REPHY time series from other sites

September 6, 2017

1 Choices

1.1 Filtering in Quadrige

We queried the Quadrige database on the 18th of April 2017. We filtered data to consider only results from the REPHY network (REPHY-Etudes and REPHYTOX were ignored as they are either too short or focused on toxin concentrations) from 1987 to the end of 2016. Among the 1077 sites where such measurements have been made, we chose to focus on the 31 sites which were already used in **(author?)** [5]. Dunkerque, Boulogne and Somme sites were not all present in the database as most of them actually belonged to the SRN network (Suivi Régional des Nutriments). The SRN time series were downloaded on the 11th of May 2017. In the following, we excluded one of the SRN sampling site (Mismer) from further analyses as it often changed location and bathymetry without notifications. We can see the impact of using both REPHY and SRN data for the same sampling sites (Point 1 Dunkerque, Point 1 Boulogne, At So and Bif) on Fig. 1. For most sites (Dunkerque, At So, Bif), the SRN data complete and sometimes duplicate the observations included in the REPHY programme. However, observations in Boulogne are very different for two sampling dates in SRN and REPHY. In all cases, we only take SRN values when none exists for the same dates in the REPHY.

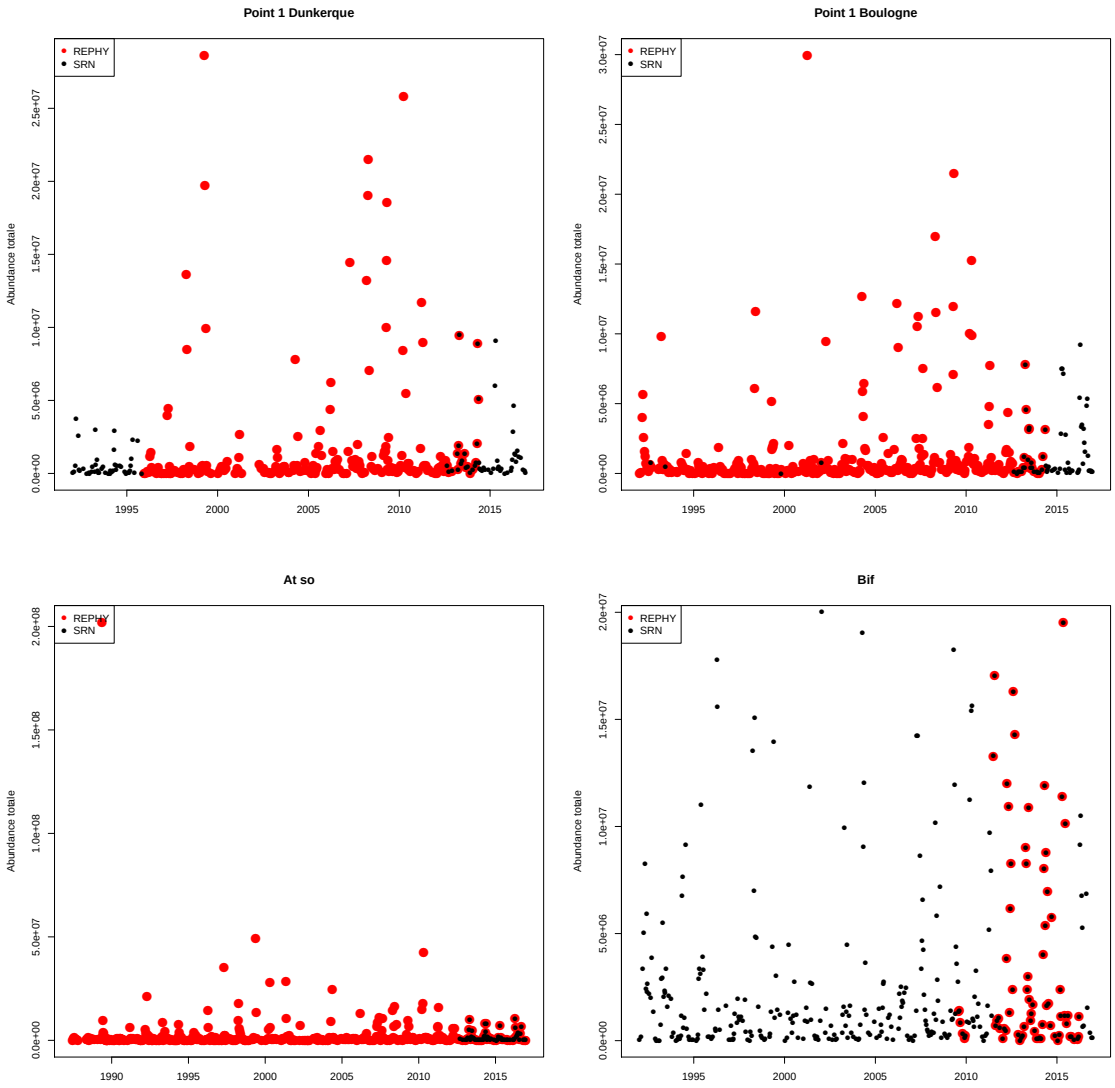


Figure 1: Comparison of total phytoplankton abundance for each site monitored by both REPHY and SRN.

In addition to the sites presented in **(author?)** [5], we looked for overseas sites: in Mayotte, there is one monitored site (Passi Keli), and there were four of them on the Réunion island. Thau-Crique de l'Angle was also retrieved.

