

Prediction of *Alexandrium* and *Dinophysis* algal blooms and shellfish contamination in French Mediterranean Lagoons using decision trees and linear regression: a result of 10 years of sanitary monitoring

Bouquet Aurélien ^{1,*}, Laabir Mohamed ², Rolland Jean-Luc ¹, Chomérat Nicolas ³, Reynes Christelle ^{4,5}, Sabatier Robert ^{4,5}, Felix Christine ⁶, Berteaux Tom ⁷, Chiantella Claude ⁷, Abadie Eric ^{1,8}

¹ MARBEC, Université de Montpellier, CNRS, Ifremer, IRD, 87 Avenue Jean Monnet, 34200 Sète, France

² Université de Montpellier, MARBEC, CNRS, Ifremer, IRD, 34095 Montpellier, France

³ IFREMER, Station de Biologie Marine, Place de la Croix, BP 40537, 29185 Concarneau Cedex, France

⁴ Institut de Génomique Fonctionnelle, IGF, Univ. Montpellier, CNRS, INSERM, 34094 Montpellier, France

⁵ Faculté de Pharmacie, Univ. Montpellier 34093 Montpellier, France

⁶ Université de Montpellier, MARBEC, CNRS, Ifremer, IRD, 87 Avenue Jean Monnet, 34200 Sète, France

⁷ Ifremer, Laboratoire environnement ressources du Languedoc Roussillon, 34200 Sète, France

⁸ IFREMER, Biodivenv, 79 route de pointe fort, 97231 Le Robert, France

* Corresponding author : Aurélien Bouquet, email address : aurelien.bouquet@ifremer.fr

mohamed.laabir@umontpellier.fr ; jean.luc.rolland@ifremer.fr ; nicolas.chomerat@ifremer.fr ; christelle.reynes@umontpellier.fr ; robert.sabatier@umontpellier.fr ; christine.felix@umontpellier.fr ; eric.abadie@ifremer.fr ; eric.abadie@ifremer.fr ; eric.abadie@ifremer.fr

Abstract :

French Mediterranean lagoons are frequently subject to shellfish contamination by Diarrhetic Shellfish Toxins (DSTs) and Paralytic Shellfish Toxins (PSTs). To predict the effect of various environmental factors (temperature, salinity and turbidity) on the abundance of the major toxins producing genera, *Dinophysis* and *Alexandrium*, and the link with shellfish contamination, we analysed a 10-year dataset collected from 2010 to 2019 in two major shellfish farming lagoons, Thau and Leucate, using two methods: decision trees and Zero Inflated Negative Binomial (ZINB) linear regression models. Analysis of these decision trees revealed that the highest risk of *Dinophysis* bloom events occurred at temperature <16.3°C and salinity <27.8, and of *Alexandrium* at temperature ranging from 10.4 to 21.5°C and salinity >39.2. The highest risk of shellfish contaminations by DSTs and PSTs occurred during the set of conditions associated with high risk of bloom events. Linear regression prediction enables us to understand whether temperature and salinity influence the presence of *Alexandrium* and affect its

abundance. However, *Dinophysis* linear regression could not be validated due to overdispersion issues. This work demonstrates the tools which could help sanitary management of shellfish rearing areas.

Highlights

► Robust predictions of the risk of HAB in French Mediterranean lagoons are provided ► Decision trees and linear regressions are used as complementary tools ► Temperature and salinity ranges can promote *Alexandrium* and *Dinophysis* blooms ► HAB and shellfish contamination occurs for the same set of environmental conditions

Keywords : *Alexandrium*, *Dinophysis*, Toxic blooms, Temperature, Salinity, Prediction

I. Introduction

The geographical expansion of harmful algae and their durable settlement in many marine ecosystems has become a major concern all over the world (Hallegraeff et al. 2003; Drake et al. 2005; Hallegraeff 2010; Casas-Monroy et al. 2015; Glibert 2020). Harmful algae (HA) proliferation or HABs (Harmful Algal Blooms) leads to observable negative effects on marine resources, economics and human health (Granéli and Turner, 2006). Many HA produce toxins that can contaminate marine resources and lead to human intoxication (Pulido 2016). For this reason, the proliferation of HAs in water is frequently monitored near shellfish farming areas (Anderson et al. 2001). Studies showed that blooms can be impacted by many environmental factors such as temperature and salinity (Anderson and Keafer, 1987; Laabir et al. 2011), light (Figuroa et al. 2018), oxygen concentration in sediment and nutrient availability (Collos et al. 2007, 2009; Xu et al. 2010; Ishikawa et al. 2014). Blooms are generally triggered by the synergic combination of several factors (Accoroni et al. 2015; Raso et al. 2012) among which temperature and salinity play a key role as they can influence every phases of bloom dynamics, including initiation, growth and termination (Figuroa et al. 2018; Granéli and Turner 2006). Many dinoflagellates including *Alexandrium* show planktonic and benthic forms, and initiation of the blooms depends on the success of cysts germination which occurs after a dormancy period. The duration of this phase was shown to be influenced by temperature, and some HA cysts germinate only within specific ranges of temperature (Anderson and Keafer 1987; Genovesi et al. 2009). Moreover, germination can be linked with water agitation leading to sediment resuspension (Laanaia et al. 2013) which is generally associated with variations in water temperature and salinity. Growth corresponds to biomass increase which duration and maximum concentration can be influenced by temperature and salinity along with other factors such as nutrients (Bill et al. 2016; Collos et al. 2007; Davis et al. 2015; Ishikawa et al. 2014; Laabir et al. 2011; Ralston et al. 2015; Raso et al. 2012; Xu

et al. 2010). Bloom termination happens when environmental conditions are no longer favourable, which can be attributed to nutrient depletion or inadequate temperature leading to mortality or encystment (Ralston et al. 2014; Verity et al. 1988).

As a consequence of these various and contrasted impacts of environmental factors on HAB dynamic, efforts have been made to develop early HABs warning systems based on environmental predictors. Conventionally, statistical numerical modelling has been used for HABs prediction (González Vilas et al. 2014). Through linear regression, these models can efficiently predict phytoplankton dynamics because they are capable of integrating many predictors and can be adapted to different data distributions (Davidson et al. 2016). Given the cyclic and seasonal pattern in HABs, they have mainly been investigated through time series modelling based on a range of variables measured over time (Cruz et al., 2021). For instance, several studies have applied linear regression to forecasting the blooms of *Pseudo-nitzschia* (Anderson et al. 2010; Lane et al. 2010; Seubert et al. 2013). The interest in alternatives like Machine Learning (ML) models has recently increased. ML methods consist of using computer systems that are able to learn and draw inference from data patterns without explicit instructions. One of the many advantages of ML methods is their ability to predict highly dynamic and complex phenomena and to handle big and non-linear data. A wide variety of machine learning approaches has been used over the past years to forecast HABs (Franks 2018; Cruz et al. 2021). Among them, decision trees (DTs) seem to be highly relevant tools. A DT is a tree-like structure in which the leaves represent outcome labels and the branches represent conjunctions of the input features that resulted in those outcomes. Aside from overcoming the problems due to missing data, they are easy to interpret since they provide concrete thresholds that can be used by marine stakeholders. In South Korea, Shin (2017) used DTs to forecast cyanobacteria blooms and (Park et al. 2011) applied this method to predict red tide blooms.

French Mediterranean lagoons are special ecosystems as they are semi-enclosed marine systems, containing many shellfish farming areas, and they are frequently subject to HAB events and particularly proliferation of the dinoflagellates *Alexandrium pacificum* and *Dinophysis acuminata* complex responsible for intoxication events threatening human health and resulting in economic losses (Deslous-Paoli et al. 1998; Collos et al. 2009; Laabir et al. 2011; Laanaia et al. 2013; Belin et al. 2021; Séchet et al. 2021). Until now, no machine learning approach has been developed to quantify and predict the risks of HABs proliferations and shellfish contamination in these ecosystems. Based on a ten-year dataset of parameters (temperature, salinity and turbidity) and of HA taxa, using both decision trees and a Zero-Inflated Negative Binomial regression model, this study aimed at providing robust tools to predict the risk of bloom of the two major HA genera in French Mediterranean lagoons, *Dinophysis* and *Alexandrium*, and the subsequent shellfish contamination in exploited semi-enclosed marine systems.

2. Materials and methods

2.1. Data recovery

Data were collected from the framework of the REPHY and REPHYTOX monitoring networks (REPHY - French Observation and Monitoring program for Phytoplankton and Hydrology in coastal waters 2019; REPHYTOX - French Monitoring program for Phytotoxins in marine organisms 2019).

