



HAL
open science

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

Aurélien Bouquet, Mohamed Laabir, Jean Luc Rolland, Nicolas Chomérat, Christelle Reynes, Robert Sabatier, Christine Felix, Tom Berteau, Claude Chiantella, Eric Abadie

► To cite this version:

Aurélien Bouquet, Mohamed Laabir, Jean Luc Rolland, Nicolas Chomérat, Christelle Reynes, et al.. 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. *Harmful Algae*, 2022, 115, pp.102234. 10.1016/j.hal.2022.102234 . hal-03685552

HAL Id: hal-03685552

<https://hal.science/hal-03685552>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

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

4 Aurélien Bouquet^{a*}, Mohamed Laabir^b, Jean Luc Rolland^a, Nicolas Chomérat^c, Christelle
5 Reynes^{d,e}, Robert Sabatier^{d,e}, Christine Felix^f, Tom Berteau^g, Claude Chiantella^g, Eric
6 Abadie^{a,h}

7 ^a MARBEC, Université de Montpellier, CNRS, Ifremer, IRD, 87 Avenue Jean Monnet, 34200
8 Sète, France

9 ^b Université de Montpellier, MARBEC, CNRS, Ifremer, IRD, 34095 Montpellier, France

10 ^c IFREMER, Station de Biologie Marine, Place de la Croix, BP 40537, 29185 Concarneau
11 Cedex, France

12 ^d Institut de Génomique Fonctionnelle, IGF, Univ. Montpellier, CNRS, INSERM, 34094
13 Montpellier, France

14 ^e Faculté de Pharmacie, Univ. Montpellier 34093 Montpellier, France

15 ^f Université de Montpellier, MARBEC, CNRS, Ifremer, IRD, 87 Avenue Jean Monnet, 34200
16 Sète, France

17 ^g Ifremer, Laboratoire environnement ressources du Languedoc Roussillon, 34200 Sète,
18 France

* Corresponding author at IFREMER, Laboratoire Environnement Ressources Languedoc-Roussillon, 87 Av. Jean Monnet, 34200 Sète, France.

E-mail addresses : aurelien.bouquet@ifremer.fr (A. Bouquet), eric.abadie@ifremer.fr (E. Abadie),
nicolas.chomerat@ifremer.fr (N. Chomérat), christelle.reynes@umontpellier.fr (C. Reynes),
robert.sabatier@umontpellier.fr (R. Sabatier), mohamed.laabir@umontpellier.fr (M. Laabir),
christine.felix@umontpellier.fr (C. Felix), jean.luc.rolland@ifremer.fr (JL. Rolland), eric.abadie@ifremer.fr (T. Berteau), eric.abadie@ifremer.fr (C. Chiantella)

19 ^h IFREMER, Biodivenv, 79 route de pointe fort, 97231 Le Robert, France

20 **Abstract**

21 French Mediterranean lagoons are frequently subject to shellfish contamination by Diarrheic
22 Shellfish Toxins (DSTs) and Paralytic Shellfish Toxins (PSTs). To predict the effect of
23 various environmental factors (temperature, salinity and turbidity) on the abundance of the
24 major toxins producing genera, *Dinophysis* and *Alexandrium*, and the link with shellfish
25 contamination, we analysed a 10-year dataset collected from 2010 to 2019 in two major
26 shellfish farming lagoons, Thau and Leucate, using two methods: decision trees and Zero
27 Inflated Negative Binomial (ZINB) linear regression models. Analysis of these decision trees
28 revealed that the highest risk of *Dinophysis* bloom events occurred at temperature $<16.3^{\circ}\text{C}$
29 and salinity <27.8 , and of *Alexandrium* at temperature ranging from 10.4 to 21.5°C and
30 salinity >39.2 . The highest risk of shellfish contaminations by DSTs and PSTs occurred
31 during the set of conditions associated with high risk of bloom events. Linear regression
32 prediction enables us to understand whether temperature and salinity influence the presence of
33 *Alexandrium* and affect its abundance. However, *Dinophysis* linear regression could not be
34 validated due to overdispersion issues. This work demonstrates the tools which could help
35 sanitary management of shellfish rearing areas.

36 **Keywords**

37 *Alexandrium*, *Dinophysis*, Toxic blooms, Temperature, Salinity, Prediction

38 **I. Introduction**

Abbreviations : DT: Decision tree, IRR : Incidence Risk Ratio, LD : Limit of Detection, LQ: Limit of
Quantification, OR: Odds Ratio, REPHY: French Observation and Monitoring program for Phytoplankton and
Hydrology in coastal waters, REPHYTOX: French Monitoring Program for Phytotoxins in marine organisms,
ZINB: Zero Inflated Negative Binomial

39 The geographical expansion of harmful algae and their durable settlement in many marine
40 ecosystems has become a major concern all over the world (Hallegraeff et al. 2003; Drake et
41 al. 2005; Hallegraeff 2010; Casas-Monroy et al. 2015; Glibert 2020). Harmful algae (HA)
42 proliferation or HABs (Harmful Algal Blooms) leads to observable negative effects on marine
43 resources, economics and human health (Granéli and Turner, 2006). Many HA produce toxins
44 that can contaminate marine resources and lead to human intoxication (Pulido 2016). For this
45 reason, the proliferation of HAs in water is frequently monitored near shellfish farming areas
46 (Anderson et al. 2001). Studies showed that blooms can be impacted by many environmental
47 factors such as temperature and salinity (Anderson and Keafer, 1987; Laabir et al. 2011), light
48 (Figuroa et al. 2018), oxygen concentration in sediment and nutrient availability (Collos et
49 al. 2007, 2009; Xu et al. 2010; Ishikawa et al. 2014). Blooms are generally triggered by the
50 synergic combination of several factors (Accoroni et al. 2015; Raso et al. 2012) among which
51 temperature and salinity play a key role as they can influence every phases of bloom
52 dynamics, including initiation, growth and termination (Figuroa et al. 2018; Granéli and
53 Turner 2006). Many dinoflagellates including *Alexandrium* show planktonic and benthic
54 forms, and initiation of the blooms depends on the success of cysts germination which occurs
55 after a dormancy period. The duration of this phase was shown to be influenced by
56 temperature, and some HA cysts germinate only within specific ranges of temperature
57 (Anderson and Keafer 1987; Genovesi et al. 2009). Moreover, germination can be linked with
58 water agitation leading to sediment resuspension (Laanaia et al. 2013) which is generally
59 associated with variations in water temperature and salinity. Growth corresponds to biomass
60 increase which duration and maximum concentration can be influenced by temperature and
61 salinity along with other factors such as nutrients (Bill et al. 2016; Collos et al. 2007; Davis
62 et al. 2015; Ishikawa et al. 2014; Laabir et al. 2011; Ralston et al. 2015; Raso et al. 2012; Xu
63 et al. 2010). Bloom termination happens when environmental conditions are no longer

64 favourable, which can be attributed to nutrient depletion or inadequate temperature leading to
65 mortality or encystment (Ralston et al. 2014; Verity et al. 1988).

66 As a consequence of these various and contrasted impacts of environmental factors on HAB
67 dynamic, efforts have been made to develop early HABs warning systems based on
68 environmental predictors. Conventionally, statistical numerical modelling has been used for
69 HABs prediction (González Vilas et al. 2014). Through linear regression, these models can
70 efficiently predict phytoplankton dynamics because they are capable of integrating many
71 predictors and can be adapted to different data distributions (Davidson et al. 2016). Given the
72 cyclic and seasonal pattern in HABs, they have mainly been investigated through time series
73 modelling based on a range of variables measured over time (Cruz et al., 2021). For instance,
74 several studies have applied linear regression to forecasting the blooms of *Pseudo-nitzschia*
75 (Anderson et al. 2010; Lane et al. 2010; Seubert et al. 2013). The interest in alternatives like
76 Machine Learning (ML) models has recently increased. ML methods consist of using
77 computer systems that are able to learn and draw inference from data patterns without explicit
78 instructions. One of the many advantages of ML methods is their ability to predict highly
79 dynamic and complex phenomena and to handle big and non-linear data. A wide variety of
80 machine learning approaches has been used over the past years to forecast HABs (Franks
81 2018; Cruz et al. 2021). Among them, decision trees (DTs) seem to be highly relevant tools.
82 A DT is a tree-like structure in which the leaves represent outcome labels and the branches
83 represent conjunctions of the input features that resulted in those outcomes. Aside from
84 overcoming the problems due to missing data, they are easy to interpret since they provide
85 concrete thresholds that can be used by marine stakeholders. In South Korea, Shin (2017)
86 used DTs to forecast cyanobacteria blooms and (Park et al. 2011) applied this method to
87 predict red tide blooms.

88 French Mediterranean lagoons are special ecosystems as they are semi-enclosed marine
89 systems, containing many shellfish farming areas, and they are frequently subject to HAB
90 events and particularly proliferation of the dinoflagellates *Alexandrium pacificum* and
91 *Dinophysis acuminata* complex responsible for intoxication events threatening human health
92 and resulting in economic losses (Deslous-Paoli et al. 1998; Collos et al. 2009; Laabir et al.
93 2011; Laanaia et al. 2013; Belin et al. 2021; Séchet et al. 2021). Until now, no machine
94 learning approach has been developed to quantify and predict the risks of HABs proliferations
95 and shellfish contamination in these ecosystems. Based on a ten-year dataset of parameters
96 (temperature, salinity and turbidity) and of HA taxa, using both decision trees and a Zero-
97 Inflated Negative Binomial regression model, this study aimed at providing robust tools to
98 predict the risk of bloom of the two major HA genera in French Mediterranean lagoons,
99 *Dinophysis* and *Alexandrium*, and the subsequent shellfish contamination in exploited semi-
100 enclosed marine systems.

101 **2. Materials and methods**

102 *2.1. Data recovery*

103 Data were collected from the framework of the REPHY and REPHYTOX monitoring net-
104 works (REPHY - French Observation and Monitoring program for Phytoplankton and Hy-
105 drology in coastal waters 2019; REPHYTOX - French Monitoring program for Phytotoxins in
106 marine organisms 2019).