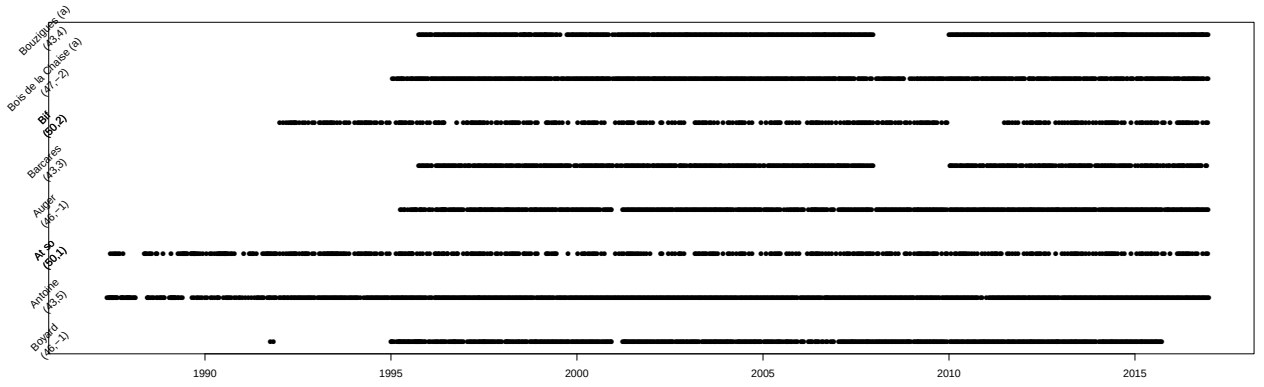
Several observation networks could complement the REPHY network. We also queried data from:

- INRH: a phytoplankton monitoring along the Moroccan coast. However, none of the 66 concerned sites could be retrieved
- OBSERVATOIRE-EPOC: only the Comprians site is shown, without giving access to data
- DYMPAHY: Ferry-Box observation network, none of the 184 home ports forming part of this study gives access to data
- RINBIO, anounced as a 'biological variable' monitoring, does not follow phytoplankton abundance
- AAMP: phytoplanktonic surveillance, leads to a void extraction

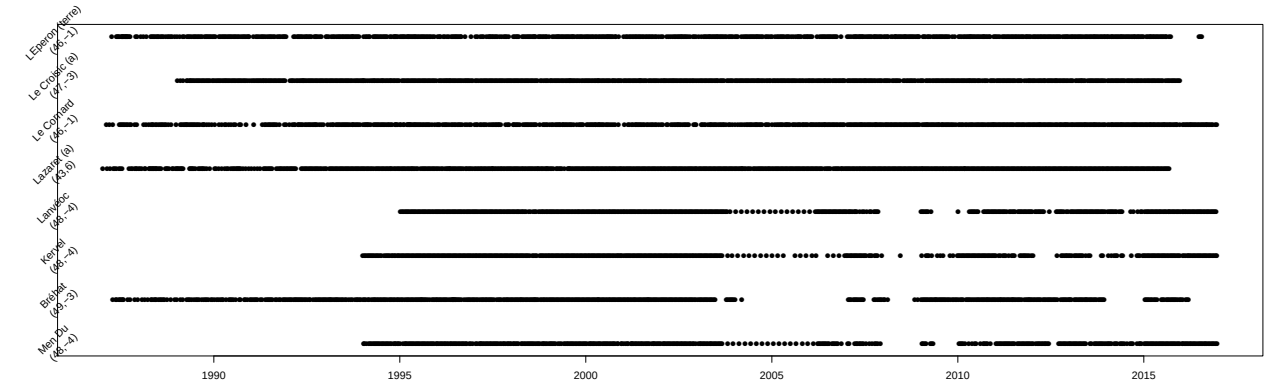
We only used data which were either qualified as 'good' or 'unqualified'. For certain sites and certain periods of times, plankton samples were collected from both the top and bottom of the water column. We only kept values from the top of the water column[1].

[1] We can discuss this. We can average them, we can use only bottom, we can begin with a study for each site of the assemblages for

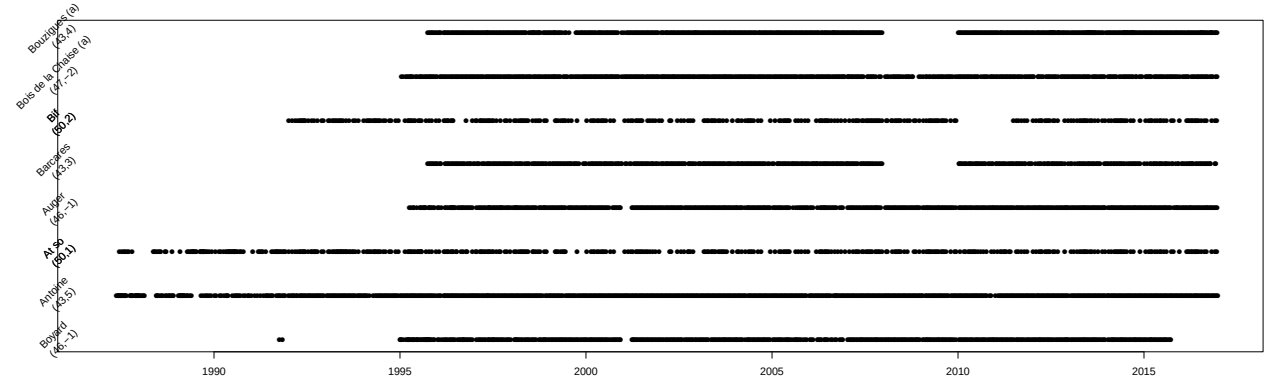
1.2 Filtering "best" time series to work with