Briefly, the data comes from field surveys carried out by Ifremer in two exploited French Mediterranean lagoons, Thau (7000ha, mean depth 4m, water temperature range : 3.9-29.6°C; salinity range : 27.4-42.9) and Leucate (5400ha, mean depth 1.9m water temperature range 0.9-29.3°C, salinity range: 8.4-42.3) (Fiandrino et al. 2017; Ladagnous and Le Bec 1997) from January 2010 to December 2020 (Fig 1). The concentration of phytoplankton species

was measured using the Utermöhl method (Utermöhl 1931) with a 10 mL Hydrobios chamber. The limit of quantification (LQ) and the threshold limit of detection (LD) are both 100 cell/L. Temperature, salinity, turbidity, total phytoplankton including the lipophilic toxins producer *Dinophysis* spp. and the paralytic shellfish toxins producer *Alexandrium* spp., were monitored twice a month at three stations in shellfish farming areas in the Thau lagoon (Marseillan, Bouzigues and Crique de l'Angle), and two stations in the Leucate lagoon (Parc and Grau) (Fig 1). Mussels and oysters were sampled in these stations twice a month. At least 1 kg of shellfish, composed of a minimum of ten individuals, were sampled randomly and Lipophilic toxins (Okadaic Acid (OA), Dinophysistoxins (DTXs) and Pectenotoxins (PTXs)) were measured by Liquid Chromatography coupled to tandem Mass Spectrometry (LC-MS / MS) (ANSES LSA-INS-0147, Standards and Reference Material were provided by the NRC-CNRC, Marine Analytical Chemistry Standards Program. Institute for Marine Biosciences). Each of the constitutive toxins was corrected by a TEF (Toxic Equivalent Factor) to account for its potential toxicity. The quantification of PSTs was carried out by mouse bioassay, calibrated and validated to give quantitative results (Anses PBM BM LSA-INS-0143).

2.3. Data processing and analyses

2.3.1. Preliminary processing

All data analyses were performed using R software (R Core Team 2020). Since *Alexandrium pacificum* and *Alexandrium tamarense* are difficult to distinguish by light microscopy, abundances of both species were added to each other and were thereafter named *Alexandrium tamarense/pacificum*. As for *Dinophysis*, regulatory monitoring considers all species that can be identified in the lagoons because they are all likely to contaminate shellfish. Thus, even though the *D. accuminata* complex represents the vast majority of the species found in Thau and Leucate, we chose to consider data at the genus level. The value of toxin concentrations

below the detectable threshold was estimated to be the detectable threshold divided by 2 since a non-detectable value does not mean that there is no toxin in shellfish. When toxins were detectable but not quantifiable, the value was estimated by the means of the detectable and quantifiable thresholds.

2.3.1. Exploratory analysis

Bivariate analyses were performed in order to identify which factors were mainly related to phytoplankton and toxin concentrations. Pairwise Spearman and Pearson coefficients were calculated between phytoplankton and environmental factors, and between toxins and environmental factors.

2.3.2. Building decision trees

Variable selection and encoding

Decision trees (DT) were used to predict the risks of HABs and of bivalve contamination considering temperature, salinity and turbidity as predictors. Prior to the analysis, *Dinophysis* and *A. tamarensis/pacificum* concentrations were converted to binary variables: the abundance of cells was encoded by 1 when the concentration exceeded 1000 cells/L for *A. tamarensis/pacificum* (value inducing the sanitary warning protocol) and 500 cells/L for *Dinophysis* (estimated value indicating a bloom, the sanitary warning protocol being induced by a concentration of cells \geq LD), and by 0 when the abundances were beyond these concentrations. Toxin concentrations were converted to categorical variables based on the regulatory thresholds: they were encoded by 0 when concentrations were below 80 μg per kg of shellfish flesh) for lipophilic toxins and below 340 $\mu\text{g}/\text{kg}$ for PSTs, by 1 when they were between 80 $\mu\text{g}/\text{kg}^{-1}$ and 160 $\mu\text{g}/\text{kg}$ for lipophilic toxins, and between 340 $\mu\text{g}/\text{kg}$ and 800 $\mu\text{g}/\text{kg}$ for PSTs, and by 2 when they were above 160 $\mu\text{g}/\text{kg}$ for lipophilic toxins and above 800 $\mu\text{g}/\text{kg}$ for PSTs.

Decision tree analysis

The DTs building process is described in Fig 2. DTs were set up with the *rpart package* (Therneau and Atkinson 2019) using the CART algorithm (Breiman et al. 1984). As a supplementary constraint, a minimum of 10 observations was imposed for a branch to become a branch node and there had to be at least 5 observations per leaf node. A maximum of 10 nodes per tree was imposed. In order to set up a robust tool: for each predicted variable, the experiment with DT was repeated 100 times; each time, the data set was split into two sets, 80 % made up the training set and 20% the testing set. The split algorithm allowed all the different pairs of training and testing sets to have approximately the same proportion of 0, 1 (and 2 for the toxins) in the predicted variables. The structures of the 100 trees were then compared in order to find the most common one. For each node the process consisted of:

- Finding the most frequent variable
- Finding the most frequent threshold for this variable
- Selecting the trees with this variable and threshold
- Repeating the process at the next children nodes for the selected trees only.

Secondly, the most common structure was determined and an average of the trees presenting this structure was then calculated. To this end, the mean of the number of observations for each node and the average proportions of 0, 1 or 2 were calculated. In order to test the trees' performance outside the training data sets, the selected structures were tested on the testing data sets. The average testing trees were then processed in the same way as for the average learning trees.

Finally, pruning analysis was applied in order to prevent overfitting by reducing the size of the trees and to prevent overtraining. Pruning a tree consists in removing the redundant and/or

less important nodes to improve its predictive accuracy. The pruning process followed these rules:

- The proportion of 1 or 2 in at least one of the nodes of the testing trees should be above 20%
- The number of observations in each node of the testing trees should be above 3

2.3.3. Linear models

Type of regression

Mixed linear models were set up to predict phytoplankton concentrations in the Thau and Leucate lagoons based on hydrological factors. The distributions of HA abundance were characterized with a very high frequency of 0, meaning that HA were not detected most of the time. In addition, abundances are discrete variables, hence the Zero Inflated Negative Binomial (ZINB) model (Greene 1994) appeared to be the most suitable regression for this dataset. ZINB regression is used for data that exhibits excess zeros and over dispersion. In Zero Inflated (ZI) models (Lambert 1992), theory suggests that the excess zeros are generated by a separate process from the count values and that the excess zeros can be modelled independently. Thus, the ZINB model has two parts, the negative binomial count model and the logit model for predicting excess zeros. The process explaining the excess zeros could correspond to an absence of HA cyst germination, when count values would represent the proliferation of vegetative cells once germination had occurred. Type 2 ZINB were used for the models because the Akaike Information Criteria (AIC) (Akaike 1974) were lower than type 1 ZINB. As the variables were time series data sampled in different stations, it was necessary to check the influence of the stations, and to check that the relationships between the variables would still be significant without this influence. Mixed models were thus performed, using the stations as grouping factors. Only random-intercept models were performed because random-slope models did not converge. The significance of Random effects was tested by performing an Anova between the models with and without random

effects. Random effects were kept for the logistic and/or the negative binomial part only when it was significant it when it allowed the model to fit the data better; this was shown on the residuals graphs.

Variable selection and encoding

The model building process is described in Fig 3. A model was set up to predict HA concentrations using the hydrological factors as predictors (i.e the predictive variables) and showing a Spearman or a Pearson coefficient with a p -value lower than 20%. The linearities of the relationships were tested by performing maximum likelihood tests between models before and after having transformed the factors to categorical variables. When non-linear relationships were found, the variables were encoded as categorical variables, the levels being determined by the ranges depicted by the DT and by visual examination of the distribution graphics. Several combinations of variables and of levels were tested and the AIC was used to select the best fitting model. The Incident Risk Ratio (IRR) was calculated for the negative binomial part ($IRR = \exp(\text{coefficient})$). An $IRR > 1$ means an increase of the average microalgae's abundance. As for the logit part, a particularity in the ZI model is that the coefficient and the Odds Ratio ($OR = \exp(\text{coefficient})$) characterize the probability of zero, in this case the probability of absence of the HABs. To make it easier to understand, we choose to depict the Odds Ratio of the presence of the algae: $OR(\text{Presence}) = 1/OR$. An $OR(\text{presence}) > 1$ means an increase of the probability of presence of the alga.

Validation

The validation of the model was based on the standardized residuals simulated with the *DHARMA* package (Hartig 2020). The resulting simulated residuals graphics obtained with this package are interpretable the same way as the residuals graphics for simple linear models (Hartig 2020). Specialized goodness-of-fit tests were performed on the simulated residuals

with the DHARMA package: the Kolmogorov-Smirnov test, a non-parametric dispersion test based on the standard deviation of the simulated residuals, testing whether the simulated dispersion is equal to the observed dispersion, and an outlier test that tests whether there are more simulation outliers than expected. Finally, Nakagawa's conditional and marginal R^2 were calculated to show how much of the variability was explained by the global model and by random effects.