107 Briefly, the data comes from field surveys carried out by Ifremer in two exploited French
108 Mediterranean lagoons, Thau (7000ha, mean depth 4m, water temperature range : 3.9-29.6°C;
109 salinity range : 27.4-42.9) and Leucate (5400ha, mean depth 1.9m water temperature range
110 0.9-29.3°C, salinity range: 8.4-42.3) (Fiandrino et al. 2017; Ladagnous and Le Bec 1997)
111 from January 2010 to December 2020 (Fig 1). The concentration of phytoplankton species

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

127 2.3. Data processing and analyses

128 2.3.1. Preliminary processing

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

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

140 *2.3.1. Exploratory analysis*

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

145 *2.3.2. Building decision trees*

146 *Variable selection and encoding*

147 Decision trees (DT) were used to predict the risks of HABs and of bivalve contamination
148 considering temperature, salinity and turbidity as predictors. Prior to the analysis, *Dinophysis*
149 and *A. tamarensis/pacificum* concentrations were converted to binary variables: the abundance
150 of cells was encoded by 1 when the concentration exceeded 1000 cells/L for *A.*
151 *tamarensis/pacificum* (value inducing the sanitary warning protocol) and 500 cells/L for
152 *Dinophysis* (estimated value indicating a bloom, the sanitary warning protocol being induced
153 by a concentration of cells \geq LD), and by 0 when the abundances were beyond these
154 concentrations. Toxin concentrations were converted to categorical variables based on the
155 regulatory thresholds: they were encoded by 0 when concentrations were below 80 μg per kg
156 of shellfish flesh) for lipophilic toxins and below 340 $\mu\text{g}/\text{kg}$ for PSTs, by 1 when they were
157 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
158 $\mu\text{g}/\text{kg}$ for PSTs, and by 2 when they were above 160 $\mu\text{g}/\text{kg}$ for lipophilic toxins and above
159 800 $\mu\text{g}/\text{kg}$ for PSTs.

160 *Decision tree analysis*

161 The DTs building process is described in Fig 2. DTs were set up with the *rpart package*
162 (Therneau and Atkinson 2019) using the CART algorithm (Breiman et al. 1984). As a
163 supplementary constraint, a minimum of 10 observations was imposed for a branch to become
164 a branch node and there had to be at least 5 observations per leaf node. A maximum of 10
165 nodes per tree was imposed. In order to set up a robust tool: for each predicted variable, the
166 experiment with DT was repeated 100 times; each time, the data set was split into two sets, 80
167 % made up the training set and 20% the testing set. The split algorithm allowed all the
168 different pairs of training and testing sets to have approximately the same proportion of 0, 1
169 (and 2 for the toxins) in the predicted variables. The structures of the 100 trees were then
170 compared in order to find the most common one. For each node the process consisted of:

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

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

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

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

185 - The proportion of 1 or 2 in at least one of the nodes of the testing trees should be above 20%

186 - The number of observations in each node of the testing trees should be above 3

187 *2.3.3. Linear models*

188 *Type of regression*

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

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

211 *Variable selection and encoding*

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

227 *Validation*

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

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

238 **3. Results**

239 *3.1. Exploratory data analysis*

240 Correlation coefficients between phytoplankton abundance, abiotic factors and toxins in Thau
241 and Leucate lagoons were calculated as shown in Tab 1A and 1B respectively. In both
242 lagoons, the coefficients showed significant correlations between phytoplankton and abiotic
243 factors: In Thau, the Spearman coefficient showed a positive correlation ($p.val < 0.001$)
244 between *A. tamarense/pacificum* and salinity (Sal.). In Leucate, *Dinophysis* was negatively
245 correlated to temperature (Temp.) and Sal. ($p.val < 0.001$), whereas it was positively correlated
246 to Temp. in Thau ($p.val < 0.001$) (Tab. 1). *Dinophysis* was also positively correlated to
247 turbidity (Turb.) (Spearman coefficient 0.15, $p.val < 0.001$) (Tab. 1). When significant, the
248 correlations between *A. tamarense/pacificum*, *Dinophysis* and abiotic factors were quite low,
249 absolute values ranged from 0.15 to 0.31 (Tab. 1). In both lagoons, whenever significant, the
250 Spearman coefficients between the environment and toxic phytoplankton species were always
251 stronger than the Pearson coefficients, suggesting non-linear relationships. Furthermore,
252 toxins were highly correlated to the abundance of the microalgae which produce them. The
253 significant correlations between *A. tamarense/pacificum* and PSTs in Thau ranged from 0.41
254 to 0.58 and the Pearson coefficients were stronger than the Spearman coefficients, with
255 $p.val < 0.001$, suggesting a linear relationship (Tab. 1). The lipophilic toxins (OA, PTX, DTX)

256 were positively correlated with the abundance of *Dinophysis*, the significant coefficients
257 ranged from 0.45 to 0.69 ($p.val < 0.00$) (Tab 1). Furthermore, abiotic factors seemed to have
258 similar impact on the toxins as on the microalgae producing them. In Thau, PSTs were
259 positively correlated to Sal. (significant correlation coefficients ranged from 0.3 to 0.41,
260 $p.val < 0.05$) (Tab. 1). In Leucate, lipophilic toxins were negatively correlated to both Temp.
261 (coefficients ranging from -0.38 to -0.57) and Sal. (significant coefficients ranging from -0.14
262 to -0.35) and positively correlated to Turb. (significant coefficient ranging from 0.27 to 0.28,
263 $p.val < 0.001$) (Tab. 1). The Spearman coefficients showed that *A. tamarensis/pacificum* and
264 *Dinophysis* abundances were non-linearly but positively correlated with the abundance of
265 dinoflagellates and diatoms, and that these correlations were stronger for *A.*
266 *tamarensis/pacificum* (coefficients ranging from 0.18 to 0.29, $p.val < 0.001$) than for
267 *Dinophysis* (coefficients ranging from 0.08 to 0.09, $p.val < 0.05$) (Tab. 1).

268 3.2. Decision trees

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

281 Decision trees were able to predict when shellfish contaminations could occur (Fig 5). The
282 trees showed that toxins were usually found in animals in the conditions promoting a bloom
283 event (Fig 5). Again, Turb. was never kept in the pruned DTS. In Thau, the risk of
284 contamination of mussels and oysters by PSTs was significant for seawater Sal. > 37.7 and
285 39.9 , respectively (Fig 5 A/B). Low Temp. also promoted risks of shellfish contamination.
286 For mussels, Temp. $< 18.5^{\circ}\text{C}$ combined with Sal. > 37.7 yielded a total mean risk of
287 contamination of 92% (44.9% between $340\ \mu\text{g}/\text{kg}$ and $800\ \mu\text{g}/\text{kg}$ (labelled 1 on the figure)
288 and 47.1% exceeding $800\ \mu\text{g}/\text{kg}$ (labelled 2 on the testing data sets). For oysters, Temp.
289 $< 16^{\circ}\text{C}$ combined with Sal. > 39.9 led to a 57% mean risks of contamination, the latter being
290 mainly comprised between $340\ \mu\text{g}/\text{kg}$ and $800\ \mu\text{g}/\text{kg}$ (52.3% of the observations in the testing
291 dataset). It is possible that the contaminations were rare below a low Temp. threshold, but
292 were not represented on the decision trees because of a low number of observations. In
293 Leucate, the risks of mussel and oyster lipophilic contamination was higher when Temp. is
294 $< 16.3^{\circ}\text{C}$ and $< 13.5^{\circ}\text{C}$, respectively (Fig 5 C/D). The 16.3°C threshold showed that mussels
295 would become contaminated as soon as Temp. allowed *Dinophysis* blooms. Under this Temp.,
296 the risk of contamination was high independently of the Sal. (56.7% and 31.1% risk of a
297 contamination over $160\ \mu\text{g}/\text{kg}$ (labelled 2 on the figure) for a Sal. respectively $>$ and < 33.5). In
298 oysters, contamination by lipophilic toxins happened at Temp. $< 13.5^{\circ}\text{C}$ and Sal. < 27.8 that
299 led to 30.7% mean risk of contamination (over $160\ \mu\text{g}/\text{kg}$) in the testing data sets. In all, these
300 trees showed that the conditions in which oysters were contaminated were more restrictive
301 than for mussels: the Temp. and Sal. ranges were systematically narrower for oysters.

302 3.3. Linear regression

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

306 thanks to the results of decision trees and from visual examination of data distribution. The
307 negative binomial part of the model, which depicted how predictors impact cell abundance,
308 showed that the reference Temp. level, 14.6 to 21.5 °C, was associated with the highest
309 average microalgal abundance. The IRR revealed that the average abundance was multiplied
310 by 0.15 when Temp. was >25°C ($p.val$ 4.38e-3), and by 0.010 when Temp. was <10.4°C
311 ($p.val$ <7.73e-5). A Sal. >36 was associated with the highest average abundance of *A.*
312 *tamarensis/pacificum*: the average abundance was multiplied by 0.25 when Sal. was <36
313 compared with a Sal. >39.2, but no difference was shown when Sal. was comprised between
314 36 and 39.2. Turb. did not significantly impact *A. tamarensis/pacificum* abundance. It should
315 be noted, before interpreting the logit part of the model, the coefficients were conversely
316 related to the presence of the alga. In order to depict how the predictors affect the probability
317 of presence of the alga, we choose to calculate OR (presence) = $1/(exp(coefficients))$. This part
318 of the model showed that the reference Temp. level was associated with the highest
319 probability of presence of *A. tamarensis/pacificum*. For instance, the odds of its presence were
320 multiplied by 0.14 when the temperature was <10.4°C and by 0.15 when it was >25. The
321 reference Sal. (over 39.2) led to the highest likelihood of the presence of *A.*
322 *tamarensis/pacificum*: the odds of the presence were multiplied by 0.21 when Sal. was <36
323 and by 0.44 when it was between 36 and 39.2. Turb. was not associated with the presence of
324 this alga. Overall, what emerged from the two parts of the models was that the conditions in
325 which *A. tamarensis/pacificum* usually reached higher abundances were generally the same as
326 those in which the probability of finding it were the highest.