(a) Overseas sites + Thau



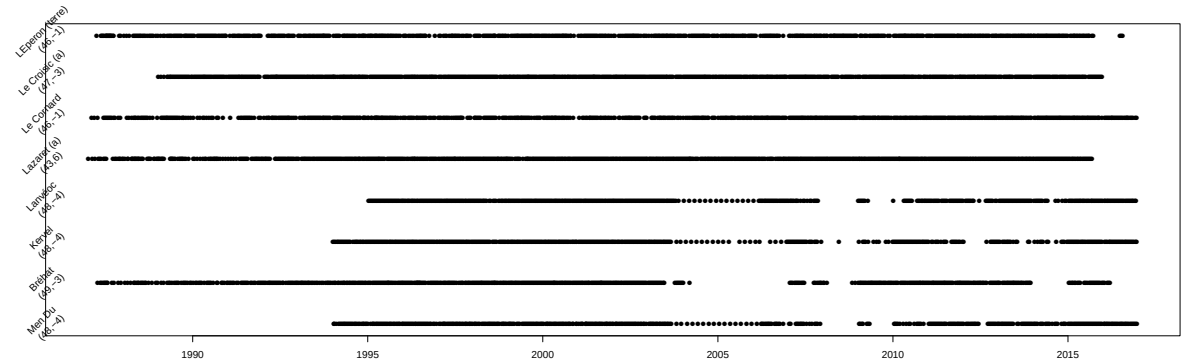
(b) Overseas sites + Thau



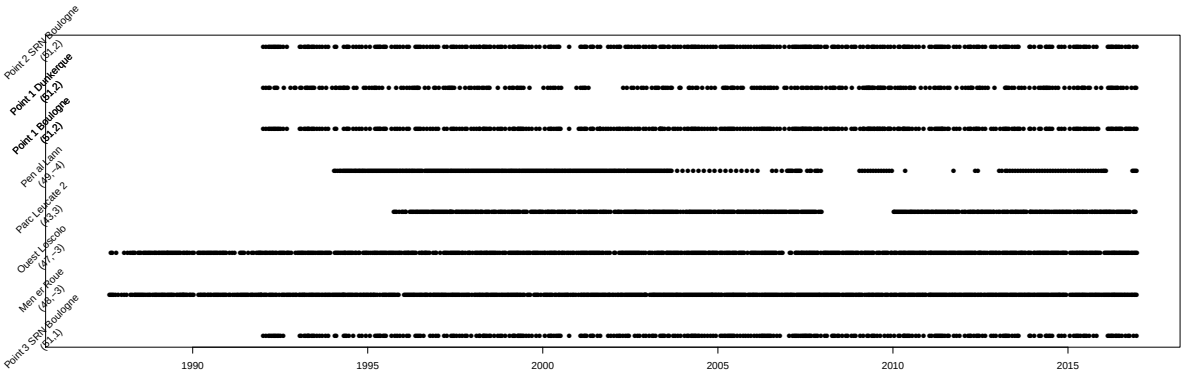
(c) REPHY + SRN (from Tania)

Figure 2: Sampling dates belonging to the REPHY observation network for all sites retrieved from Quadrige, all variables cofounded. All sites are identified by their name and coordinates (latitude,longitude).

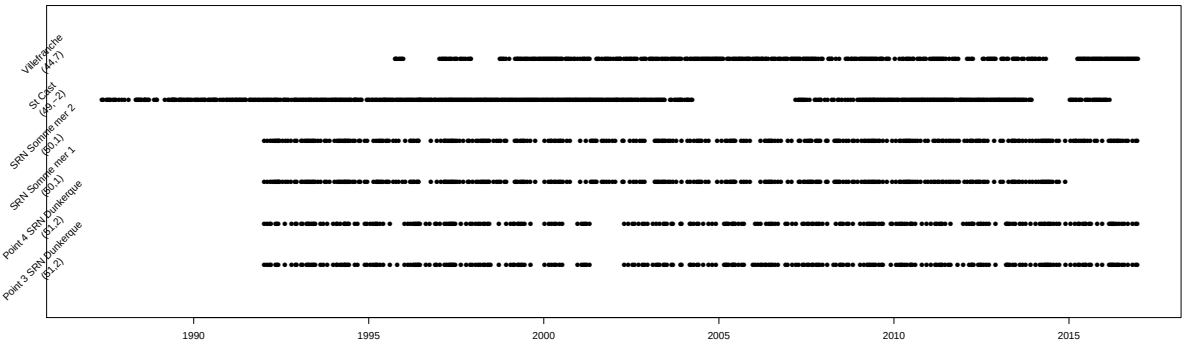
both bottom and top...



(d) REPHY + SRN (from Tania)



(e) REPHY + SRN (from Tania)

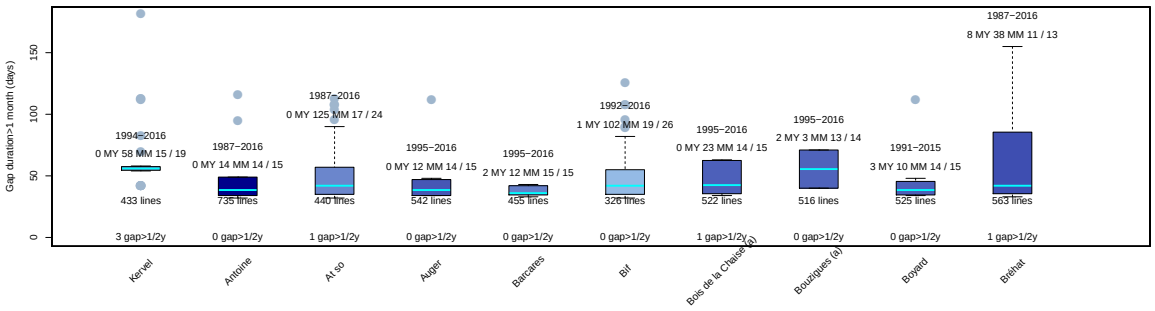


(f) REPHY + SRN (from Tania)

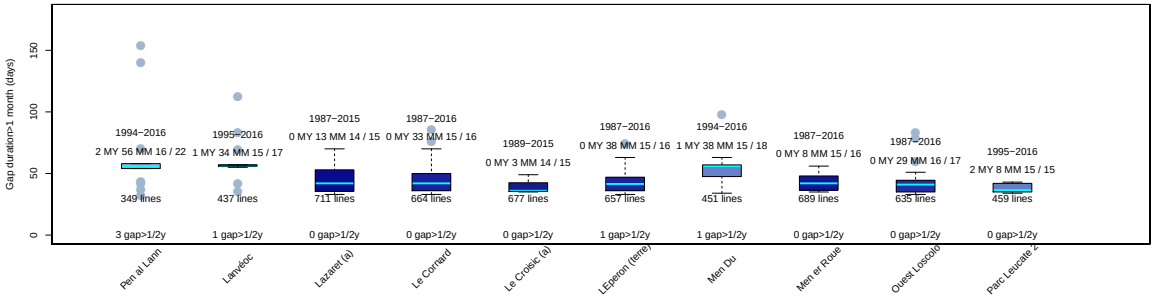
Figure 2

In Fig. 2, we can see that most time series are incomplete, but some of them are really gappy. All overseas time series would be hard to analyze within an autoregressive framework. When considering each variable for each site [2], we can also see that Thau has no monitoring of the whole phytoplanktonic population (FLORTOT variable) and that apart from planktonic groups, only temperature and salinity are common variables for all sampling sites.