3. Results

3.1. Exploratory data analysis

Correlation coefficients between phytoplankton abundance, abiotic factors and toxins in Thau and Leucate lagoons were calculated as shown in Tab 1A and 1B respectively. In both lagoons, the coefficients showed significant correlations between phytoplankton and abiotic factors: In Thau, the Spearman coefficient showed a positive correlation ($p.val < 0.001$) between *A. tamarense/pacificum* and salinity (Sal.). In Leucate, *Dinophysis* was negatively correlated to temperature (Temp.) and Sal. ($p.val < 0.001$), whereas it was positively correlated to Temp. in Thau ($p.val < 0.001$) (Tab. 1). *Dinophysis* was also positively correlated to turbidity (Turb.) (Spearman coefficient 0.15, $p.val < 0.001$) (Tab. 1). When significant, the correlations between *A. tamarense/pacificum*, *Dinophysis* and abiotic factors were quite low, absolute values ranged from 0.15 to 0.31 (Tab. 1). In both lagoons, whenever significant, the Spearman coefficients between the environment and toxic phytoplankton species were always stronger than the Pearson coefficients, suggesting non-linear relationships. Furthermore, toxins were highly correlated to the abundance of the microalgae which produce them. The significant correlations between *A. tamarense/pacificum* and PSTs in Thau ranged from 0.41 to 0.58 and the Pearson coefficients were stronger than the Spearman coefficients, with $p.val < 0.001$, suggesting a linear relationship (Tab. 1). The lipophilic toxins (OA, PTX, DTX)

were positively correlated with the abundance of *Dinophysis*, the significant coefficients ranged from 0.45 to 0.69 ($p.val < 0.00$) (Tab 1). Furthermore, abiotic factors seemed to have similar impact on the toxins as on the microalgae producing them. In Thau, PSTs were positively correlated to Sal. (significant correlation coefficients ranged from 0.3 to 0.41, $p.val < 0.05$) (Tab. 1). In Leucate, lipophilic toxins were negatively correlated to both Temp. (coefficients ranging from -0.38 to -0.57) and Sal. (significant coefficients ranging from -0.14 to -0.35) and positively correlated to Turb. (significant coefficient ranging from 0.27 to 0.28, $p.val < 0.001$) (Tab. 1). The Spearman coefficients showed that *A. tamarensis/pacificum* and *Dinophysis* abundances were non-linearly but positively correlated with the abundance of dinoflagellates and diatoms, and that these correlations were stronger for *A. tamarensis/pacificum* (coefficients ranging from 0.18 to 0.29, $p.val < 0.001$) than for *Dinophysis* (coefficients ranging from 0.08 to 0.09, $p.val < 0.05$) (Tab. 1).

3.2. Decision trees

Decision trees were able to predict when blooms of *A. tamarensis/pacificum* and *Dinophysis* could occur (Fig 4). Turb. was never kept in pruned DTs and appeared thus to be a non-determinant risk factor in *Dinophysis* and *A. tamarensis/pacificum* blooms. Results showed that, for Temp. between 10.4°C and 21.5 °C and Sal. >39.2, the mean risk of a bloom of *A. tamarensis/pacificum* occurrence was over 50% (Fig 4A). The mean risk was less than 12% when Sal. is <39.2 and, even with a high salinity, it was near 0% when Temp. was <10.4 °C. The decision tree for *Dinophysis* (Fig 4B) indicated that Sal. <27.8 and Temp <16.3°C led to a mean risk of bloom over 50%. This situation was however quite rare (only 21 observations over the 10 years monitoring). Sal. >32.0 along with Temp. <16.3°C also led to a high probability of bloom (in average 26% of the testing datasets). Temp. >16.3 °C led to a low risk (on average 4.2% of the testing data sets). No low threshold of temperature was found for *Dinophysis*.

Decision trees were able to predict when shellfish contaminations could occur (Fig 5). The trees showed that toxins were usually found in animals in the conditions promoting a bloom event (Fig 5). Again, Turb. was never kept in the pruned DTS. In Thau, the risk of contamination of mussels and oysters by PSTs was significant for seawater Sal. > 37.7 and 39.9 , respectively (Fig 5 A/B). Low Temp. also promoted risks of shellfish contamination. For mussels, Temp. $< 18.5^{\circ}\text{C}$ combined with Sal. > 37.7 yielded a total mean risk of contamination of 92% (44.9% between $340\ \mu\text{g}/\text{kg}$ and $800\ \mu\text{g}/\text{kg}$ (labelled 1 on the figure) and 47.1% exceeding $800\ \mu\text{g}/\text{kg}$ (labelled 2 on the testing data sets). For oysters, Temp. $< 16^{\circ}\text{C}$ combined with Sal. > 39.9 led to a 57% mean risks of contamination, the latter being mainly comprised between $340\ \mu\text{g}/\text{kg}$ and $800\ \mu\text{g}/\text{kg}$ (52.3% of the observations in the testing dataset). It is possible that the contaminations were rare below a low Temp. threshold, but were not represented on the decision trees because of a low number of observations. In Leucate, the risks of mussel and oyster lipophilic contamination was higher when Temp. is $< 16.3^{\circ}\text{C}$ and $< 13.5^{\circ}\text{C}$, respectively (Fig 5 C/D). The 16.3°C threshold showed that mussels would become contaminated as soon as Temp. allowed *Dinophysis* blooms. Under this Temp., the risk of contamination was high independently of the Sal. (56.7% and 31.1% risk of a contamination over $160\ \mu\text{g}/\text{kg}$ (labelled 2 on the figure) for a Sal. respectively $>$ and < 33.5). In oysters, contamination by lipophilic toxins happened at Temp. $< 13.5^{\circ}\text{C}$ and Sal. < 27.8 that led to 30.7% mean risk of contamination (over $160\ \mu\text{g}/\text{kg}$) in the testing data sets. In all, these trees showed that the conditions in which oysters were contaminated were more restrictive than for mussels: the Temp. and Sal. ranges were systematically narrower for oysters.

3.3. Linear regression

In order to predict *A. tamarense/pacificum* abundance in Thau, a ZINB model was set up using two categorical predictors (temperature, split into four levels, and salinity, split into three levels) and one continuous predictor, turbidity (Tab 2). The splits have been selected

thanks to the results of decision trees and from visual examination of data distribution. The negative binomial part of the model, which depicted how predictors impact cell abundance, showed that the reference Temp. level, 14.6 to 21.5 °C, was associated with the highest average microalgal abundance. The IRR revealed that the average abundance was multiplied by 0.15 when Temp. was >25°C (p .val 4.38e-3), and by 0.010 when Temp. was <10.4°C (p .val <7.73e-5). A Sal. >36 was associated with the highest average abundance of *A. tamarensis/pacificum*: the average abundance was multiplied by 0.25 when Sal. was <36 compared with a Sal. >39.2, but no difference was shown when Sal. was comprised between 36 and 39.2. Turb. did not significantly impact *A. tamarensis/pacificum* abundance. It should be noted, before interpreting the logit part of the model, the coefficients were conversely related to the presence of the alga. In order to depict how the predictors affect the probability of presence of the alga, we choose to calculate OR (presence) = $1/(\exp(\text{coefficient}))$. This part of the model showed that the reference Temp. level was associated with the highest probability of presence of *A. tamarensis/pacificum*. For instance, the odds of its presence were multiplied by 0.14 when the temperature was <10.4°C and by 0.15 when it was >25. The reference Sal. (over 39.2) led to the highest likelihood of the presence of *A. tamarensis/pacificum*: the odds of the presence were multiplied by 0.21 when Sal. was <36 and by 0.44 when it was between 36 and 39.2. Turb. was not associated with the presence of this alga. Overall, what emerged from the two parts of the models was that the conditions in which *A. tamarensis/pacificum* usually reached higher abundances were generally the same as those in which the probability of finding it were the highest.

Random effects were significant for both parts of the model (p .val of the Anovas describing random effects for the log and negative binomial parts were respectively 1.82e-7 and 4.88e-3) which indicated that a part of the variability in *A. tamarensis/pacificum*'s presence and abundance was due to the location. The conditional Nakagawa R^2 was 0.402 whereas the

marginal Nakagawa R^2 was 0.206. The model thus described almost 40.2% of the data variability, with 19.6% (0.402-0.206) being explained by random effects.

The study of residuals (Fig 6) showed no significant problem: the QQplot curve was straight, the Kolmogorov-Smirnov and Dispersion tests were not significant (respectively p .val: 0.36 and 0.056). The graph of residuals against predicted values showed that residuals followed a uniform law (no particular shape was revealed) and the predicted quantile curves were aligned with the observed quantile curves. Five outliers were revealed by this graph, but did not invalidate the model (p .val of the outlier test: 0.088). No significant problems were detected through these graphics.

4. Discussion and conclusion

4.1. Discussion

In this work, we developed an innovative approach to forecast the risk of HABs and shellfish contamination occurring in the marine environment. Our study was based on a solid long-term database integrating ten years of sanitary monitoring results. After having explored the data through exploratory bivariate analyses, we used two complementary methods, DTs and statistical ZINB models, in order to reveal the relationships between environmental parameters such as temperature and salinity, HABs and shellfish contamination.

First, the exploratory bivariate analyses indicated an impact of temperature and salinity on HABs. These results were in accordance with several studies showing that temperature and salinity influence the occurrence of *Alexandrium* and *Dinophysis* blooms (Abdenadher et al. 2012; Anderson and Keafer 1987; Gomis et al. 1996; Itakura et al. 2002; Ninčević Gladan et al. 2008; Swan et al. 2018). The correlation coefficients were medium to low, but were not expected to be much higher as they depicted the influence of only one variable on the HA

abundances variability. The exploratory analyses showed thus that temperature and salinity were to be adequate predictors in following multivariate analyses.