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

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

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

340 **4. Discussion and conclusion**

341 *4.1. Discussion*

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

348 First, the exploratory bivariate analyses indicated an impact of temperature and salinity on
349 HABs. These results were in accordance with several studies showing that temperature and
350 salinity influence the occurrence of *Alexandrium* and *Dinophysis* blooms (Abdenadher et al.
351 2012; Anderson and Keafer 1987; Gomis et al. 1996; Itakura et al. 2002; Ninčević Gladan et
352 al. 2008; Swan et al. 2018). The correlation coefficients were medium to low, but were not
353 expected to be much higher as they depicted the influence of only one variable on the HA

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

356 By using decision trees, we determined the range of optimum temperature and salinity
357 associated with high risks of blooms. The proportion of blooms with a salinity above 39.2 and
358 temperature ranging between 10.4 and 21.5 for *A. tamarense/pacificum*, and when salinity
359 was below 27.8 for *Dinophysis* spp. exceeded 50%, which could trigger safeguard measures
360 of the reared shellfish. If the optimal ranges of temperature and salinity may appear wide, they
361 only represent 9.7% and 3.1% of the data in Thau and Leucate, respectively, which has to do
362 with the broad ranges of temperature and salinity in Mediterranean lagoons. The temperatures
363 were congruent with previous observations showing that *A. tamarense/pacificum* blooms
364 mainly occur from October to December and April to June but were rarely observed in
365 summer and winter (Laabir et al. 2011, 2012). Laanaia (2013) highlighted that *Alexandrium*
366 blooms occurred mainly during periods when salinity was below 39.3 from 2002 to 2006.
367 Even though substantial risks could subsist outside the optimum ranges depicted by the DTs
368 (11.9% risk of *A. tamarense/pacificum* bloom below 39.2), this study could thus reveal a
369 change in *A. tamarense* populations composition together with a variation in intra-species
370 physiological constants in Thau over the past few years. Our work also showed positive
371 correlations between toxins in bivalves and the HAs producing them, as well as with the
372 environmental factors that impact these algae. The optimal temperature and salinity ranges
373 associated with high risks of shellfish contamination by phycotoxins were congruent with the
374 ranges increasing the risks of blooms. In Mediterranean lagoons, all referenced *Dinophysis*
375 species were potentially toxic. Recently, qPCR analyses the *Alexandrium* species in Thau
376 lagoon showed that the proportion of non-toxic *Alexandrium* species during a bloom were
377 very low (Sectox project, Eric Abadie, unpublished data). This could explain the high
378 correlations between algae and toxins. The correlation coefficients and the DTs showed that

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

386 Through linear regression, the model confirmed the DT results obtained with *A.*
387 *tamarensis/pacificum* and showed how abiotic factors influence either its presence or its cell
388 abundance when present. The optimal set of parameters chosen by the CART algorithm was
389 significantly associated with the maximum average abundance and odds of presence of the
390 algae. Moreover, this model specified the sets of parameters more accurately by
391 distinguishing significantly more levels of temperature and salinity than the DT. It also
392 revealed that the optimal sets of parameters are similar for the presence of the microalgae and
393 for their abundance when they were present: this could indicate that germination and
394 proliferation occurred in the same conditions (Temperature from 14.6 to 21.5°C and Salinity
395 >39.2). Actually, from 2010 to 2020, most of the substantial blooms occurred within these
396 conditions, as the bloom of October 2015 (820000 cells/mL) and October 2017 (35000
397 cells/mL). Few other blooms occurred for slightly lower salinities, as in June 2011 (250000
398 cells/mL; Salinity 36.6). This is however not in total contradiction with the model since the
399 latter shows that, even though a salinity over 39.2 increases the risk of *Alexandrium* presence,
400 its abundance when it's present could be increased by a salinity over 36. This brings to light
401 that, if this model shows conditions where risks are very important, it does not mean that
402 blooms cannot occur outside them. The model predicting *Dinophysis* spp. abundance could
403 not be validated due to overdispersion issues. In contrast to *Alexandrium*, *Dinophysis* feeds on

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

412 Several other environmental factors could have been integrated to increase the model's
413 performances regarding the mechanisms of blooms. Indeed, temperature and salinity changes
414 may occur along with variations in other factors such as nutrients, water column conditions
415 and residence time which can trigger or deplete blooms. In addition to the direct impact of
416 temperature toward *Alexandrium* cyst germination (Anderson and Keafer 1987; Genovesi et
417 al. 2009), this factor is likely to be related with water stability or turbulence. In this regard,
418 winds seem to be a critical factor in lagoons by promoting turbulence and resuspension and
419 decreasing temperature (Laanaia et al. 2013). As instance, an important bloom of
420 *Alexandrium* occurred in 2004 for a salinity of 36.5 (outside the optimal salinities depicted by
421 the DTs): this bloom was triggered by cysts germination resulting from water column
422 agitation by strong winds, also inducing a drop in temperature (Laanaia et al. 2013).
423 Moreover, salinity and temperature may be associated with nutrient availability as they can be
424 linked with nitrogenous and phosphorous fluxes between compartments (organic and
425 inorganic nutrients in water column, organisms and sediments) (Bougis 1974; Boynton and
426 Kemp 1985; Grenz et al. 1992). Many studies showed that nutrient availability is positively
427 associated with *Dinophysis* (Ajani et al., 2016) and *Alexandrium* growth (Collos et al., 2007;
428 2009; Natsuike et al., 2018) and that a lack of nutrients could lead to sexuality, and thus to the

429 decline of blooms (Anderson and Lindquist 1985). Unfortunately, nutrient measurements by
430 the REPHY network were not performed at the same time and stations than those of this
431 work. Further studies including nutrient availability in water would allow to understand if
432 temperature and salinity have an indirect impact on cells abundance as their variation occur
433 along nutrients fluxes. Oxygen and light intensity are also known to influence *Alexandrium*
434 proliferations (Anderson and Keafer 1987; Keafer et al. 1992; Ajani et al. 2016; Valbi et al.
435 2019) and light increases toxin production by HA (Ogata et al. 1987). As nutrients, these
436 factors were not measured during the period or sites considered here but could be integrated in
437 further studies. We tried to integrate turbidity, although the linear regression showed no
438 significant impact and it never ended up as a remaining variable in the pruned DTs. In the
439 Mediterranean lagoons as in all aquatic ecosystems, turbidity can result from a large panel of
440 environmental variations such as water turbulence (Baas et al. 2005) and surface winds (Arfi
441 et al. 1993). These different factors could have various impacts on *Alexandrium* and
442 *Dinophysis* distribution, which may explain why turbidity appears to be a bad predictor of the
443 concentration of these species.

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

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

457 4.2. Conclusion

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

467 **Acknowledgements**

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

473 **References**

474 Abdenadher, M., Hamza, A., Fekih, W., Hannachi, I., Bellaaj, A.Z., Bradai, M.N., and Aleya,
475 L. 2012. "Factors determining the dynamics of toxic blooms of *Alexandrium minutum*
476 during a 10-year study along the shallow southwestern mediterranean coasts."
477 *Estuarine, Coastal and Shelf Science* 106:102–11. doi: 10.1016/j.ecss.2012.04.029.

- 478 Accoroni, S., Glibert, P.M., Pichierri, S., Romagnoli, T., Marini, M., and Totti, C. 2015. “A
479 conceptual model of annual *Ostreopsis cf. ovata* blooms in the northern Adriatic Sea
480 based on the synergic effects of hydrodynamics, temperature, and the N:P ratio of
481 water column Nutrients.” *Harmful Algae* 45:14–25. doi:
482 <http://dx.doi.org/10.1016/j.hal.2015.04.002>.
- 483 Ajani, P., Larsson, M.E., Rubio, A., Bush, S., Brett, S., and Farrell, H. 2016. “Modelling
484 bloom formation of the toxic dinoflagellates *Dinophysis acuminata* and *Dinophysis*
485 *caudata* in a highly modified estuary, South Eastern Australia.” *Estuarine, Coastal*
486 *and Shelf Science* 183:95–106. doi: 10.1016/j.ecss.2016.10.020.
- 487 Akaike, H. 1974. “A new look at the statistical model identification.” *IEEE Transactions on*
488 *Automatic Control* 19:716–23.
- 489 Anderson, D.M., and Lindquist, N.L. 1985. “Time-course measurements of phosphorus
490 depletion and cyst formation in the dinoflagellate *Gonyaulax tamarensis* Lebour.”
491 *Journal of Experimental Marine Biology and Ecology* 86(1):1–13. doi: 10.1016/0022-
492 0981(85)90039-5.
- 493 Anderson, D. M., and Keafer, B.A. 1987. “An endogenous annual clock in the toxic marine
494 dinoflagellate *Gonyaulax tamarensis*.” *Nature* 325(6105):616–17. doi:
495 10.1038/325616a0.
- 496 Anderson, D.M., Bricelj, V., Cullen, J., and Rensel, J. 2001. *Monitoring and management*
497 *strategies for harmful algal blooms in coastal waters*, pp. 47-51.
- 498 Anderson, D.M., Sapiano M.R.P, Prasad, M.B.K., Long, W., Tango, P.J., Brown, C.W., and
499 Murtugudde, R. 2010. “Predicting potentially toxigenic *Pseudo-nitzschia* blooms in
500 the Chesapeake Bay.” *Journal of Marine Systems* 83(3):127–40. doi:
501 10.1016/j.jmarsys.2010.04.003.
- 502 Arfi, R., Guiral, D., and Bouvy, M. 1993. “Wind induced resuspension in a shallow tropical
503 lagoon.” *Estuarine, Coastal and Shelf Science* 36(6):587–604. doi:
504 10.1006/ecss.1993.1036.
- 505 Baas, J.H., McCaffrey, W.D., Haughton, P., and Choux, C. 2005. “Coupling between
506 suspended sediment distribution and turbulence structure in a laboratory turbidity
507 current.” *Journal of Geophysical Research: Oceans* 110(C11). doi:
508 <https://doi.org/10.1029/2004JC002668>.
- 509 Belin, C., Soudant, D., and Amzil, Z. 2021. “Three decades of data on phytoplankton and
510 phycotoxins on the french coast: lessons from REPHY and REPHYTOX.” *Harmful*
511 *Algae* 102:101733. doi: 10.1016/j.hal.2019.101733.
- 512 Bill, B.D., Moore, S.K., Hay, L.R., Anderson, D.M., and Trainer, V.L. 2016. “Effects of
513 temperature and salinity on the growth of *Alexandrium* (Dinophyceae) isolates from
514 the Salish Sea.” *Journal of Phycology* 52(2):230–38. doi: 10.1111/jpy.12386.
- 515 Bougis, P. 1974. *Écologie du plancton marin: Le phytoplancton*. Masson. pp. 38-80.

