–430.
18. Hattenrath-Lehmann, T.K.; Marcoval, M.A.; Mittlesdorf, H.; Goleski, J.A.; Wang, Z.; Haynes, B.; Morton, S.L.; Gobler, C.J. Nitrogenous nutrients promote the growth and toxicity of *Dinophysis acuminata* during estuarine bloom events. *PLoS ONE* **2015**, *10*, e0124148. [CrossRef]
19. Velo-Suárez, L.; Gutiérrez-Estrada, J.C. Artificial neural network approaches to one-step weekly prediction of *Dinophysis acuminata* blooms in Huelva (Western Andalucía, Spain). *Harmful Algae* **2007**, *6*, 361–371. [CrossRef]
20. Valbi, E.; Ricci, F.; Capellacci, S.; Casabianca, S.; Scardi, M.; Penna, A. A model predicting the PSP toxic dinoflagellate *Alexandrium minutum* occurrence in the coastal waters of the NW Adriatic Sea. *Sci. Rep.* **2019**, *9*, 4166. [CrossRef]
21. Yasumoto, T.; Murata, M.; Oshima, Y.; Sano, M.; Matsumoto, G.K.; Clardy, J. Diarrhetic shellfish toxins. *Tetrahedron* **1985**, *41*, 1019–1025. [CrossRef]
22. Ajani, P.; Larsson, M.E.; Rubio, A.; Bush, S.; Brett, S.; Farrell, H. Modelling bloom formation of the toxic dinoflagellates *Dinophysis acuminata* and *Dinophysis caudata* in a highly modified estuary, southeastern Australia. *Estuar. Coast. Shelf Sci.* **2016**, *183*, 95e106. [CrossRef]
23. Harred, L.B.; Campbell, L. Predicting harmful algal blooms: A case study with *Dinophysis ovum* in the Gulf of Mexico. *J. Plankton Res.* **2014**, *36*, 1434–1445. [CrossRef]
24. Caroppo, C. Autoecology and morphological variability of *Dinophysis sacculus* (Dinophyceae: Dinophysiaceae) in a Mediterranean lagoon. *J. Mar. Biolog. Assoc.* **2001**, *81*, 11–21. [CrossRef]
25. Lessard, E.J. The trophic role of heterotrophic dinoflagellates in diverse marine environments. *Mar. Microb. Food Webs* **1991**, *5*, 49–58.
26. Di Cicco, A.; Sammartino, M.; Marullo, S.; Santoleri, R. Regional Empirical Algorithms for an Improved Identification of Phytoplankton Functional Types and Size Classes in the Mediterranean Sea Using Satellite Data. *Front. Mar. Sci.* **2017**, *4*, 126. [CrossRef]
27. Jeni, L.A.; Cohn, J.F.; De La Torre, F. Facing Imbalanced Data—Recommendations for the Use of Performance Metrics. In Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction 2013, Geneva, Switzerland, 2–5 September 2013; pp. 245–251.
28. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *JAIR* **2002**, *16*, 321–357. [CrossRef]
29. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
30. Yun, J.; Guangyu, S.; Quanche, C.; Min, Z.; Huixian, Z.; Mujeeb, U.R. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. *IEEE Access* **2019**, *7*, 28309–28318.
31. Rossetti, E.; (Consorzio Cooperative Pescatori del Polesine, Scardovari, Italy). Personal communication, 2023.
32. Martini, A.; Calì, M.; Capoccioni, F.; Martinoli, M.; Pulcini, D.; Buttazzoni, L.; Moranduzzo, T.; Pirlo, G. Environmental performance and shell formation-related carbon flows for mussel farming systems. *Sci. Total Environ.* **2022**, *831*, 154891. [CrossRef] [PubMed]
33. D'Costa, P.M.; Anil, A.C.; Patil, J.S.; Hegde, S.; D'Silva, M.S.; Chourasia, M. Dinoflagellates in a mesotrophic, tropical environment influenced by monsoon. *Estuar. Coast. Shelf Sci.* **2018**, *77*, 77–90. [CrossRef]

34. Singh, A.; Harding, K.; Reddy, H.R.V.; Godhe, A. An assessment of *Dinophysis* blooms in the coastal Arabian Sea. *Harmful Algae* **2014**, *34*, 29–35. [CrossRef]
35. Hou, L.T.; Wang, B.S.; Lai, C.C.; Chen, T.Y.; Shih, Y.Y.; Shiah, F.K.; Ko, C.Y. Effects of mixed layer depth on phytoplankton biomass in a tropical marginal ocean: A multiple timescale analysis. *Earth's Future* **2022**, *10*, e2020EF001842. [CrossRef]
36. Luz, L.F.G.; Noernberg, M.A. Spectral changing during a massive bloom event of *Dinophysis* cf. *acuminata* in south Brazil. *Reg. Stud. Mar. Sci.* **2022**, *56*, 102681.