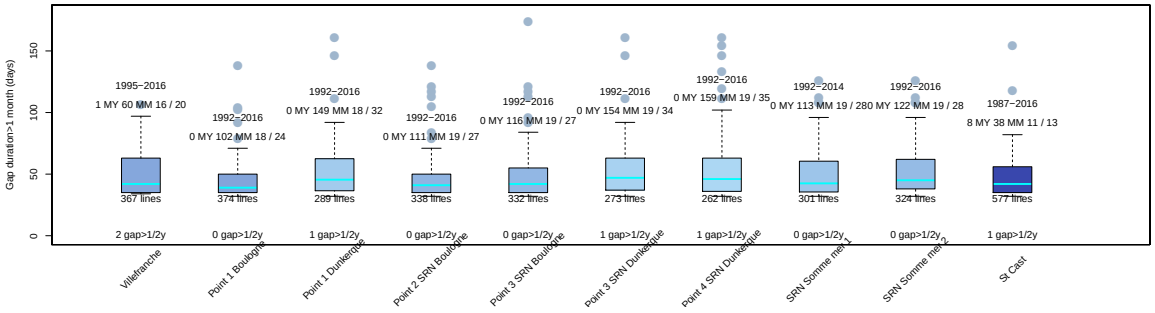
[2]Not included in the document, but all figures can be found in the graph directory under the name *_var_time.pdf



(g) Distribution of the duration between two sampling dates for 10 sites monitored under the REPHY and SRN surveys (more details at the bottom of the figure)



(h) Distribution of the duration between two sampling dates for 10 sites monitored under the REPHY and SRN surveys (more details at the bottom of the figure)



(i) Distribution of the duration between two sampling dates for 10 sites monitored under the REPHY and SRN surveys (more details at the bottom of the figure)

Figure 3: Distribution of the duration of time gaps superior to 1 month and inferior to 1 year for each sampling site (in days). For visual comparison, only gaps lasting less than 175 days are shown in the figure, but this information is included is the text below the boxplots. The text at the top indicate the starting and ending years (or the last available year on the Quadrige database), the number of missing years (MY), missing month (MM) and the average duration between two sampling dates, only taking durations under 1 month into account (corresponding to the prescribed survey conditions), versus the average duration between dates in the whole time series. Finally, the background color of each boxplot ranges from light blue to dark blue according to the total number of points in the times series (this information is also shown at the bottom of each boxplot).

The search for long (more than 10 years after 1995/1996, when an important formation took place for plankton observers, whose traces can be more or less conspicuous in all time series), continuous (no missing year), regularly sampled (average time between sampling around 15 days) time series can be seen on Fig. 3[3]. Finally, the following sites were chosen: Antoine, Lazaret (Mediterranean sea), Auger, L'Eperon, Le Cornard (Pertuis Charentais), Le Croisic, Men er Roue, Loscolo (South Brittany). They are located on Fig. 4.[4]

[3]Not included in the document, but the distribution of gaps according to season is shown in gap_per_season.pdf. An example of bias can be seen for At so, for which autumn sampling seems difficult

[4]We might want to add a sampling site in the North, close to the English Channel. The database from At so is the richest, but it is still gappy. There are other defects: dinoflagellates counts are given to the unit (seems impossible to me), counting zeros which is not the case at other places.

With a closer look on the quality of monitoring for each site[5], we first observe that the 1995 training that was relevant for the Arcachon Bay actually impacts most of the other sampling sites. I would advise to take results only after this date (see for example the strange peaks in abundance at Antoine, PSE observations at Loscolo, or RHI observations at Cornard...). What is more, Bois de la Chaise has been strongly impacted by its location shift in 2007: either we consider only values before this date, or we just don't use it (we already have data from South Brittany). I also have doubts for Croisic site: the RHI and CRY groups may indicate a change in observer thay may strongly impact the quality of the time series.

Planktonic groups vary slightly between sites. While CHA is a common species at each sampling site, the abundance of AST varies between sites and it is not always a dominant group, which makes the comparison of interactions between these two groups more difficult. Cryptophytes and Euglenophytes are not followed with the same rigour than in Arcachon Bay and we may want to ignore them in further analyses, especially as we revealed no significant effect of these groups in our previous study.

As far as methods are concerned, we can see that some sites followed PSE as a toxic species, which means that the observer looked for this specific group and included '0' instead of missing value when it could not be found (therefore there are both '0' and missing values in the time series). This is the same for AST in Bois de la Chaise and dinoflagellates in At So, but there is no "toxicity excuse" for this. What is more, some sites have a less-than-usual variability for lower abundance, especially for dinoflagellates (see for instance Auger): we might think that it is also observer-related and we may want to introduce a small noise around this value.

I suggest that we work on Auger, L'Eperon, Cornard in Marennes Oléron, Antoine, Lazaret in the Mediterranean Sea, and Loscolo, Men er Roue and Croisic in Brittany.

1.3 Which groups should we focus on

1.3.1 Taxonomy

The number of observations for each site is presented in Table 1. Except for Antoine, the length of the time series and the number of different taxa is homogeneous among sites. When pooling all sites, 660 different taxa have been identified during the REPHY monitoring. Retrieving phylogenetic information for each organism is the first step towards the construction of relevant groups for community dynamics analysis[6]. We used WORMS database[7] with an automated access thanks to the taxize package in R, so that updates in phylogeny, which can be common in plankton classification, can be passed on to our own database[8].