By using decision trees, we determined the range of optimum temperature and salinity associated with high risks of blooms. The proportion of blooms with a salinity above 39.2 and temperature ranging between 10.4 and 21.5 for *A. tamarense/pacificum*, and when salinity was below 27.8 for *Dinophysis* spp. exceeded 50%, which could trigger safeguard measures of the reared shellfish. If the optimal ranges of temperature and salinity may appear wide, they only represent 9.7% and 3.1% of the data in Thau and Leucate, respectively, which has to do with the broad ranges of temperature and salinity in Mediterranean lagoons. The temperatures were congruent with previous observations showing that *A. tamarense/pacificum* blooms mainly occur from October to December and April to June but were rarely observed in summer and winter (Laabir et al. 2011, 2012). Laanaia (2013) highlighted that *Alexandrium* blooms occurred mainly during periods when salinity was below 39.3 from 2002 to 2006. Even though substantial risks could subsist outside the optimum ranges depicted by the DTs (11.9% risk of *A. tamarense/pacificum* bloom below 39.2), this study could thus reveal a change in *A. tamarense* populations composition together with a variation in intra-species physiological constants in Thau over the past few years. Our work also showed positive correlations between toxins in bivalves and the HAs producing them, as well as with the environmental factors that impact these algae. The optimal temperature and salinity ranges associated with high risks of shellfish contamination by phycotoxins were congruent with the ranges increasing the risks of blooms. In Mediterranean lagoons, all referenced *Dinophysis* species were potentially toxic. Recently, qPCR analyses the *Alexandrium* species in Thau lagoon showed that the proportion of non-toxic *Alexandrium* species during a bloom were very low (Sectox project, Eric Abadie, unpublished data). This could explain the high correlations between algae and toxins. The correlation coefficients and the DTs showed that

relationships between phytoplankton and toxins were stronger in mussels than in oysters. Mussels are considered as sentinel organisms for lipophilic toxins as they are known to be contaminated more easily than the other bivalves (Levinton et al. 2002). However, a strong inter-individual variability in the accumulation of toxins has been documented in oysters (Pousse et al. 2018), meaning that sanitary risks remain substantial even in shellfish other than mussels. Together, our results suggested that the DT is a tool which can provide relevant information that will be useful in sanitary monitoring.

Through linear regression, the model confirmed the DT results obtained with *A. tamarense/pacificum* and showed how abiotic factors influence either its presence or its cell abundance when present. The optimal set of parameters chosen by the CART algorithm was significantly associated with the maximum average abundance and odds of presence of the algae. Moreover, this model specified the sets of parameters more accurately by distinguishing significantly more levels of temperature and salinity than the DT. It also revealed that the optimal sets of parameters are similar for the presence of the microalgae and for their abundance when they were present: this could indicate that germination and proliferation occurred in the same conditions (Temperature from 14.6 to 21.5°C and Salinity >39.2). Actually, from 2010 to 2020, most of the substantial blooms occurred within these conditions, as the bloom of October 2015 (820000 cells/mL) and October 2017 (35000 cells/mL). Few other blooms occurred for slightly lower salinities, as in June 2011 (250000 cells/mL; Salinity 36.6). This is however not in total contradiction with the model since the latter shows that, even though a salinity over 39.2 increases the risk of *Alexandrium* presence, its abundance when it's present could be increased by a salinity over 36. This brings to light that, if this model shows conditions where risks are very important, it does not mean that blooms cannot occur outside them. The model predicting *Dinophysis* spp. abundance could not be validated due to overdispersion issues. In contrast to *Alexandrium*, *Dinophysis* feeds on

living preys and acquires its kleptochloroplasts originating from a cryptophyte through predation on the ciliate *Mesodinium rubrum* (Park et al. 2006; Reguera et al. 2012). Hence, its growth depends on the presence of cryptophyte preys and ciliates, which themselves depend on environmental factors like light intensity (Kim et al. 2008; González-Gil et al. 2010). The mixotrophic status of *Dinophysis* spp. and the influence of its predator make its dynamics more complex and difficult to integrate in a linear regression model. Further studies integrating more variables, abiotic as well as biotic, may allow to validate a model for *Dinophysis* spp.

Several other environmental factors could have been integrated to increase the model's performances regarding the mechanisms of blooms. Indeed, temperature and salinity changes may occur along with variations in other factors such as nutrients, water column conditions and residence time which can trigger or deplete blooms. In addition to the direct impact of temperature toward *Alexandrium* cyst germination (Anderson and Keafer 1987; Genovesi et al. 2009), this factor is likely to be related with water stability or turbulence. In this regard, winds seem to be a critical factor in lagoons by promoting turbulence and resuspension and decreasing temperature (Laanaia et al. 2013). As instance, an important bloom of *Alexandrium* occurred in 2004 for a salinity of 36.5 (outside the optimal salinities depicted by the DTs): this bloom was triggered by cysts germination resulting from water column agitation by strong winds, also inducing a drop in temperature (Laanaia et al. 2013). Moreover, salinity and temperature may be associated with nutrient availability as they can be linked with nitrogenous and phosphorous fluxes between compartments (organic and inorganic nutrients in water column, organisms and sediments) (Bougis 1974; Boynton and Kemp 1985; Grenz et al. 1992). Many studies showed that nutrient availability is positively associated with *Dinophysis* (Ajani et al., 2016) and *Alexandrium* growth (Collos et al., 2007; 2009; Natsuike et al., 2018) and that a lack of nutrients could lead to sexuality, and thus to the

decline of blooms (Anderson and Lindquist 1985). Unfortunately, nutrient measurements by the REPHY network were not performed at the same time and stations than those of this work. Further studies including nutrient availability in water would allow to understand if temperature and salinity have an indirect impact on cells abundance as their variation occur along nutrients fluxes. Oxygen and light intensity are also known to influence *Alexandrium* proliferations (Anderson and Keafer 1987; Keafer et al. 1992; Ajani et al. 2016; Valbi et al. 2019) and light increases toxin production by HA (Ogata et al. 1987). As nutrients, these factors were not measured during the period or sites considered here but could be integrated in further studies. We tried to integrate turbidity, although the linear regression showed no significant impact and it never ended up as a remaining variable in the pruned DTs. In the Mediterranean lagoons as in all aquatic ecosystems, turbidity can result from a large panel of environmental variations such as water turbulence (Baas et al. 2005) and surface winds (Arfi et al. 1993). These different factors could have various impacts on *Alexandrium* and *Dinophysis* distribution, which may explain why turbidity appears to be a bad predictor of the concentration of these species.

The approach developed here provided robust, easily interpretable, reliable results. At first, the DTs highlighted the set of parameters in which blooms and shellfish contamination were the most likely to happen, and provided an estimation of the risk within these sets. The models statistically confirmed the set for *A. tamarense/pacificum*, specified it by providing narrower ranges, and linked them to the mechanisms of blooms. The robustness of the DTs was ensured by repeating each experiment 100 times and calculating average DTs on both learning and testing datasets, thus providing a very robust tool. The number of observations used for toxins DTs was quite low, but the sets of parameters for phytoplankton and toxins DTs were congruent which tends to validate the results. By validating the model's predictions thanks to residuals graphs and to adequation, dispersion and outlier tests, this work confirmed

that temperature and salinity did have an impact on *A. tamarense/pacificum*'s presence and abundance. Moreover, including random effects in the model by grouping on the sampling sites allowed us to exclude the variability due to location, thus increasing its reliability.

4.2. Conclusion

This study is a first step in forecasting the risks due to HABs in French Mediterranean lagoons. Few variables were used, as they came from a network thought to trigger sanitary alerts based on few parameters. Models were set up for two lagoons, including Thau which is the fourth oyster farming area in France. They provided ranges of temperature and salinity triggering blooms, and these ranges could be linked to the mechanisms of blooms. The robust results could help stakeholders to manage the sanitary regulation of these lagoons in the future according to network measurements. The methods developed in this study could be used to investigate other data sets describing different ecosystems and integrating more biotic and abiotic variables, which could allow to increase the model's performance.

Acknowledgements

The authors would like to thank all members of the Laboratoire Environnement Ressource Languedoc Roussillon for their work during the REPHY (Observation and Surveillance Network for Phytoplankton and Hydrology in coastal waters) and REPHYTOX (Monitoring Network for Phycotoxins in marine organisms) surveys. The REPHY and REPHYTOX are supported by the French Ministry of Agriculture and Food.



[Download : Download high-res image \(2MB\)](#)

[Download : Download full-size image](#)

Fig. 1. Location of the REPHY and REPHYTOX monitoring stations in the French Thau and Leucate lagoons.