- 516 Boynton, W.R., and Kemp, W.M. 1985. "Nutrient regeneration and oxygen consumption by
517 sediments along an estuarine salinity gradient." *Marine Ecology Progress Series*
518 23:45–55. doi: 10.3354/meps023045.
- 519 Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984. "Classification and
520 regression trees." *CERN Document Server*. Retrieved November 27, 2020
521 (<https://cds.cern.ch/record/2253780>).
- 522 Casas-Monroy, O., Linley, R.D., Adams, J.K., Chan, F.T., Drake, D.A.R., and Bailey, S.A.
523 2015. "Relative invasion risk for plankton across marine and freshwater systems:
524 examining efficacy of proposed international ballast water discharge standards." *PLOS*
525 *ONE* 10(3):e0118267. doi: 10.1371/journal.pone.0118267.
- 526 Collos, Y., Vaquer, A., Laabir, M., Abadie, E., Laugier, T., Pastoureaud, A., and Souchu, P.
527 2007. "Contribution of several nitrogen sources to growth of *Alexandrium catenella*
528 during blooms in Thau lagoon, southern France." *Harmful Algae* 6(6):781–89. doi:
529 10.1016/j.hal.2007.04.003.
- 530 Collos, Y., Bec, B., Jauzein, C., Abadie, E., Laugier, T., Lautier, J., Pastoureaud, A., Souchu,
531 P., and Vaquer, A. 2009. "Oligotrophication and emergence of picocyanobacteria and
532 a toxic dinoflagellate in Thau lagoon, southern France." *Journal of Sea Research*
533 61(1–2):68–75. doi: <http://dx.doi.org/10.1016/j.seares.2008.05.008>.
- 534 Cruz, R.C., Costa, P.R., Vinga, S., Krippahl, L., and Lopes, M.B. 2021. "A review of recent
535 machine learning advances for forecasting harmful algal blooms and shellfish
536 contamination." *Journal of Marine Science and Engineering* 9(3):283. doi:
537 10.3390/jmse9030283.
- 538 Davidson, K., Anderson, D.M., Mateus, M., Reguera, B., Silke, J., Sourisseau, M., and
539 Maguire, J. 2016. "Forecasting the risk of harmful algal blooms." *Harmful Algae*
540 53:1–7. doi: 10.1016/j.hal.2015.11.005.
- 541 Davis, T.W., Bullerjahn, G.S., Tuttle, T., McKay, R.M., and Watson, S.B. 2015. "Effects of
542 increasing nitrogen and phosphorus concentrations on phytoplankton community
543 growth and toxicity during *Planktothrix* blooms in Sandusky Bay, Lake Erie."
544 *Environmental Science & Technology* 49(12):7197–7207. doi:
545 10.1021/acs.est.5b00799.
- 546 Deslous-Paoli, J., Souchu, P., Mazouni, N., Juge, C., and Dagault, F. 1998. "Relationship
547 between environment and resources: impact of shellfish farming on a Mediterranean
548 lagoon (Thau, France)." *Oceanologica Acta* 6(21):831–43.
- 549 Drake, L.A., Meyer, A.E., Forsberg, R.L., Baier, R.E., Doblin, M.A., Heinemann, S.,
550 Johnson, W.P., Koch, M., Rublee, P.A., and Dobbs, F.C. 2005. "Potential invasion of
551 microorganisms and pathogens via 'Interior hull fouling': biofilms inside ballast water
552 tanks." *Biological Invasions* 7(6):969–82. doi: 10.1007/s10530-004-3001-8.
- 553 Fiandrino, A., Ouisse, V., Dumas, F., Lagarde, F., Pete, R., Malet, N., Le Noc, S., and De
554 Wit, R. 2017. "Spatial patterns in coastal lagoons related to the hydrodynamics of
555 seawater intrusion." *Marine Pollution Bulletin* 119(1):132–44. doi:
556 10.1016/j.marpolbul.2017.03.006.

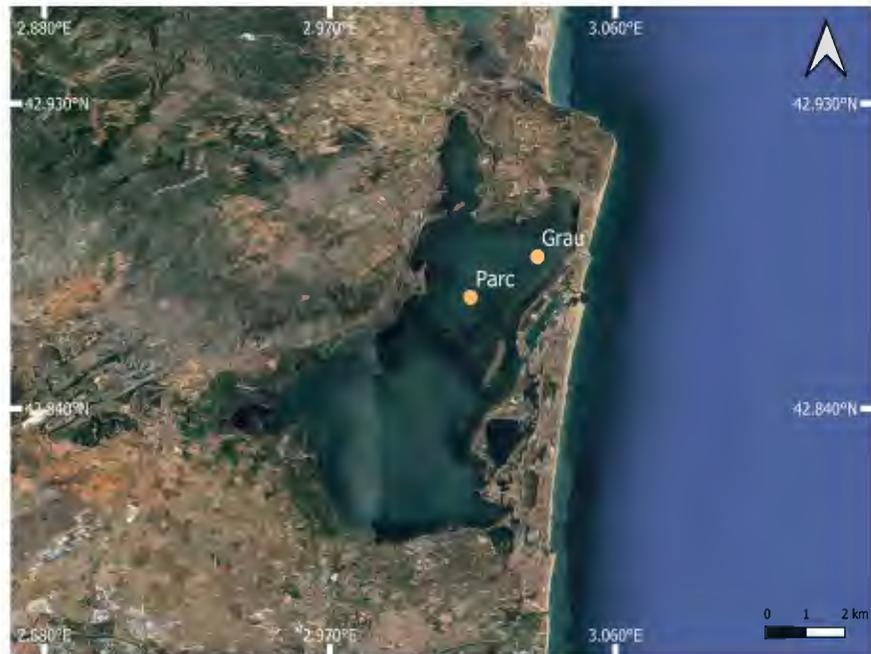
- 557 Figueroa, R.I., Estrada, M., and Garcés, E. 2018. “Life histories of microalgal species causing
558 harmful blooms: Haploids, diploids and the relevance of benthic stages.” *Harmful*
559 *Algae* 73:44–57. doi: 10.1016/j.hal.2018.01.006.
- 560 Franks, P.J.S. 2018. “Recent advances in modelling of harmful algal blooms.” Pp. 359–77 in
561 *Global Ecology and Oceanography of Harmful Algal Blooms, Ecological Studies*,
562 edited by P. M. Glibert, E. Berdalet, M. A. Burford, G. C. Pitcher, and M. Zhou.
563 Cham: Springer International Publishing.
- 564 Genovesi, B., Laabir, M., Masseret, E., Collos, Y., Vaquer, A., and Grzebyk, D. 2009.
565 “Dormancy and germination features in resting cysts of *Alexandrium tamarense*
566 species complex (*Dinophyceae*) can facilitate bloom formation in a shallow lagoon
567 (Thau, southern France).” *Journal of Plankton Research* 31(10):1209–24. doi:
568 10.1093/plankt/fbp066.
- 569 Gomis, C., Alcober, J., and Bernabeu, A. 1996. “Seguimiento de las poblaciones
570 fitoplanctónicas en las bateas marjillónicas del puerto de Valencia 1991–1994.”
571 *Matamoros E, Delgado M (Eds) IV Reunion Iberica de Fitoplancton Toxicó y*
572 *Biotoxinas St. Carles de La Rapita (Tarragona): Generalitat de Catalunya.*
573 *Department d’Agricultura, Ramaderia i Pesca* 29–38.
- 574 Glibert, P.M. 2020. “Harmful algae at the complex nexus of eutrophication and climate
575 change.” *Harmful Algae* 91:101583. doi: 10.1016/j.hal.2019.03.001.
- 576 González-Gil, S., Velo-Suárez, L., Gentien, P., Ramilo, I., and Reguera, B. 2010.
577 “Phytoplankton assemblages and characterization of a *Dinophysis acuminata*
578 population during an upwelling–downwelling cycle.” *Aquatic Microbial Ecology*
579 58:273–86. doi: 10.3354/ame01372.
- 580 González Vilas L., Spyrakos, E., Torres Palenzuela J.M., and Pazos, Y. 2014. “Support
581 Vector Machine-Based method for predicting *Pseudo-nitzschia* spp. blooms in coastal
582 waters (Galician Rias, NW Spain).” *Progress in Oceanography* 124:66–77. doi:
583 10.1016/j.pcean.2014.03.003.
- 584 Granéli, E., and Turner, J.T. eds. 2006. *Ecology of Harmful Algae: 15 Tables*. Berlin:
585 Springer. pp. 37-52.
- 586 Greene, W.H. 1994. Accounting for excess zeros and sample selection in Poisson and
587 Negative Binomial Regression Models. *SSRN Scholarly Paper*. ID 1293115.
588 Rochester, NY: Social Science Research Network.
- 589 Grenz, C., Alliot, E., Baudinet, D., Helis, L., and Masse, H. 1992 “Influence des opérations de
590 dévasage sur les flux de nutriments à l’interface eau-sédiment (Bassin de Thau -
591 France).” 9.
- 592 Hallegraeff, G.M., Anderson, D.M., and Cembella, A.D. 2003. “Harmful Algal Blooms : a
593 global overview.” *Manual on Harmful Marine Microalgae* 11.
- 594 Hallegraeff, G.M. 2010. “Ocean climate change, phytoplankton community responses and
595 Harmful Algal Blooms: a formidable predictive challenge.” *Journal of Phycology*
596 46(2):220–35. doi: 10.1111/j.1529-8817.2010.00815.x.