	Dates	Taxa	Observations
Men er Roue	689	241	15828
Loscolo	635	251	15622
Large Croisic	677	291	20408
Eperon	657	243	14799
Cornard	664	240	13996
Auger	542	255	14691
Antoine	380	153	7315
Lazaret	711	194	a12568

Table 1: Number of sampling dates and different taxa for each site. Each observation corresponds to the detection of a taxon on a specific date. The generic term taxon can represent a species, a group of species, a genus, or even a phylum.

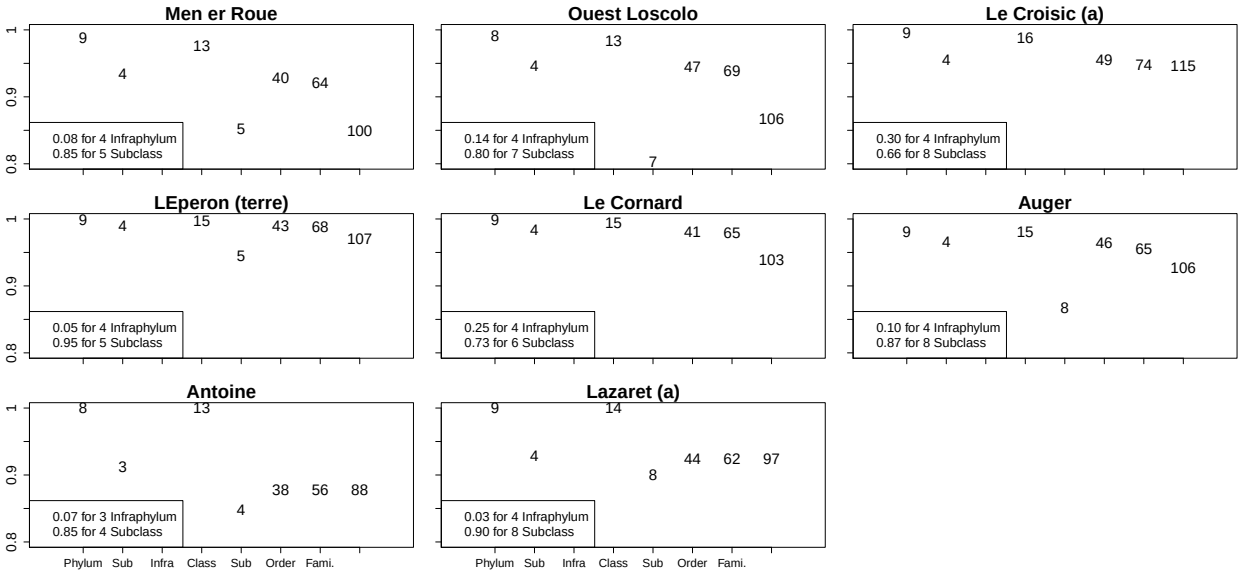
We can see in Fig. 5 that we can identify more than 85% of the observed cells at the genus level at all sites.

[5]Not included in the main document, can be found under the name *_per_species.pdf

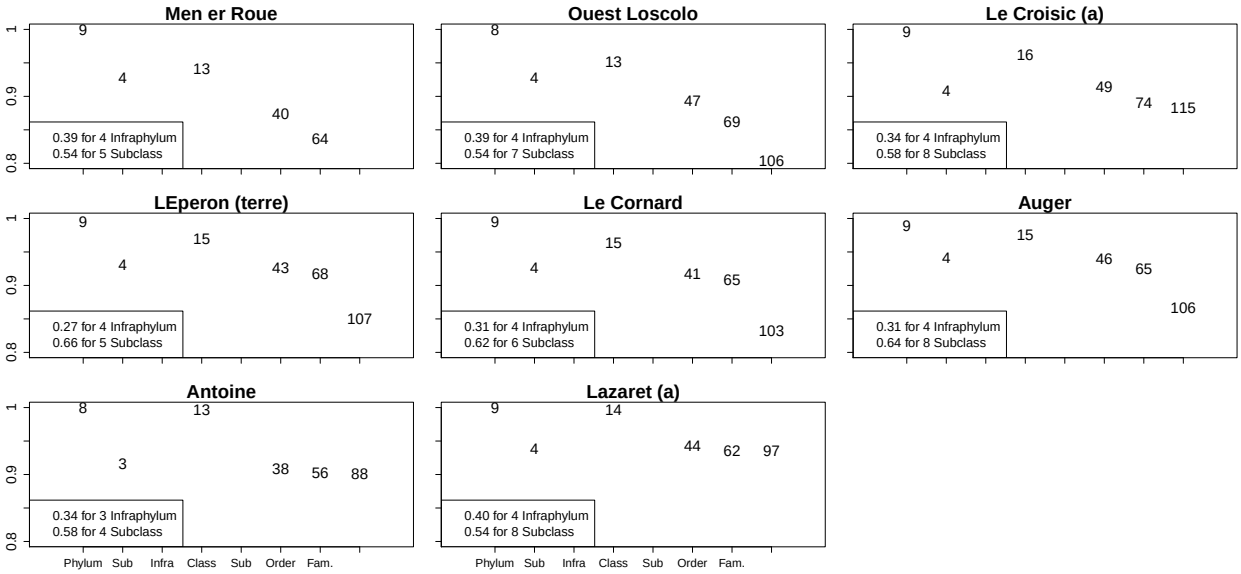
[6]We could also try functional groups... if we had such information

[7]http://www.marinespecies.org/

[8]This led to the 'db_taxonomy_REPHY.csv' file which translates each identifier used in the REPHY monitoring to its classification in WORMS and can be found in the same folder



(a) Relative abundance of identified taxa for different taxonomic levels for each site. Numbers indicate the different groups which can be found for a taxonomic level. For example, in Men er Roue, 4 different subphyla are identified throughout the whole time series, and about 95% of all taxa are identified at the subphylum level whereas 85% of all taxa are identified at the genus level, among 100 genera or groups of genera. The same information is given in boxes at the bottom left of each graph for infraphylum and subclass levels for graphical purposes (values were much lower).