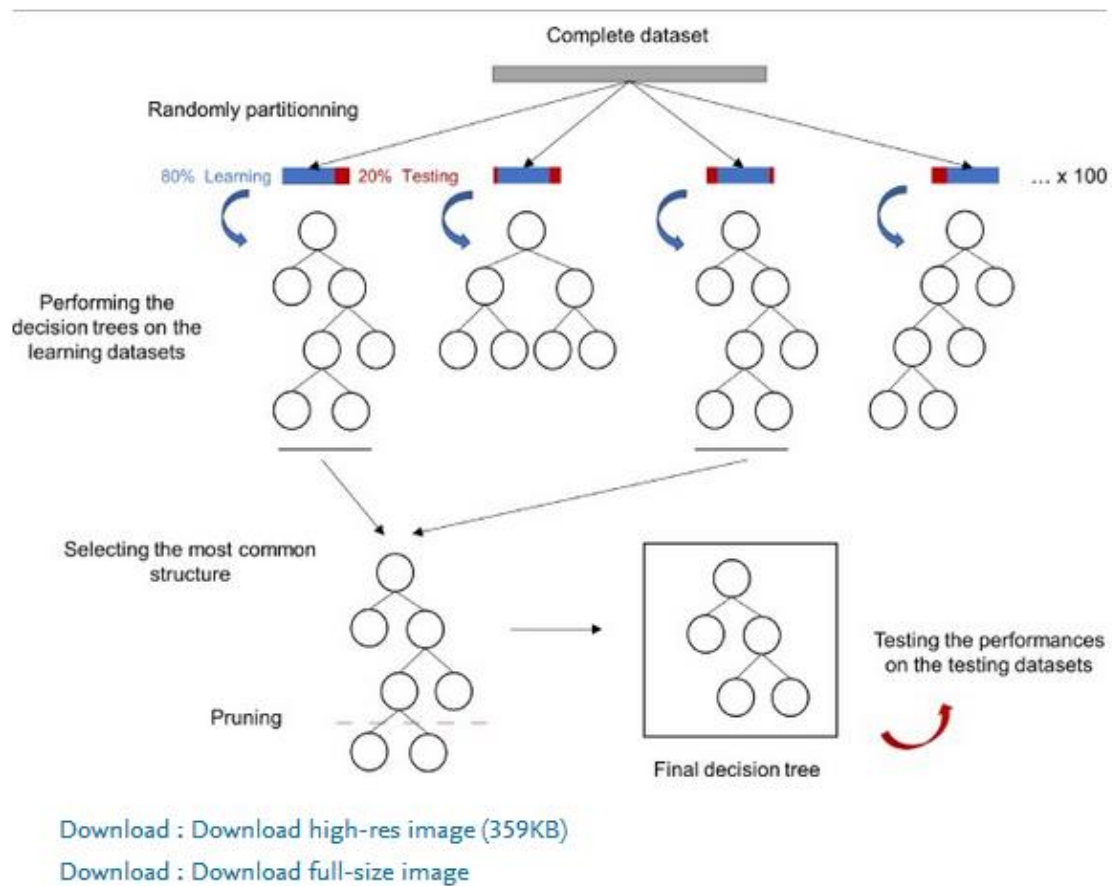


Fig. 2. Decision trees building process

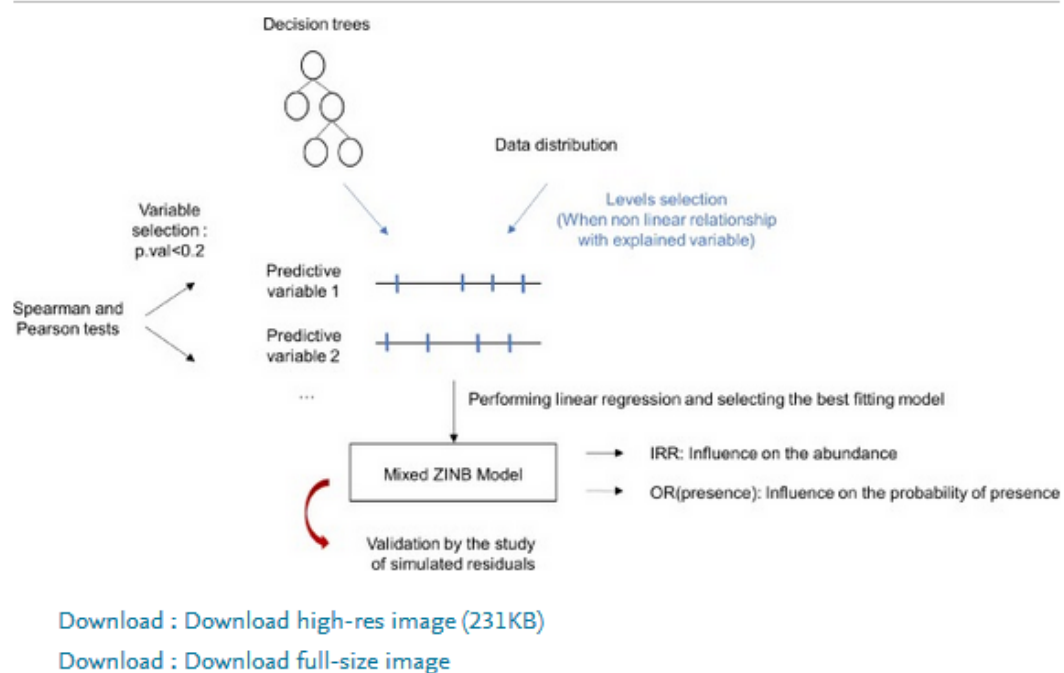


Fig. 3. Mixed Zero Inflated negative binomial model (ZINB) building process. IRR: Incident Risk Ratio. OR: Odds Ratio

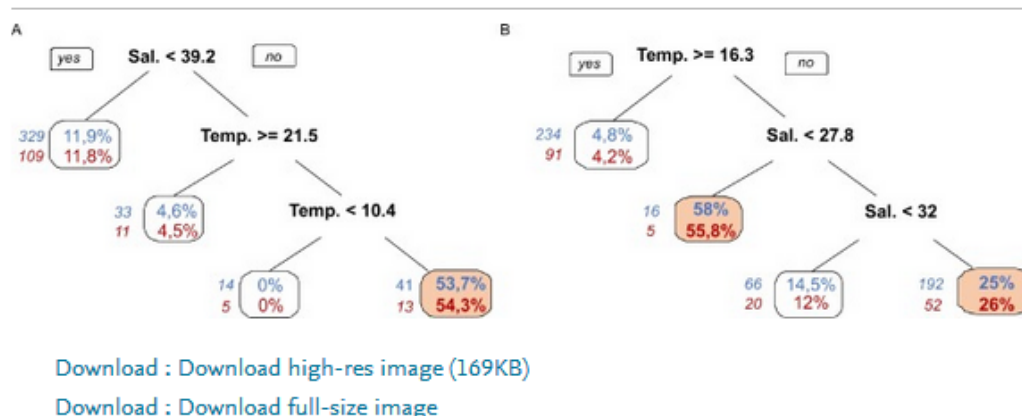
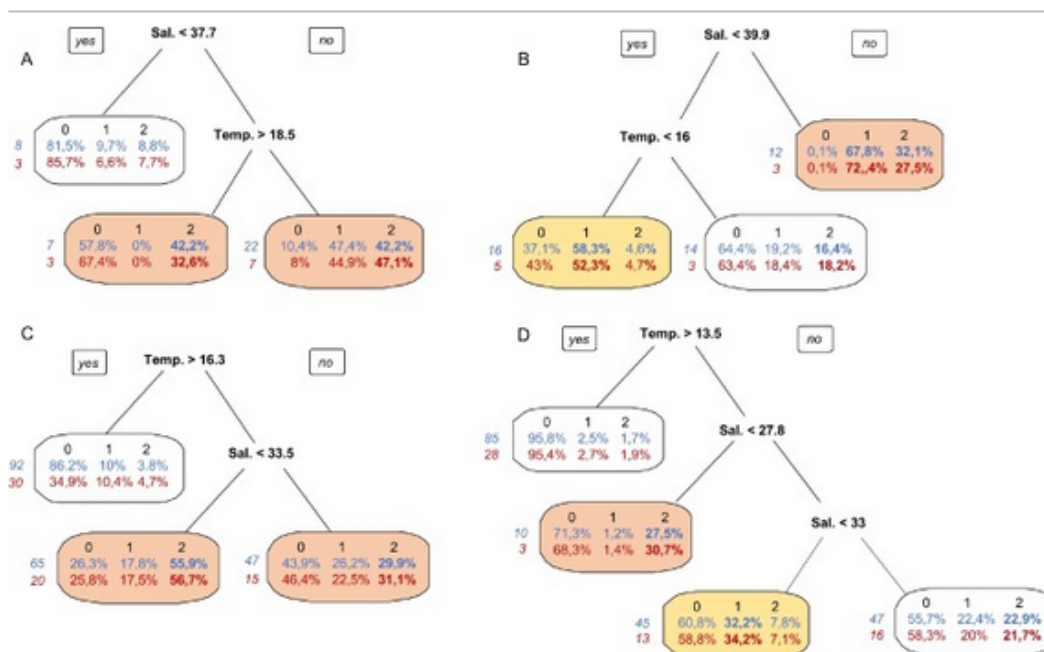


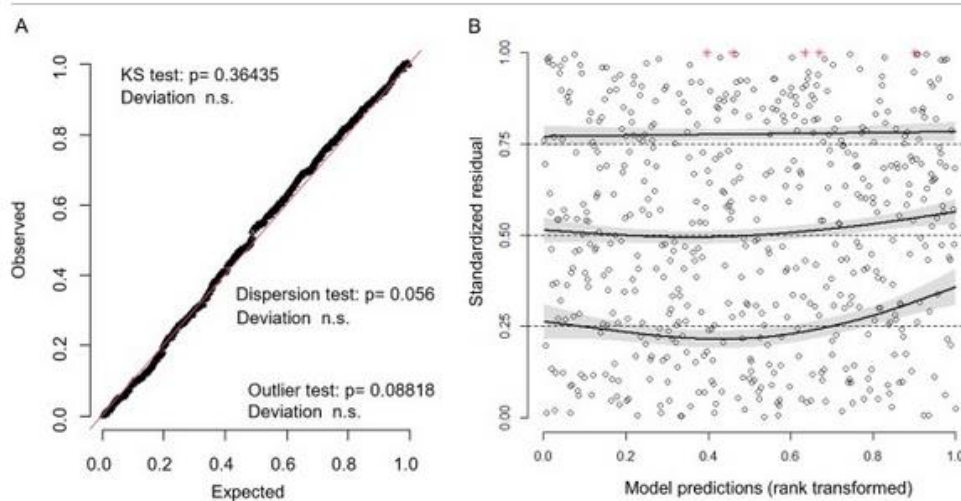
Fig. 4. Decision tree prediction for HABs events from 2010 to 2019 in French Mediterranean lagoons. Predicted variables for *A. tamarense/pacificum* in Thau (A) and for *Dinophysis* in Leucate (B). The mean numbers of observations in the learning and testing data sets are shown at each node, and the mean percentages of blooms within these observations. The percentages correspond to the mean proportion of bloom in the observations. The figures in italics show the mean number of observations at each leaf. The figures in blue and red show the average numbers and proportions in the learning and the testing datasets, respectively. Leafs are coloured in orange when the percentages of blooms exceed 25%. Sal.: Salinity. Temp.: Temperature.