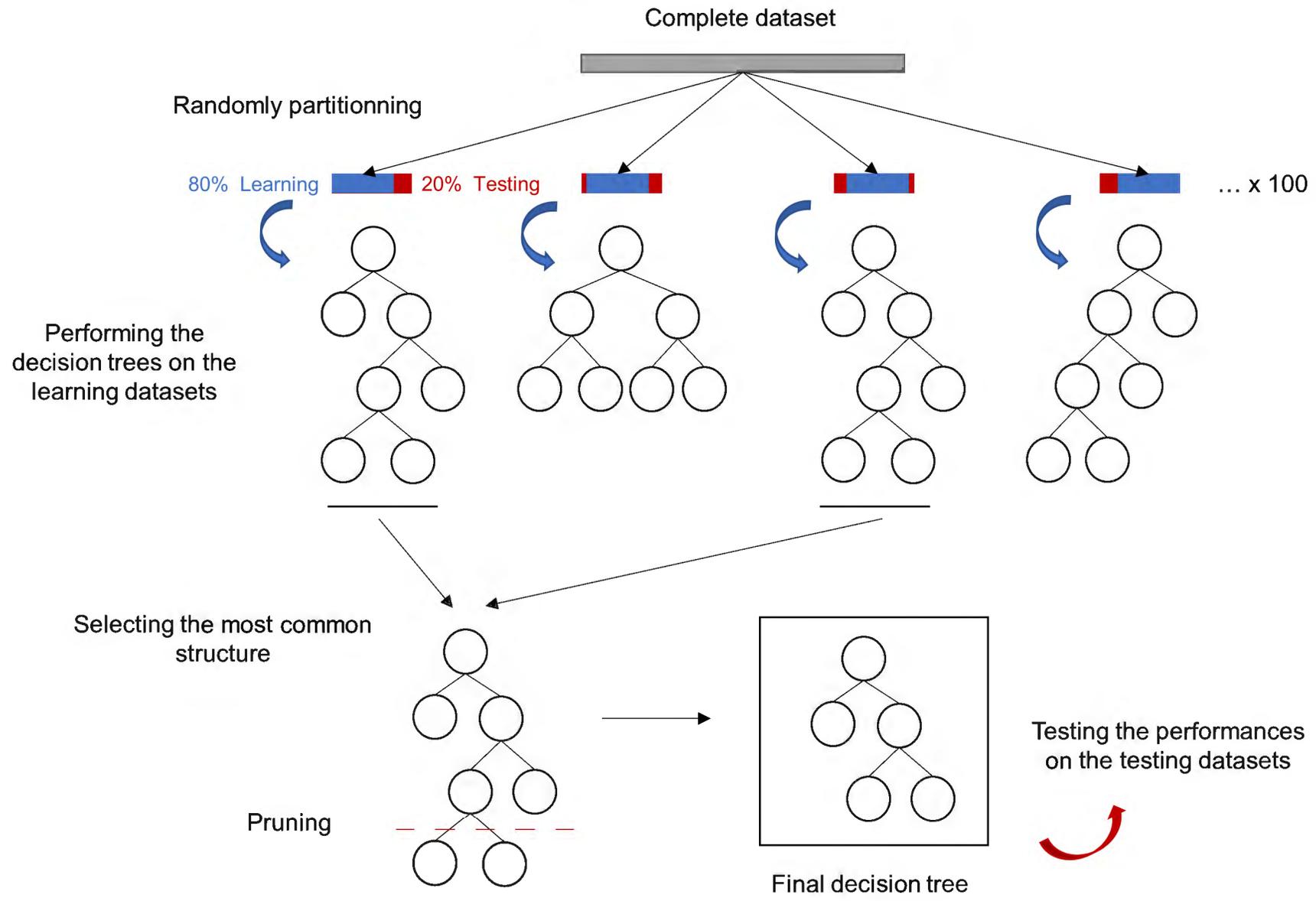
- 597 Hartig, F. 2020. "DHARMA: Residual Diagnostics for Hierarchical (Multi-level / Mixed)
598 Regression Models. R Package Version 0. 3. 3. 0."
- 599 Ishikawa, A., Hattori, M., Ishii, K.I., Kulis, D.M., Anderson, D.M., and Imai, I. 2014. "In situ
600 dynamics of cyst and vegetative cell populations of the toxic dinoflagellate
601 *Alexandrium catenella* in Ago Bay, central Japan." *Journal of Plankton Research*
602 36(5):1333–43. doi: 10.1093/plankt/fbu048.
- 603 Itakura, S., Yamaguchi, M., YOSHIDA, M., and Fukuyo, Y. 2002. "The seasonal occurrence
604 of *Alexandrium tamarense* (Dinophyceae) vegetative cells in Hiroshima Bay, Japan."
605 *Fisheries Science* 68:77–86. doi: 10.1046/j.1444-2906.2002.00392.x.
- 606 Keafer, B.A., Buesseler, K.O., and Anderson, D.M. 1992. "Burial of living dinoflagellate
607 cysts in estuarine and nearshore sediments." *Marine Micropaleontology* 20(2):147–61.
608 doi: 10.1016/0377-8398(92)90004-4.
- 609 Kim, S., Kang, Y.G., Kim, HS., Yih, W., and Coats, D. 2008. "Growth and grazing responses
610 of the mixotrophic dinoflagellate *Dinophysis acuminata* as functions of light intensity
611 and prey concentration." *Aquatic Microbial Ecology - AQUAT MICROB ECOL*
612 51:301–10. doi: 10.3354/ame01203.
- 613 Laabir, M., Jauzein, C., Genovesi, B., Masseret, E., Grzebyk, D., Cecchi, P., Vaquer, A.,
614 Perrin, Y., and Collos, Y. 2011. "Influence of temperature, salinity and irradiance on
615 the growth and cell yield of the harmful red tide dinoflagellate *Alexandrium catenella*
616 colonizing mediterranean waters." *Journal of Plankton Research* 33(10):1550–63. doi:
617 10.1093/plankt/fbr050.
- 618 Laabir, M., Barre, N., Franco, J., Brunet, C., Masseret, E., and Collos, Y. 2012.
619 "Morphological, biochemical and growth characteristics of *Alexandrium catenella*
620 (Whedon & Kofoid) Balech, a toxic dinoflagellate expanding in mediterranean
621 waters." *Cahiers De Biologie Marine* 53(3):365–72.
- 622 Laanaia, N., Vaquer, A., Fiandrino, A., Genovesi, B., Pastoureaud, A., Cecchi, P., and Collos,
623 Y. 2013. "Wind and temperature controls on *Alexandrium* blooms (2000-2007) in
624 Thau lagoon (Western Mediterranean)." *Harmful Algae* 28:31–36. doi:
625 10.1016/j.hal.2013.05.016.
- 626 Lambert, D. 1992. "Zero-Inflated Poisson regression, with an application to defects in
627 manufacturing." *Technometrics* 34(1):1–14. doi: 10.2307/1269547.
- 628 Ladagnous H., Le Bec C. 1997. Lagune de Salses – Leucate. I – Analyse bibliographique.
629 R.INT.DEL/97.02/SETE. <https://archimer.ifremer.fr/doc/00073/18422/>
- 630 Lane, J.Q., Roddam, C.M., Langlois, G.W., and Kudela, R.M. 2010. "Application of Solid
631 Phase Adsorption Toxin Tracking (SPATT) for field detection of the hydrophilic
632 phycotoxins domoic acid and saxitoxin in coastal California." *Limnology and*
633 *Oceanography-Methods* 8:645–60. doi: 10.4319/lom.2010.8.645.
- 634 Levinton, J., Ward, J., and Shumway, S. 2002. "Feeding responses of the bivalves
635 *Crassostrea gigas* and *Mytilus trossulus* to chemical composition of fresh and aged
636 kelp detritus." *Marine Biology* 141(2):367–76. doi: 10.1007/s00227-002-0830-9.