(b) Relative number of observations for different taxonomic levels for each site. For example, 80% of the observations in Ouest Loscolo identify the genus of the observed organism (note that, when focusing on abundance, this value is above 85%).

Figure 5: Percentage of the total abundance (a) and the total number of observations (b) that can be kept according to the required taxonomic level.

Genus level Using groups of genera, we still have to choose which genus we should focus on.

Arcachon :

In the folder 'graphe/Arcachon'[9], we used the same planktonic groups as in our paper for Arcachon (defined in Table 2)

Code	Genus
AST	Asterionella+Asterionellopsis+Asteroplanus
CHA	Chaetoceros
CRY	Cryptophytes
EUG	Euglenophytes
GUI	Guinardia
GYM	Gymnodiniaceae+Amphidinium+Gyrodinium+Katodinium
LEP	Leptocylindrus
NIT	Ceratoneis+Nitzschia+Hantzschia+Bacillaria
PRP	Protoperidinium+Peridinium
PSE	Pseudo-nitzschia
RHI	Rhizosolenia+Neocalyptrella
SKE	Skeletonema

Table 2: Definition of groups of genera for 'Arcachon' classification

However, we can see that there are much more missing observations than in Arcachon Bay for certain groups, especially AST, CHA and EUG. Replacing missing values would be more challenging if we kept these groups.

[9]I did not include the files in the main document as there are many, but they can be found in the graph/Arcachon/* folder. The caption is always the same : Time series for log abundance of observed groups of genera at X sampling site (X given by the name of the file). The color of the point corresponds to the relative abundance of the group at a given sampling date: red corresponds to 95-to-100% of a given group whereas cyan corresponds to 25-to-35% of the overall abundance, the exact number being given in the top left corner, below the percentage of missing values. Th. NEI always corresponds to 'Not Elsewhere Identified'.

Hernandez : **(author?)** [4] defined much more groups of genera than us, based on several REPHY sites (Table 3). We used exactly the same definitions to see if some groups should be kept[10].

Code	Genus
AST	Asterionella+Asterionellopsis+Asteroplanus
BID	Biddulphia+Trigonium
CER	Cerataulina
CHA	Chaetoceros
COS	Coscinodiscus+Stellarima
DAC	Dactyliosolen
DIT	Ditylum
EUC	Eucampia+Climacodium
GUI	Guinardia
LAU	Lauderia+Schroederella
LEP	Leptocylindrus
NAV	Navicula+Fallacia+Haslea+Lyrella+Petroneis
NIT	Nitzschia+Hantzschia
ODO	Odontella
PARs	Paralia
PLE	Pleurosigma+Gyrosigma
PSE	Pseudo-nitzschia
RHI	Rhizosolenia+Neocalyptrella
SKE	Skeletonema
THP	Thalassiosira+Porosira
THL	Thalassionema+Lioloma
ALE	Alexandrium
CEI	Ceratium+Neoceratium
DIP	Diplopsalis+Diplopelta+Diplopsalopsis+Preperidinium+Oblea
GON	Gonyaulax
GYM	Gymnodinium+Gyrodinium
HET	Heterocapsa
KAT	Katodinium
NOC	Noctiluca
POL	Polykrikos
PRO	Prorocentrum
PRP	Protoperidinium+Archaeperidinium+Peridinium
SCR	Scrippsiella+Ensiculifera+Pentapharsodinium+Bysmatrum
PHA	Phaeocystis
DIC	Dictyocha

Table 3: Definition of groups of genera according to Hernandez classification

We present in Fig. 6 the number of groups we could use in the following analyses according to the requirements we have on the time series and the abundance of the considered groups.

[10]Same as before: time series can be found in the graphe/Hernandez folder

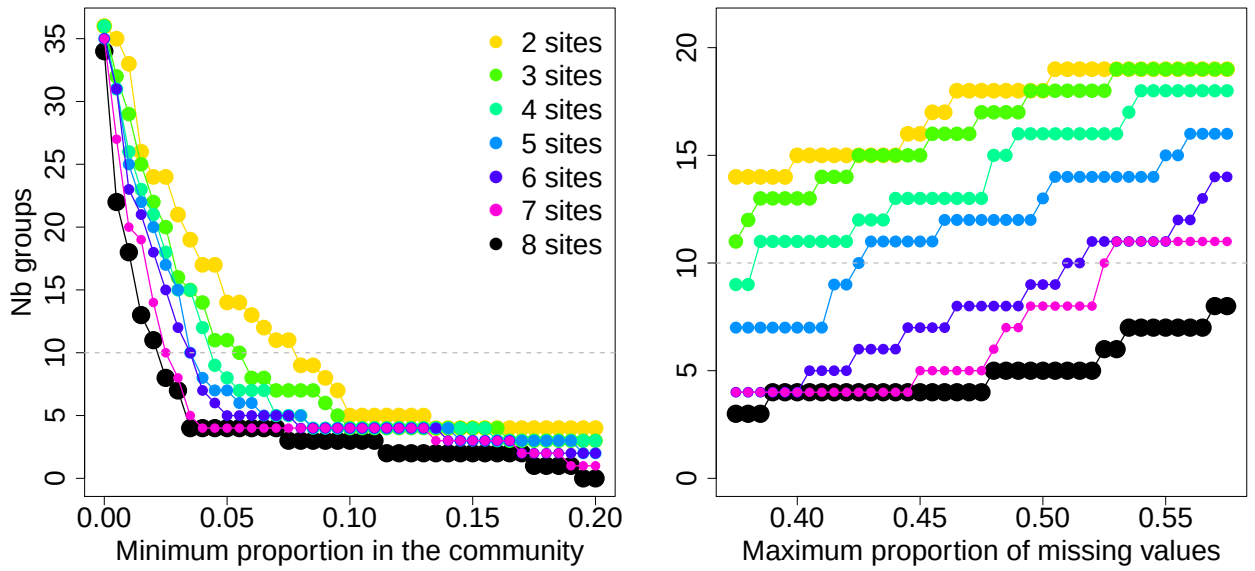


Figure 6: Number of planktonic groups whose average proportion in the community is above a threshold given on the x-axis (left) or whose maximum proportion of missing values is below a threshold on the x-axis (right) for 2 to 8 out of the 8 sites we consider. For example, black dots show that a minimum proportion of 0.05% in the planktonic community at all sites can be reached for only 4 groups, whereas if we relax the condition to a minimum proportion of 0.05% in 3 sites, 11 groups are concerned (green dots). Note that the 'NEI' group, while not very accurate for interaction analyses, is also included.