37. Gentien, P.; Lunven, M.; Lehaitre, M.; Duvent, J.L. In situ depth profiling of particle sizes. *Deep-Sea Res.* **1995**, *42*, 1297–1312. [CrossRef]
38. Pollinger, U.; Zemel, E. In situ and experimental evidence of the influence of turbulence on cell division processes of *Peridinium cinctum* forma *westii* (Lemm.) Lefevre. *Br. J. Phycol.* **1981**, *16*, 281–287. [CrossRef]
39. Thomas, W.H.; Gibson, C.H. Effects of small-scale turbulence on microalgae. *J. Appl. Phycol.* **1990**, *2*, 71–77. [CrossRef]
40. Berdalet, E. Effects of turbulence on the marine dinoflagellate *Gymnodinium nelsoni*. *J. Phycol.* **1992**, *28*, 267–272. [CrossRef]
41. Lima, M.J.; Relvas, P.; Barbosa, A.B. Variability patterns and phenology of harmful phytoplankton blooms off southern Portugal: Looking for region-specific environmental drivers and predictors. *Harmful Algae* **2022**, *116*, 102254. [CrossRef] [PubMed]
42. Hällfors, H.; Hajdu, S.; Kuosa, H.; Larsson, U. Vertical and temporal distribution of the dinoflagellates *Dinophysis acuminata* and *D. norvegica* in the Baltic Sea. *Boreal Environ. Res.* **2011**, *16*, 121–135.
43. Granéli, E.; Ferreira, C.E.L.; Yasumoto, T.; Rodrigues, E.M.; Neves, B. Sea urchins poisoning by the benthic dinoflagellate *Ostreopsis ovata* on the Brazilian Coast. In Proceedings of the 10th International Conference on Harmful Algae, St. Pete Beach, FL, USA, 20–25 October 2002; p. 113.
44. Ciminiello, P.; Dell'Aversano, C.; Fattorusso, E.; Forino, M.; Magno, G.S.; Tartaglione, L.; Grillo, C.; Melchiorre, N. The Genoa 2005 outbreak. Determination of putative palytoxin in Mediterranean *Ostreopsis ovata* by a new liquid chromatography tandem mass spectrometry method. *Anal. Chem.* **2006**, *78*, 6153–6159. [CrossRef]
45. Accoroni, S.; Glibert, P.M.; Pichierri, S.; Romagnoli, T.; Marini, M.; Totti, C. A conceptual model of annual *Ostreopsis* cf. *ovata* blooms in the northern Adriatic Sea based on the synergic effects of hydrodynamics, temperature, and the N:P ratio of water column nutrients. *Harmful Algae* **2015**, *45*, 14–25. [CrossRef]
46. Telesh, I.; Schubert, H.; Skarlato, S. Abiotic stability promotes dinoflagellate blooms in marine coastal ecosystems. *Estuar. Coast. Shelf Sci.* **2021**, *251*, 107239. [CrossRef]
47. Jeong, H.J.; Kim, J.S.; Kim, J.H.; Kim, S.T.; Seong, K.A. Feeding and grazing impact of the newly described heterotrophic dinoflagellate *Stoeckeria algicida* on the harmful alga *Heterosigma akashiwo*. *Mar. Ecol. Prog. Ser.* **2005**, *295*, 69–78. [CrossRef]
48. Jeong, H.J.; Yoo, Y.D.; Park, J.Y.; Song, J.Y.; Kim, S.T.; Lee, S.H.; Kim, K.Y.; Yih, W.H. Feeding by phototrophic red-tide dinoflagellates: Five species newly revealed and six species previously known to be mixotrophic. *Aquat. Microb. Ecol.* **2005**, *40*, 133–150. [CrossRef]
49. Park, M.G.; Kim, S.; Kang, Y.G.; Yih, W. First successful culture of the marine dinoflagellate *Dinophysis acuminata*. *Aquat. Microb. Ecol.* **2006**, *45*, 101–106. [CrossRef]
50. Harley, J.R.; Lanphier, K.; Kennedy, E.; Whitehead, C.; Bidlack, A. Random Forest classification to determine environmental drivers and forecast paralytic shellfish toxins in Southeast Alaska with high temporal resolution. *Harmful Algae* **2020**, *99*, 101918. [CrossRef] [PubMed]
51. Grasso, I.; Archer, S.D.; Burnell, C.; Tupper, B.; Rauschenber, C.; Kanwit, K.; Record, N.R. The hunt for red tides: Deep learning algorithm forecasts shellfish toxicity at site scales in coastal Maine. *Ecosphere* **2019**, *10*, e02960. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.