[Download : Download high-res image \(419KB\)](#)

[Download : Download full-size image](#)

Fig. 5. Decision trees predicting intoxications in Thau and Leucate from 2010 to 2019. The predicted variables are A: PSTs in mussels in Thau, B: PSTs in oysters in Thau, C: lipophilic toxins in mussels in Leucate and D: lipophilic toxins in oysters in Leucate. The numbers (0,1 and 2) correspond to concentrations leading to 0: normal monitoring, 1: increased monitoring (medium contamination) and 2: closure of the farms (high contamination). The percentages correspond to the mean proportions of 0, 1 or 2 within the observations. The mean numbers of observation at each leaf are shown in italics. In blue and red figure the average numbers and proportions within the learning and the testing dataset are shown in blue and red figures respectively. Leaves are coloured in orange when the percentages of 2 exceed 25%, and in yellow when the percentage of 2 doesn't exceed 25% but the percentage of 1 does. Temp.: Temperature. Sal.: Salinity.



[Download : Download high-res image \(375KB\)](#)

[Download : Download full-size image](#)

Fig. 6. DHARMA residual diagnostics of the ZINB model predicting *A. tamarense/pacificum* abundances in the Thau lagoon. A: QQplot residuals. B: Residuals vs. predicted. KS test: Kolmogorov-Smirnov test.

Table 1 : Correlation coefficients between *A. tamarense/pacificum*, *Dinophysis*, diatoms, other dinoflagellates, abiotic factors (temperature, salinity and turbidity) and toxins in Thau and Leucate lagoons from 2010 to 2019.

Thau lagoon						
	<i>A. tamarense/pacificum</i>	Diatoms	Dinoflagellates	Sal.	Temp.	Turb.
<i>A. tamarense/pacificum</i>		-0.01 <u>0.18</u>	0.05 <u>0.29</u>	0.04 <u>0.17</u>	0.02 <u>0.02</u>	0.04 <u>0.05</u>
PST (m)	0.41 <u>0.17</u>	-0.11 <u>-0.23</u>	-0.2 <u>-0.05</u>	0.3 <u>0.33</u>	-0.18 <u>-0.38</u>	-0.14 -0.31
PST (o)	0.58 <u>0.44</u>	-0.18 <u>-0.19</u>	-0.12 <u>-0.06</u>	0.27 <u>0.41</u>	0.04 <u>-0.04</u>	0.15 <u>-0.16</u>
Leucate lagoon						
	<i>Dinophysis</i>	Diatoms	Dinoflagellates	Sal.	Temp.	Turb.
<i>Dinophysis</i>		-0.01 <u>0.09</u>	-0.04 <u>0.08</u>	-0.21 <u>-0.22</u>	-0.23 <u>-0.31</u>	0.03 <u>0.15</u>
OA (m)	0.45 <u>0.67</u>	-0.06 -0.15	-0.06 <u>-0.07</u>	-0.28 <u>-0.34</u>	-0.38 <u>-0.51</u>	0.08 <u>0.27</u>
OA+PTX+DTX (m)	0.46 <u>0.66</u>	-0.06 -0.15	-0.06 <u>-0.08</u>	-0.28 <u>-0.35</u>	-0.39 <u>-0.51</u>	0.08 <u>0.27</u>
OA (o)	0.5 <u>0.69</u>	-0.08 <u>-0.2</u>	-0.05 <u>-0.1</u>	-0.14 <u>-0.19</u>	-0.39 <u>-0.56</u>	0.13 <u>0.28</u>
OA+PTX+DTX (o)	0.52 <u>0.69</u>	-0.08 -0.22	-0.06 <u>-0.13</u>	-0.12 <u>-0.17</u>	-0.38 <u>-0.56</u>	0.12 <u>0.28</u>

Temp.: Temperature; Sal.: Salinity; Turb.: Turbidity. Non-italics: Pearson coefficients. Italics: Spearman coefficients. Bold: $p.val < 0.05$. Underlined: $p.val < 0.01$. Bold and underlined: $p.val < 0.001$. Colorized in blue: Significant positive correlation coefficients. Colorized in red: Significant negative correlation coefficients. Paralytic Shellfish Toxins (PST); Okadaic acid (OA); Pectenotoxins (PTX); Dinophysis toxins (DTX); Toxins in the mussels (m); Toxins in the oysters (o).

Table 2 : Zero inflated negative binomial model predicting *A. tamarense/pacificum* concentrations in Thau

Term	Negative Binomial regression			Logistic regression		
	Coefficient	IRR	<i>p.val</i>	Coefficient	OR(presence)	<i>p.val</i>
Intercept	8.72	6.11e3	<2e-16 ***	-0.65	1.92	0.060
Temp. <10.4	-2.28	0.010	<7.73e-5 ***	1.93	0.14	1.60e-7 ***
Temp. €[10.4; 14.6[-0.78	0.45	0.030 *	0.76	0.46	6.02e-3 **
Temp. €[21.5; 25[-1.76	0.17	3.45e-7 ***	1.15	0.31	2.03e-5 ***
Temp. > 25	-1.92	0.15	4.38e-3 **	1.90	0.15	2.66e-5 ***
Sal. < 36	-1.38	0.25	4.85e-4 ***	1.53	0.21	7.45e-7 ***
Sal. € [36; 39.2[0.22	0.81	0.49	0.81	0.44	1.47e-3 **
Turb.	0.13	1.13	0.13	-0.094	1.09	0.051

The sampling site nested are as random effects. IRR: Incident Risk Ratio. OR: Odds Ratio. Temp.: Temperature. Sal.: Salinity. Turb.: Turbidity.

References

- Abdenadher, M., Hamza, A., Fekih, W., Hannachi, I., Bellaaj, A.Z., Bradai, M.N., and Aleya, L. 2012. “Factors determining the dynamics of toxic blooms of *Alexandrium minutum* during a 10-year study along the shallow southwestern mediterranean coasts.” *Estuarine, Coastal and Shelf Science* 106:102–11. doi: 10.1016/j.ecss.2012.04.029.
- Accoroni, S., Glibert, P.M., Pichierri, S., Romagnoli, T., Marini, M., and Totti, C. 2015. “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* 45:14–25. doi: <http://dx.doi.org/10.1016/j.hal.2015.04.002>.
- Ajani, P., Larsson, M.E., Rubio, A., Bush, S., Brett, S., and Farrell, H. 2016. “Modelling bloom formation of the toxic dinoflagellates *Dinophysis acuminata* and *Dinophysis caudata* in a highly modified estuary, South Eastern Australia.” *Estuarine, Coastal and Shelf Science* 183:95–106. doi: 10.1016/j.ecss.2016.10.020.
- Akaike, H. 1974. “A new look at the statistical model identification.” *IEEE Transactions on Automatic Control* 19:716–23.
- Anderson, D.M., and Lindquist, N.L. 1985. “Time-course measurements of phosphorus depletion and cyst formation in the dinoflagellate *Gonyaulax tamarensis* Lebour.” *Journal of Experimental Marine Biology and Ecology* 86(1):1–13. doi: 10.1016/0022-0981(85)90039-5.
- Anderson, D. M., and Keafer, B.A. 1987. “An endogenous annual clock in the toxic marine dinoflagellate *Gonyaulax tamarensis*.” *Nature* 325(6105):616–17. doi: 10.1038/325616a0.
- Anderson, D.M., Bricelj, V., Cullen, J., and Rensel, J. 2001. *Monitoring and management strategies for harmful algal blooms in coastal waters*, pp. 47-51.
- Anderson, D.M., Sapiano M.R.P, Prasad, M.B.K., Long, W., Tango, P.J., Brown, C.W., and Murtugudde, R. 2010. “Predicting potentially toxigenic *Pseudo-nitzschia* blooms in the Chesapeake Bay.” *Journal of Marine Systems* 83(3):127–40. doi: 10.1016/j.jmarsys.2010.04.003.
- Arfi, R., Guiral, D., and Bouvy, M. 1993. “Wind induced resuspension in a shallow tropical lagoon.” *Estuarine, Coastal and Shelf Science* 36(6):587–604. doi: 10.1006/ecss.1993.1036.
- Baas, J.H., McCaffrey, W.D., Haughton, P., and Choux, C. 2005. “Coupling between suspended sediment distribution and turbulence structure in a laboratory turbidity current.” *Journal of Geophysical Research: Oceans* 110(C11). doi: <https://doi.org/10.1029/2004JC002668>.
- Belin, C., Soudant, D., and Amzil, Z. 2021. “Three decades of data on phytoplankton and phycotoxins on the french coast: lessons from REPHY and REPHYTOX.” *Harmful Algae* 102:101733. doi: 10.1016/j.hal.2019.101733.