Subclass level According to WORMS, there are three main subclasses among diatoms: Coscinodiscophycidae correspond to Centric organisms whereas Bacillariophycidae and Fragilariophycidae correspond to pennate organisms[11]. We can see in Fig. 5 that subclasses are identified for more than 80% of the biomass, except in Le Croisic, where this value is lower (66%). The distinction between such groups is therefore debatable. This is the closest I found to so-called 'morpho-functional groups' [8], and to the difference we found in the first paper. Times series can be seen in graphe/Subclass **Discuss this.**

2 Preliminary MAR analyses

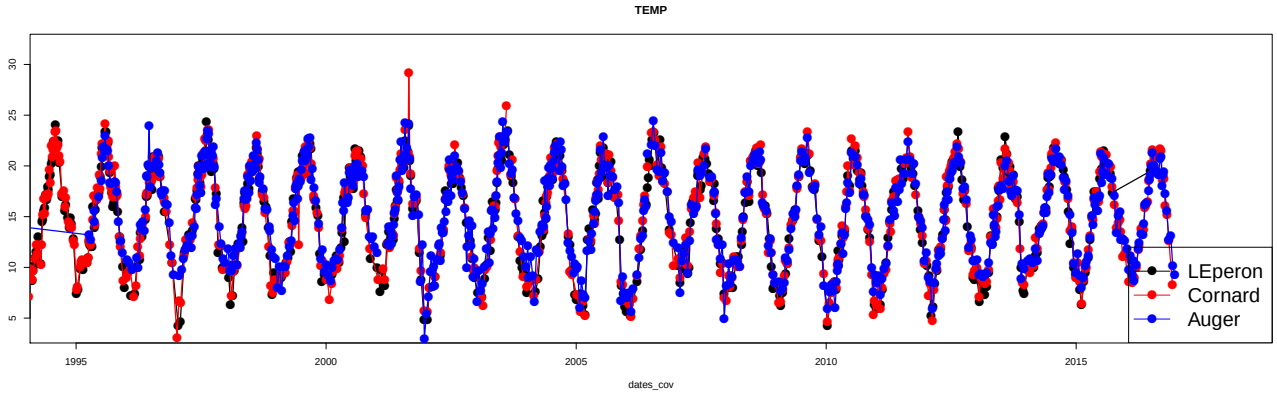
In this section, we first focus on each site separately (later, we can consider different sites as repeated observations of the same process in South Brittany -BZ/Marennes Oléron -MO/the Mediterranean Sea -SU, that can be used to estimate missing data; or just test the existence of a metapopulation. This will be discussed in next section).

2.1 Data prep

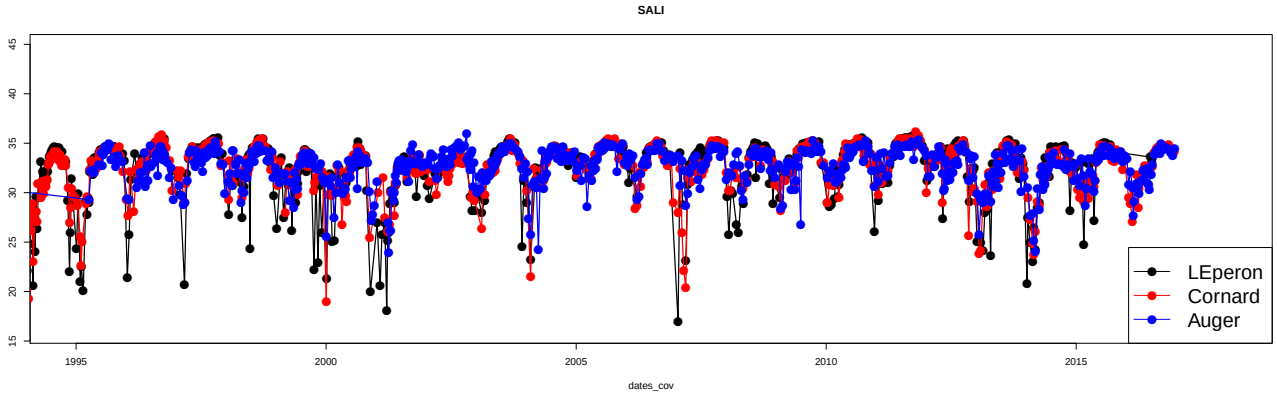
2.1.1 Abiotic variables

There are less environmental variables in Marennes Oléron than in Arcachon. We chose to focus on temperature and salinity, which were monitored at all sites and sampling dates.