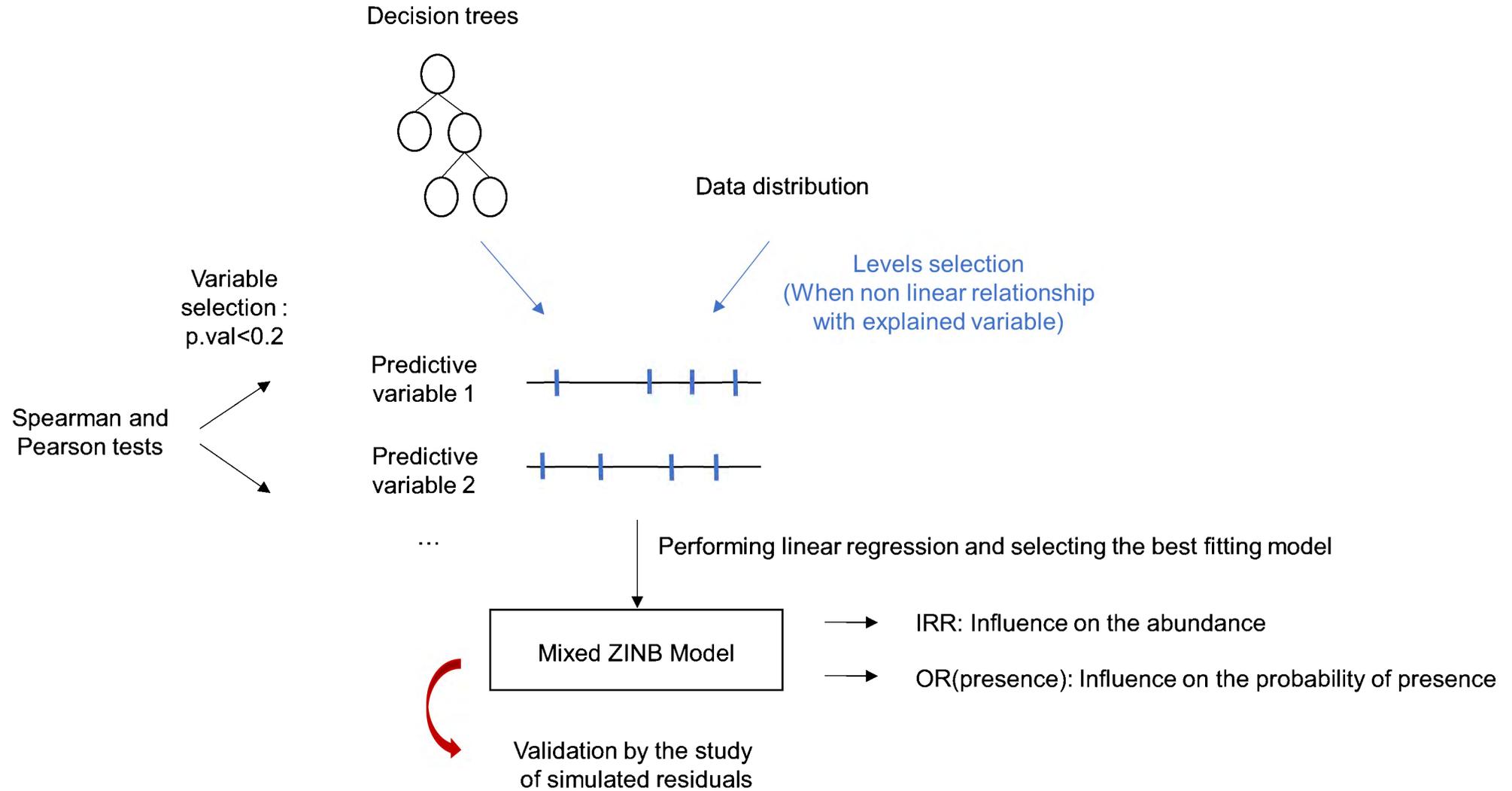
- 637 Natsuike, M., Shiraishi, T., Ishii, KI., Yamamoto, K., Nakajima, M., Sawayama, S., and Imai,
638 I. 2018. “Different nutrient availabilities of surface and bottom water under nutrient-
639 depleted conditions during bloom formation of the toxic dinoflagellate *Alexandrium*
640 *tamarense* in Osaka Bay, Japan.” 北海道大学水産科学研究彙報 68(1):7–16.
- 641 Ninčević G., Skejic, Z.S, Buzancic, M., Marasović, I., Arapov, J., Ujević, I., Bojanić, N.,
642 Grbec, B., Kušpilić, G., and Vidjak, O. 2008. “Seasonal variability in *Dinophysis* spp.
643 abundances and diarrhetic shellfish poisoning outbreaks along the eastern Adriatic
644 Coast.” *Botanica Marina - BOT MAR* 51:449–63. doi: 10.1515/BOT.2008.067.
- 645 Ogata, T., Ishimaru, T., and Kodama, M. 1987. “Effect of water temperature and light
646 intensity on growth rate and toxicity change in *Protogonyaulax tamarensis*.” *Marine*
647 *Biology* 95(2):217–20. doi: 10.1007/BF00409008.
- 648 Park, M.G., Kim, S., Kim, H.S., Myung, G., Kang, Y.G., and Yih, W. 2006. “First successful
649 culture of the marine dinoflagellate *Dinophysis acuminata*.” *Aquatic Microbial*
650 *Ecology* 45(2):101–6. doi: 10.3354/ame045101.
- 651 Park, S., Lee, S.R., Park, J.H., Park, Y., Jung, M.A., Pyo, S.J., and Kim, K.S. 2011.
652 “Prediction of red tide blooms using Decision Tree model.” Pp. 710–13 in *ICTC 2011*.
- 653 Pousse, E., Flye-Sainte-Marie, J., Alunno-Bruscia, M., Hegaret, H., and Jean, F. 2018.
654 “Sources of paralytic shellfish toxin accumulation variability in the pacific oyster
655 *Crassostrea gigas*.” *Toxicon* 144:14–22. doi: 10.1016/j.toxicon.2017.12.050.
- 656 Pulido, O. 2016. “Phycotoxins by Harmful Algal Blooms (HABS) and human poisoning: an
657 overview. *Int Clin Pathol J* 2(6): 00062. DOI:10.15406/icpjl.2016.02.00062.”
658 *International Clinical Pathology Journal* 2. doi: 10.15406/icpjl.2016.02.00062.
- 659 R Core Team. 2020. “R: A language and environment for statistical computing.”
- 660 Ralston, D.K., Keafer, B.A., Brosnahan, M.L., and Anderson, D.M. 2014. “Temperature
661 dependence of an estuarine Harmful Algal Bloom: resolving interannual variability in
662 bloom dynamics using a degree-day approach.” *Limnology and Oceanography*
663 59(4):1112–26. doi: 10.4319/lo.2014.59.4.1112.
- 664 Ralston, D.K., Brosnahan, M.L., Fox, S.E., Lee, K.D., and Anderson, D.M. 2015.
665 “Temperature and residence time controls on an estuarine Harmful Algal Bloom:
666 modeling hydrodynamics and *Alexandrium fundyense* in Nauset Estuary.” *Estuaries*
667 *and Coasts* 38(6):2240–58. doi: 10.1007/s12237-015-9949-z.
- 668 Raso, S., Van Genugten, B., Vermuë, M., and Wijffels, R.H. 2012. “Effect of oxygen
669 concentration on the growth of *Nannochloropsis* Sp. at low light intensity.” *Journal of*
670 *Applied Phycology* 24(4):863–71. doi: 10.1007/s10811-011-9706-z.
- 671 Reguera, B., Velo-Suarez, L., Raine, R., and Park, M.G. 2012. “Harmful *Dinophysis* species: a
672 review.” *Harmful Algae* 14:87–106.
- 673 REPHY - French Observation and Monitoring program for Phytoplankton and Hydrology in
674 coastal waters. 2019. “REPHY Dataset - French Observation and Monitoring Program
675 for Phytoplankton and Hydrology in Coastal Waters. 1987-2018 Metropolitan Data.”

- 676 REPHYTOX - French Monitoring program for Phycotoxins in marine organisms. 2019.
677 “REPHYTOX Dataset. French Monitoring Program for Phycotoxins in Marine
678 Organsims. Data since 1987.”
- 679 Séchet, V., Sibat, M., Billien, G., Carpentier, L., Rovillon, G-A., Raimbault, V., Malo, F.,
680 Gaillard, S., Perrière-Rumebe, M., Hess, P., and Chomérat, N. 2021. “Characterization
681 of toxin-producing strains of *Dinophysis* Spp. (Dinophyceae) isolated from french
682 coastal waters, with a particular focus on the *D. Acuminata*-Complex.” *Harmful Algae*
683 107:101974. doi: 10.1016/j.hal.2021.101974.
- 684 Seubert, E.L., Gellene, A.G., Howard, M.D.A., Connell, P., Ragan, M., Jones, B.H., Runyan,
685 J., and Caron, D.A. 2013. “Seasonal and annual dynamics of harmful algae and algal
686 toxins revealed through weekly monitoring at two coastal ocean sites off southern
687 California, USA.” *Environmental Science and Pollution Research* 20(10):6878–95.
688 doi: 10.1007/s11356-012-1420-0.
- 689 Shin, J., Yoon, S.H., and Cha, Y. 2017. “Prediction of cyanobacteria blooms in the lower Han
690 River (South Korea) using ensemble learning algorithms.” *DESALINATION AND*
691 *WATER TREATMENT* 84:31–39. doi: 10.5004/dwt.2017.20986.
- 692 Swan, S.C., Turner, A.D., Bresnan, E., Whyte, C., Paterson, R.F., McNeill, S., Mitchell, E.,
693 and Davidson, K. 2018. “*Dinophysis acuta* in scottish coastal waters and its influence
694 on diarrhetic shellfish toxin profiles.” *Toxins* 10(10). doi: 10.3390/toxins10100399.
- 695 Therneau, T., and Atkinson, B. 2019. “Rpart: Recursive Partitioning and Regression Trees. R
696 Package Version 4.1-15.”
- 697 Utermöhl, H. 1931. “Neue wege in der quantitativen erfassung des planktons: (mit besondere
698 berücksichtigung des ultraplanktons).” *Verh. Int. Ver. Theor. Angew. Limnol* 5:567–
699 95.
- 700 Valbi, E., Ricci, F., Capellacci, S., Casabianca, S., Scardi, M., and Penna, A. 2019. “A model
701 predicting the PSP toxic dinoflagellate *Alexandrium minutum* occurrence in the coastal
702 waters of the NW Adriatic Sea.” *Scientific Reports* 9(1):4166. doi: 10.1038/s41598-
703 019-40664-w.
- 704 Verity, P.G., Villareal, T.A., and Smayda, T.J. 1988. “Ecological investigations of blooms of
705 colonial *Phaeocystis pouchetti* . II. The role of life-cycle phenomena in bloom
706 termination.” *Journal of Plankton Research* 10(4):749–66. doi:
707 10.1093/plankt/10.4.749.
- 708 Xu, N., Duan, S., Li, A., Zhang, C., Cai, Z., and Hu, Z. 2010. “Effects of temperature, salinity
709 and irradiance on the growth of the harmful dinoflagellate *Prorocentrum donghaiense*
710 Lu.” *Harmful Algae* 9(1):13–17.