- Bill, B.D., Moore, S.K., Hay, L.R., Anderson, D.M., and Trainer, V.L. 2016. "Effects of temperature and salinity on the growth of *Alexandrium* (Dinophyceae) isolates from the Salish Sea." *Journal of Phycology* 52(2):230–38. doi: 10.1111/jpy.12386.
- Bougis, P. 1974. *Écologie du plancton marin: Le phytoplancton*. Masson. pp. 38-80.
- Boynton, W.R., and Kemp, W.M. 1985. "Nutrient regeneration and oxygen consumption by sediments along an estuarine salinity gradient." *Marine Ecology Progress Series* 23:45–55. doi: 10.3354/meps023045.
- Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984. "Classification and regression trees." *CERN Document Server*. Retrieved November 27, 2020 (<https://cds.cern.ch/record/2253780>).
- Casas-Monroy, O., Linley, R.D., Adams, J.K., Chan, F.T., Drake, D.A.R., and Bailey, S.A. 2015. "Relative invasion risk for plankton across marine and freshwater systems: examining efficacy of proposed international ballast water discharge standards." *PLOS ONE* 10(3):e0118267. doi: 10.1371/journal.pone.0118267.
- Collos, Y., Vaquer, A., Laabir, M., Abadie, E., Laugier, T., Pastoureaud, A., and Souchu, P. 2007. "Contribution of several nitrogen sources to growth of *Alexandrium catenella* during blooms in Thau lagoon, southern France." *Harmful Algae* 6(6):781–89. doi: 10.1016/j.hal.2007.04.003.
- Collos, Y., Bec, B., Jauzein, C., Abadie, E., Laugier, T., Lautier, J., Pastoureaud, A., Souchu, P., and Vaquer, A. 2009. "Oligotrophication and emergence of picocyanobacteria and a toxic dinoflagellate in Thau lagoon, southern France." *Journal of Sea Research* 61(1–2):68–75. doi: <http://dx.doi.org/10.1016/j.seares.2008.05.008>.
- Cruz, R.C., Costa, P.R., Vinga, S., Krippahl, L., and Lopes, M.B. 2021. "A review of recent machine learning advances for forecasting harmful algal blooms and shellfish contamination." *Journal of Marine Science and Engineering* 9(3):283. doi: 10.3390/jmse9030283.
- Davidson, K., Anderson, D.M., Mateus, M., Reguera, B., Silke, J., Sourisseau, M., and Maguire, J. 2016. "Forecasting the risk of harmful algal blooms." *Harmful Algae* 53:1–7. doi: 10.1016/j.hal.2015.11.005.
- Davis, T.W., Bullerjahn, G.S., Tuttle, T., McKay, R.M., and Watson, S.B. 2015. "Effects of increasing nitrogen and phosphorus concentrations on phytoplankton community growth and toxicity during *Planktothrix* blooms in Sandusky Bay, Lake Erie." *Environmental Science & Technology* 49(12):7197–7207. doi: 10.1021/acs.est.5b00799.
- Deslous-Paoli, J., Souchu, P., Mazouni, N., Juge, C., and Dagault, F. 1998. "Relationship between environment and resources: impact of shellfish farming on a Mediterranean lagoon (Thau, France)." *Oceanologica Acta* 6(21):831–43.
- Drake, L.A., Meyer, A.E., Forsberg, R.L., Baier, R.E., Doblin, M.A., Heinemann, S., Johnson, W.P., Koch, M., Rublee, P.A., and Dobbs, F.C. 2005. "Potential invasion of microorganisms and pathogens via 'Interior hull fouling': biofilms inside ballast water tanks." *Biological Invasions* 7(6):969–82. doi: 10.1007/s10530-004-3001-8.

- Fiandrino, A., Ouisse, V., Dumas, F., Lagarde, F., Pete, R., Malet, N., Le Noc, S., and De Wit, R. 2017. "Spatial patterns in coastal lagoons related to the hydrodynamics of seawater intrusion." *Marine Pollution Bulletin* 119(1):132–44. doi: 10.1016/j.marpolbul.2017.03.006.
- Figueroa, R.I., Estrada, M., and Garcés, E. 2018. "Life histories of microalgal species causing harmful blooms: Haploids, diploids and the relevance of benthic stages." *Harmful Algae* 73:44–57. doi: 10.1016/j.hal.2018.01.006.
- Franks, P.J.S. 2018. "Recent advances in modelling of harmful algal blooms." Pp. 359–77 in *Global Ecology and Oceanography of Harmful Algal Blooms, Ecological Studies*, edited by P. M. Glibert, E. Berdalet, M. A. Burford, G. C. Pitcher, and M. Zhou. Cham: Springer International Publishing.
- Genovesi, B., Laabir, M., Masseret, E., Collos, Y., Vaquer, A., and Grzebyk, D. 2009. "Dormancy and germination features in resting cysts of *Alexandrium tamarense* species complex (*Dinophyceae*) can facilitate bloom formation in a shallow lagoon (Thau, southern France)." *Journal of Plankton Research* 31(10):1209–24. doi: 10.1093/plankt/fbp066.
- Gomis, C., Alcober, J., and Bernabeu, A. 1996. "Seguimiento de las poblaciones fitoplanctónicas en las bateas mrjilloneras del puerto de Valencia 1991–1994." *Matamoras E, Delgado M (Eds) IV Reunion Iberica de Fitoplancton Toxic y Biotoxinas St. Carles de La Rapita (Tarragona): Generalitat de Catalunya. Department d'Agricultura, Ramaderia i Pesca* 29–38.
- Glibert, P.M. 2020. "Harmful algae at the complex nexus of eutrophication and climate change." *Harmful Algae* 91:101583. doi: 10.1016/j.hal.2019.03.001.
- González-Gil, S., Velo-Suárez, L., Gentien, P., Ramilo, I., and Reguera, B. 2010. "Phytoplankton assemblages and characterization of a *Dinophysis acuminata* population during an upwelling–downwelling cycle." *Aquatic Microbial Ecology* 58:273–86. doi: 10.3354/ame01372.
- González Vilas L., Spyarakos, E., Torres Palenzuela J.M., and Pazos, Y. 2014. "Support Vector Machine-Based method for predicting *Pseudo-nitzschia* spp. blooms in coastal waters (Galician Rias, NW Spain)." *Progress in Oceanography* 124:66–77. doi: 10.1016/j.pocean.2014.03.003.
- Granéli, E., and Turner, J.T. eds. 2006. *Ecology of Harmful Algae: 15 Tables*. Berlin: Springer. pp. 37-52.
- Greene, W.H. 1994. Accounting for excess zeros and sample selection in Poisson and Negative Binomial Regression Models. *SSRN Scholarly Paper*. ID 1293115. Rochester, NY: Social Science Research Network.
- Grenz, C., Alliot, E., Baudinet, D., Helis, L., and Masse, H. 1992 "Influence des opérations de dévasage sur les flux de nutriments à l'interface eau-sédiment (Bassin de Thau - France)." 9.
- Hallegraeff, G.M., Anderson, D.M., and Cembella, A.D. 2003. "Harmful Algal Blooms : a global overview." *Manual on Harmful Marine Microalgae* 11.

- Hallegraeff, G.M. 2010. "Ocean climate change, phytoplankton community responses and Harmful Algal Blooms: a formidable predictive challenge." *Journal of Phycology* 46(2):220–35. doi: 10.1111/j.1529-8817.2010.00815.x.
- Hartig, F. 2020. "DHARMA: Residual Diagnostics for Hierarchical (Multi-level / Mixed) Regression Models. R Package Version 0. 3. 3. 0."
- Ishikawa, A., Hattori, M., Ishii, K.I., Kulis, D.M., Anderson, D.M., and Imai, I. 2014. "In situ dynamics of cyst and vegetative cell populations of the toxic dinoflagellate *Alexandrium catenella* in Ago Bay, central Japan." *Journal of Plankton Research* 36(5):1333–43. doi: 10.1093/plankt/fbu048.
- Itakura, S., Yamaguchi, M., YOSHIDA, M., and Fukuyo, Y. 2002. "The seasonal occurrence of *Alexandrium tamarense* (Dinophyceae) vegetative cells in Hiroshima Bay, Japan." *Fisheries Science* 68:77–86. doi: 10.1046/j.1444-2906.2002.00392.x.
- Keafer, B.A., Buesseler, K.O., and Anderson, D.M. 1992. "Burial of living dinoflagellate cysts in estuarine and nearshore sediments." *Marine Micropaleontology* 20(2):147–61. doi: 10.1016/0377-8398(92)90004-4.
- Kim, S., Kang, Y.G., Kim, HS., Yih, W., and Coats, D. 2008. "Growth and grazing responses of the mixotrophic dinoflagellate *Dinophysis acuminata* as functions of light intensity and prey concentration." *Aquatic Microbial Ecology - AQUAT MICROB ECOL* 51:301–10. doi: 10.3354/ame01203.
- Laabir, M., Jauzein, C., Genovesi, B., Masseret, E., Grzebyk, D., Cecchi, P., Vaquer, A., Perrin, Y., and Collos, Y. 2011. "Influence of temperature, salinity and irradiance on the growth and cell yield of the harmful red tide dinoflagellate *Alexandrium catenella* colonizing mediterranean waters." *Journal of Plankton Research* 33(10):1550–63. doi: 10.1093/plankt/fbr050.
- Laabir, M., Barre, N., Franco, J., Brunet, C., Masseret, E., and Collos, Y. 2012. "Morphological, biochemical and growth characteristics of *Alexandrium catenella* (Whedon & Kofoid) Balech, a toxic dinoflagellate expanding in mediterranean waters." *Cahiers De Biologie Marine* 53(3):365–72.
- Laanaia, N., Vaquer, A., Fiandrino, A., Genovesi, B., Pastoureaud, A., Cecchi, P., and Collos, Y. 2013. "Wind and temperature controls on *Alexandrium* blooms (2000-2007) in Thau lagoon (Western Mediterranean)." *Harmful Algae* 28:31–36. doi: 10.1016/j.hal.2013.05.016.
- Lambert, D. 1992. "Zero-Inflated Poisson regression, with an application to defects in manufacturing." *Technometrics* 34(1):1–14. doi: 10.2307/1269547.
- Ladagnous H., Le Bec C. 1997. Lagune de Salses – Leucate. I – Analyse bibliographique. R.INT.DEL/97.02/SETE. <https://archimer.ifremer.fr/doc/00073/18422/>
- Lane, J.Q., Roddam, C.M., Langlois, G.W., and Kudela, R.M. 2010. "Application of Solid Phase Adsorption Toxin Tracking (SPATT) for field detection of the hydrophilic phycotoxins domoic acid and saxitoxin in coastal California." *Limnology and Oceanography-Methods* 8:645–60. doi: 10.4319/lom.2010.8.645.

- Levinton, J., Ward, J., and Shumway, S. 2002. "Feeding responses of the bivalves *Crassostrea gigas* and *Mytilus trossulus* to chemical composition of fresh and aged kelp detritus." *Marine Biology* 141(2):367–76. doi: 10.1007/s00227-002-0830-9.
- Natsuike, M., Shiraishi, T., Ishii, KI., Yamamoto, K., Nakajima, M., Sawayama, S., and Imai, I. 2018. "Different nutrient availabilities of surface and bottom water under nutrient-depleted conditions during bloom formation of the toxic dinoflagellate *Alexandrium tamarense* in Osaka Bay, Japan." *北海道大学水産科学研究彙報* 68(1):7–16.
- Ninčević G., Skejic, Z.S, Buzancic, M., Marasović, I., Arapov, J., Ujević, I., Bojanić, N., Grbec, B., Kušpilić, G., and Vidjak, O. 2008. "Seasonal variability in *Dinophysis* spp. abundances and diarrhetic shellfish poisoning outbreaks along the eastern Adriatic Coast." *Botanica Marina - BOT MAR* 51:449–63. doi: 10.1515/BOT.2008.067.
- Ogata, T., Ishimaru, T., and Kodama, M. 1987. "Effect of water temperature and light intensity on growth rate and toxicity change in *Protogonyaulax tamarensis*." *Marine Biology* 95(2):217–20. doi: 10.1007/BF00409008.
- Park, M.G., Kim, S., Kim, H.S., Myung, G., Kang, Y.G., and Yih, W. 2006. "First successful culture of the marine dinoflagellate *Dinophysis acuminata*." *Aquatic Microbial Ecology* 45(2):101–6. doi: 10.3354/ame045101.
- Park, S., Lee, S.R., Park, J.H., Park, Y., Jung, M.A., Pyo, S.J., and Kim, K.S. 2011. "Prediction of red tide blooms using Decision Tree model." Pp. 710–13 in *ICTC 2011*.
- Pousse, E., Flye-Sainte-Marie, J., Alunno-Bruscia, M., Hegaret, H., and Jean, F. 2018. "Sources of paralytic shellfish toxin accumulation variability in the pacific oyster *Crassostrea gigas*." *Toxicon* 144:14–22. doi: 10.1016/j.toxicon.2017.12.050.
- Pulido, O. 2016. "Phycotoxins by Harmful Algal Blooms (HABS) and human poisoning: an overview. *Int Clin Pathol J* 2(6): 00062. DOI:10.15406/Icpjl.2016.02.00062." *International Clinical Pathology Journal* 2. doi: 10.15406/icpjl.2016.02.00062.
- R Core Team. 2020. "R: A language and environment for statistical computing."
- Ralston, D.K., Keafer, B.A., Brosnahan, M.L., and Anderson, D.M. 2014. "Temperature dependence of an estuarine Harmful Algal Bloom: resolving interannual variability in bloom dynamics using a degree-day approach." *Limnology and Oceanography* 59(4):1112–26. doi: 10.4319/lo.2014.59.4.1112.
- Ralston, D.K., Brosnahan, M.L., Fox, S.E., Lee, K.D., and Anderson, D.M. 2015. "Temperature and residence time controls on an estuarine Harmful Algal Bloom: modeling hydrodynamics and *Alexandrium fundyense* in Nauset Estuary." *Estuaries and Coasts* 38(6):2240–58. doi: 10.1007/s12237-015-9949-z.
- Raso, S., Van Genugten, B., Vermuë, M., and Wijffels, R.H. 2012. "Effect of oxygen concentration on the growth of *Nannochloropsis* Sp. at low light intensity." *Journal of Applied Phycology* 24(4):863–71. doi: 10.1007/s10811-011-9706-z.
- Reguera, B., Velo-Suarez, L., Raine, R., and Park, M.G. 2012. "Harmful *Dinophysis* species: a review." *Harmful Algae* 14:87–106.

- REPHY – French Observation and Monitoring program for Phytoplankton and Hydrology in coastal waters (2021). REPHY dataset - French Observation and Monitoring program for Phytoplankton and Hydrology in coastal waters. Metropolitan data. SEANOE. <https://doi.org/10.17882/47248>
- REPHYTOX - French Monitoring program for Phycotoxins in marine organisms (2021). REPHYTOX dataset. French Monitoring program for Phycotoxins in marine organisms. Data since 1987. SEANOE. <https://doi.org/10.17882/47251>
- Séchet, V., Sibat, M., Billien, G., Carpentier, L., Rovillon, G-A., Raimbault, V., Malo, F., Gaillard, S., Perrière-Rumebe, M., Hess, P., and Chomérat, N. 2021. “Characterization of toxin-producing strains of *Dinophysis* Spp. (Dinophyceae) isolated from french coastal waters, with a particular focus on the *D. Acuminata*-Complex.” *Harmful Algae* 107:101974. doi: 10.1016/j.hal.2021.101974.
- Seubert, E.L., Gellene, A.G., Howard, M.D.A., Connell, P., Ragan, M., Jones, B.H., Runyan, J., and Caron, D.A. 2013. “Seasonal and annual dynamics of harmful algae and algal toxins revealed through weekly monitoring at two coastal ocean sites off southern California, USA.” *Environmental Science and Pollution Research* 20(10):6878–95. doi: 10.1007/s11356-012-1420-0.
- Shin, J., Yoon, S.H., and Cha, Y. 2017. “Prediction of cyanobacteria blooms in the lower Han River (South Korea) using ensemble learning algorithms.” *DESALINATION AND WATER TREATMENT* 84:31–39. doi: 10.5004/dwt.2017.20986.
- Swan, S.C., Turner, A.D., Bresnan, E., Whyte, C., Paterson, R.F., McNeill, S., Mitchell, E., and Davidson, K. 2018. “*Dinophysis acuta* in scottish coastal waters and its influence on diarrhetic shellfish toxin profiles.” *Toxins* 10(10). doi: 10.3390/toxins10100399.
- Therneau, T., and Atkinson, B. 2019. “Rpart: Recursive Partitioning and Regression Trees. R Package Version 4.1-15.”
- Utermöhl, H. 1931. “Neue wege in der quantitativen erfassung des planktons: (mit besondere berücksichtigung des ultraplanktons).” *Verh. Int. Ver. Theor. Angew. Limnol* 5:567–95.
- Valbi, E., Ricci, F., Capellacci, S., Casabianca, S., Scardi, M., and Penna, A. 2019. “A model predicting the PSP toxic dinoflagellate *Alexandrium minutum* occurrence in the coastal waters of the NW Adriatic Sea.” *Scientific Reports* 9(1):4166. doi: 10.1038/s41598-019-40664-w.
- Verity, P.G., Villareal, T.A., and Smayda, T.J. 1988. “Ecological investigations of blooms of colonial *Phaeocystis pouchetti* . II. The role of life-cycle phenomena in bloom termination.” *Journal of Plankton Research* 10(4):749–66. doi: 10.1093/plankt/10.4.749.
- Xu, N., Duan, S., Li, A., Zhang, C., Cai, Z., and Hu, Z. 2010. “Effects of temperature, salinity and irradiance on the growth of the harmful dinoflagellate *Prorocentrum donghaiense* Lu.” *Harmful Algae* 9(1):13–17.