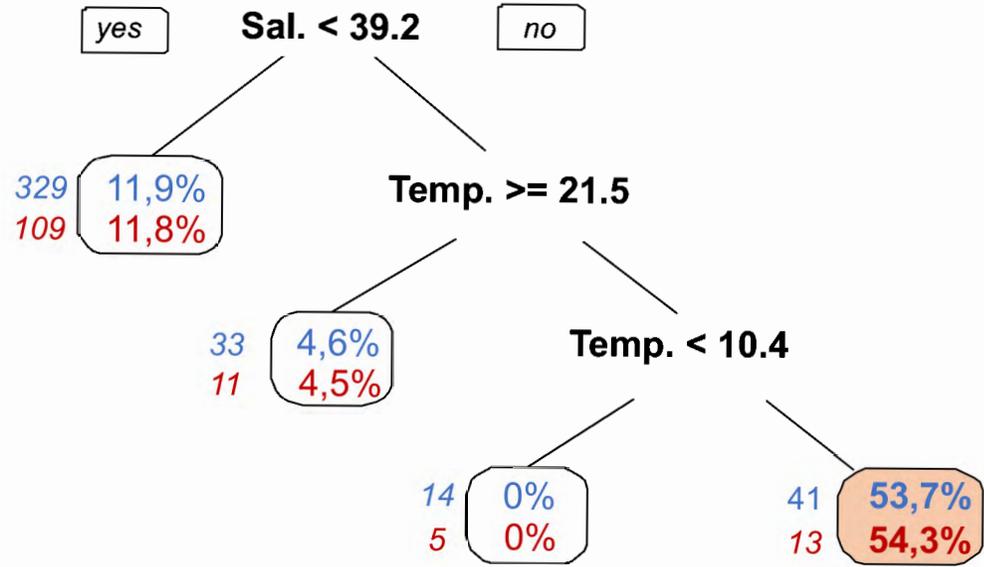
711



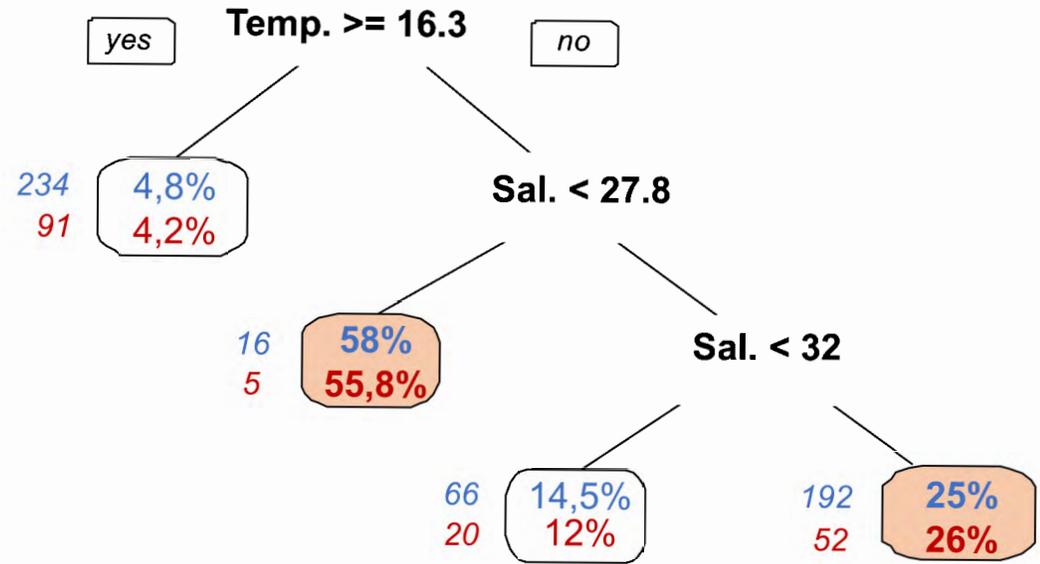


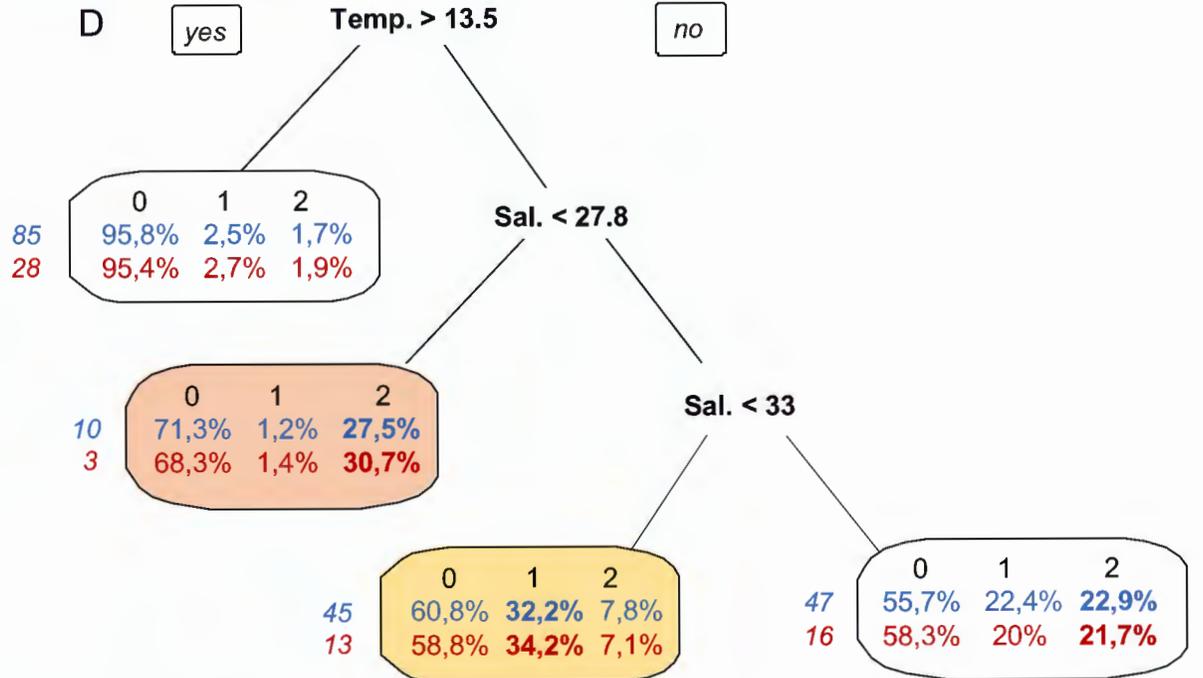
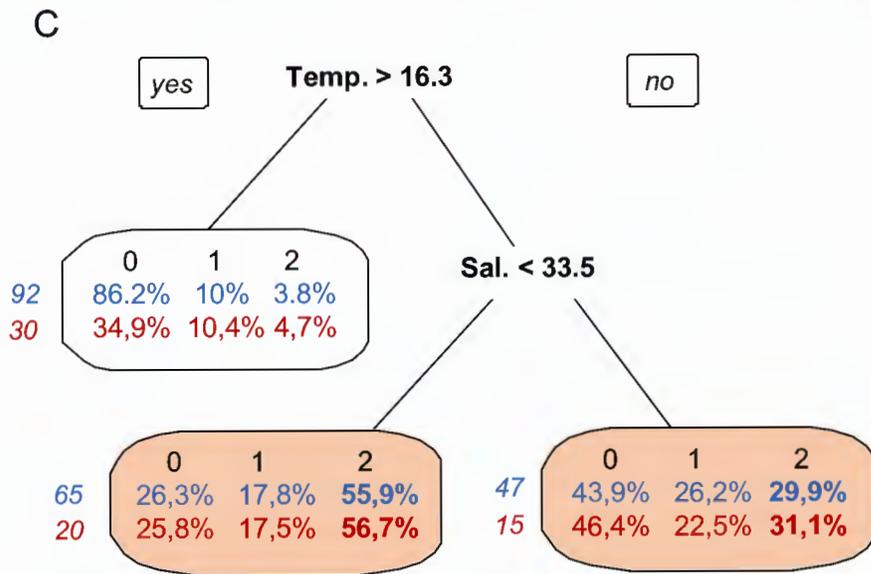
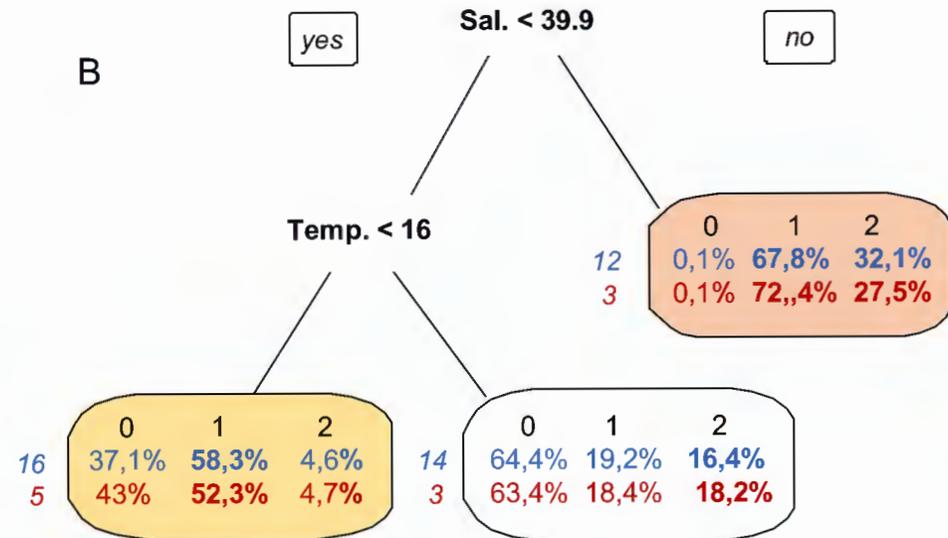
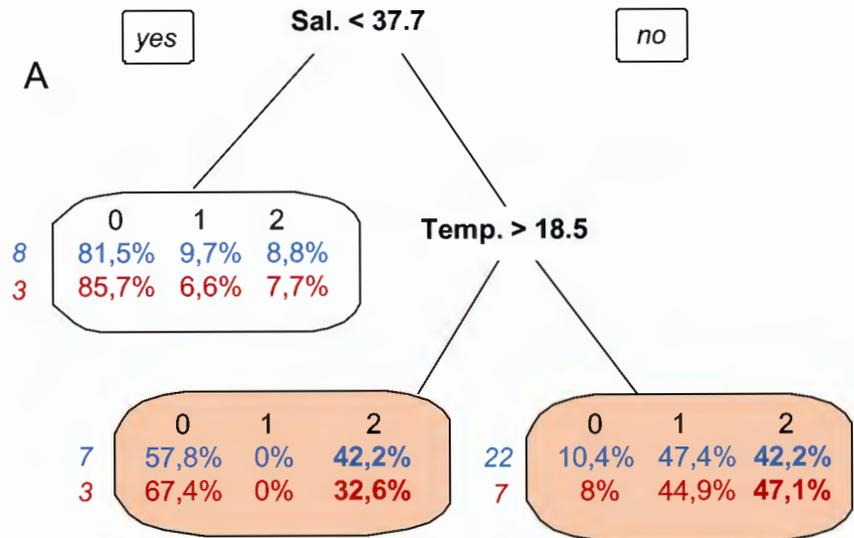


A

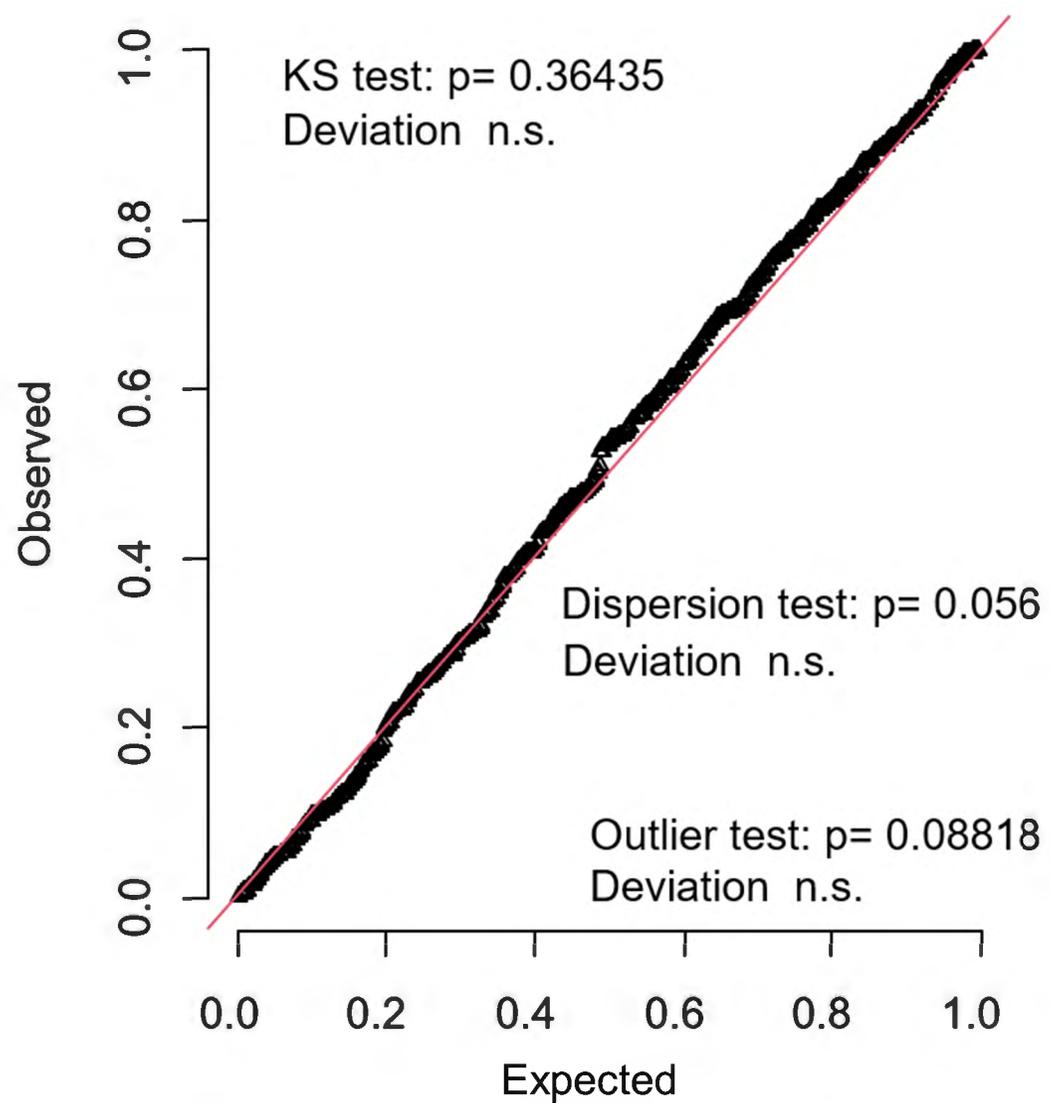


B





A



B

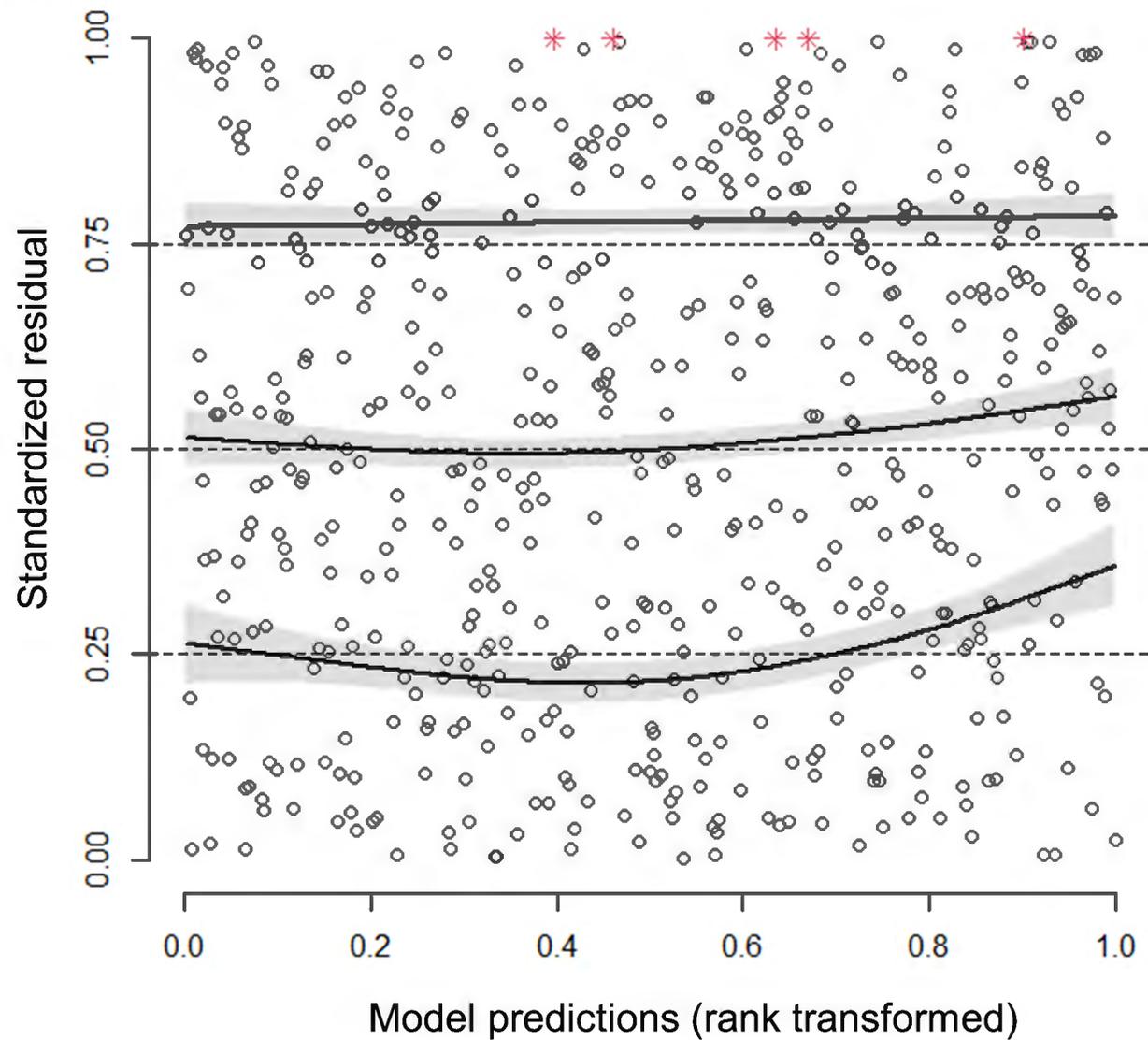


Table 1 : Correlation coefficients between *A. tamarense/pacificum*, *Dinophysis*, diatoms, other dinoflagellates,

Thau lagoon

	<i>A. tamarense/ pacificum</i>	Diatoms	Dinoflagellates	Sal.	Temp.	Turb.
<i>A. tamarense/ pacificum</i>		-0.01	0.05	0.04	0.02	0.04
		<u>0.18</u>	<u>0.29</u>	<u>0.17</u>	<i>0.02</i>	<i>0.05</i>
PST	0.41	-0.11	-0.2	0.3	-0.18	-0.14
(m)	<i>0.17</i>	-0.23	-0.05	<u>0.33</u>	<u>-0.38</u>	<u>-0.31</u>
PST	0.58	-0.18	-0.12	0.27	0.04	0.15
(o)	<u>0.44</u>	-0.19	-0.06	<u>0.41</u>	-0.04	-0.16

abiotic factors (temperature, salinity and turbidity) and toxins in Thau and Leucate lagoons from 2010 to 2019.

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).

Leucate lagoon

	<i>Dinophysis</i>	Diatoms	Dinoflagellates	Sal.	Temp.	Turb.
<i>Dinophysis</i>		-0.01	-0.04	<u>-0.21</u>	<u>-0.23</u>	0.03
		0.09	0.08	<u>-0.22</u>	<u>-0.31</u>	<u>0.15</u>
OA	0.45	-0.06	-0.06	<u>-0.28</u>	<u>-0.38</u>	0.08
(m)	<u>0.67</u>	-0.15	-0.07	<u>-0.34</u>	<u>-0.51</u>	<u>0.27</u>
OA+PTX+DTX	0.46	-0.06	-0.06	<u>-0.28</u>	<u>-0.39</u>	0.08
(m)	<u>0.66</u>	-0.15	-0.08	<u>-0.35</u>	<u>-0.51</u>	<u>0.27</u>
OA	0.5	-0.08	-0.05	-0.14	<u>-0.39</u>	0.13
(o)	<u>0.69</u>	-0.2	-0.1	<u>-0.19</u>	<u>-0.56</u>	<u>0.28</u>
OA+PTX+DTX	0.52	-0.08	-0.06	-0.12	<u>-0.38</u>	0.12
(o)	<u>0.69</u>	<u>-0.22</u>	-0.13	<u>-0.17</u>	<u>-0.56</u>	<u>0.28</u>

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. ϵ [10.4; 14.6[-0.78	0.45	0.030 *	0.76	0.46	6.02e-3 **
Temp. ϵ [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. ϵ [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.