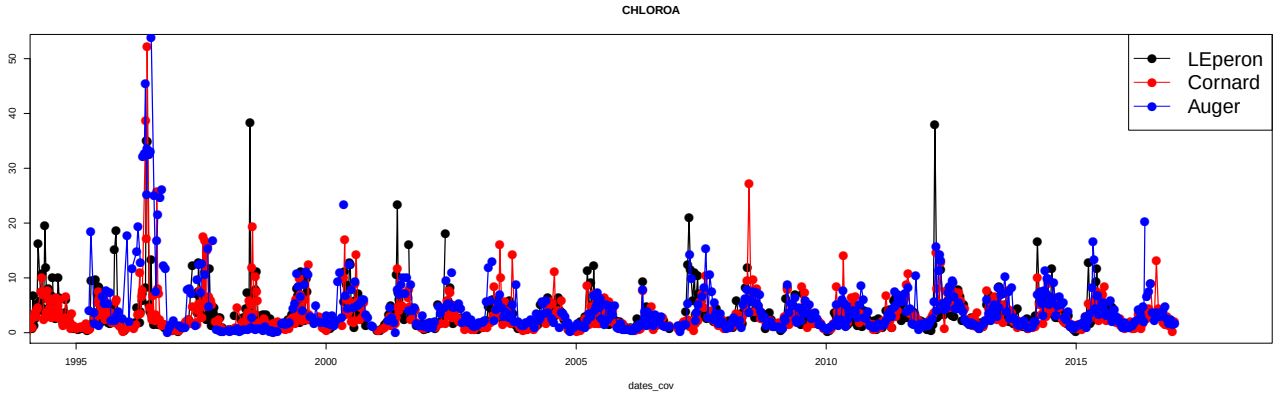
[11] See time series in the graphe/Subclass folder



(a) Temperature



(b) Salinity



(c) Chlorophyll a

Figure 7: Time series of temperature, salinity and chlorophyll a at three sampling points in Marennes Oléron.

Despite similar trends, it seems that L'Eperon and Cornard are prone to stronger decreases in salinity, maybe due to a different bathymetry or a proximity to freshwater flow that is not reflected in the salinity in Auger. A plankton proxy such as chlorophyll also shows the inter-site variability of planktonic dynamics.[12]

[12]Similar figures can be found under the names Rapport/graphe/compare_hydrology_*.pdf and Rapport/graphe/compare_plankton_species*.pdf. I think this can help explain the differences in coefficients, especillay for the effect of salinity - temperature is a bit more stable and might both cover temperature/irradiance related effects, and seasonal effects. What is more, chl a in South Brittany seems pretty high to me and might be discussed...

Spearman correlations between covariates are between 0.46 and 0.60, except for the Mediterranean sites: in Antoine, it is 0.004 and 0.04 in Lazaret (see graphe/covar_temp_sali.pdf).

A few words about Météo France These environmental values (temperature, salinity) could be completed or replaced by MeteoFrance data, using the same variables that we used in the first paper. We can see on Fig. 4 that all sampling sites seem surrounded by observation sites. However, if we want wind-related variables as well as irradiance, there are only 7 Météo France which can be used for our REPHY sampling points. In South Brittany, only Ploërmel is complete, and it is located from 55.7 to 72.8 km away from REPHY sampling sites. For Marennes-Oléron sites, both Royan and Saintes (or even Fontenay ?) can be used. For the southernmost sites, both Arles and Bormes les Mimosa can be used, but the latter is 40 km away from its closest REPHY sampling point, Lazaret. For now, it seems difficult to complete with observed data from buoys, as they mostly monitor wind and waves, but not irradiance (but nebulosity could be used as a proxy. This would mean, however, a small departure from our first analysis).

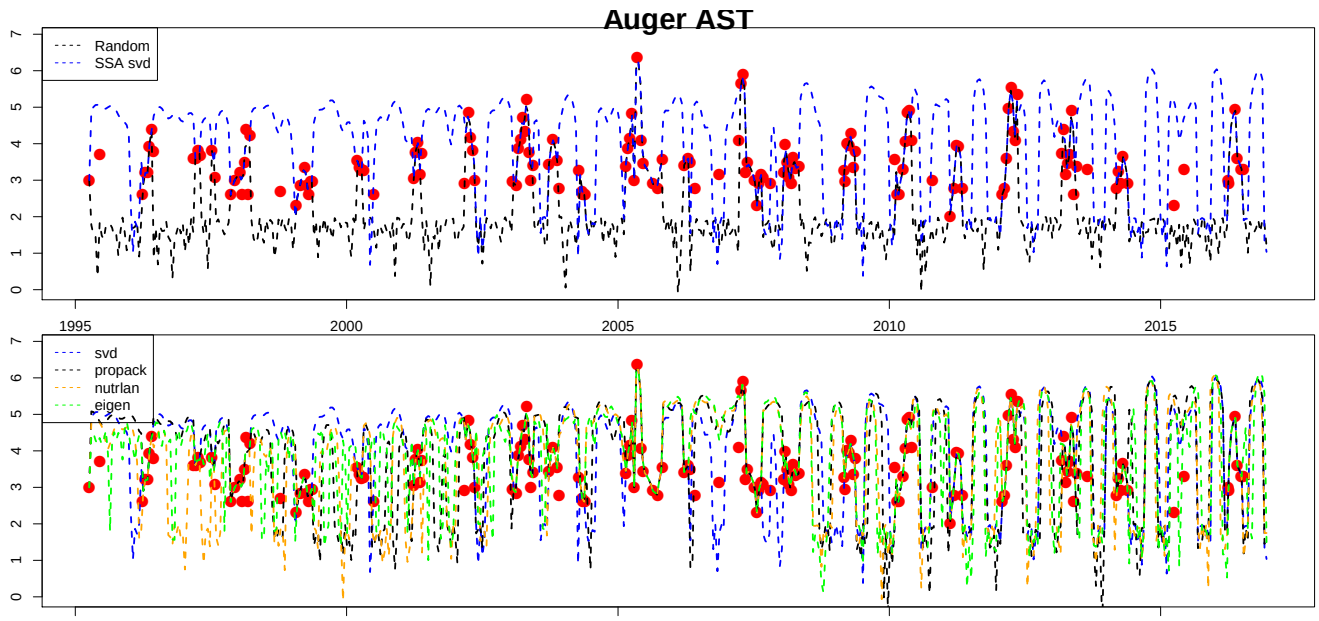
This led us to the conclusion that we can choose not to use Météo France data and keep temperature and salinity which are measured at the exact same time and place of plankton samples. Irradiance and winds can vary a lot according to the local landform, from the coast to inner land, and the distance between climatic observations and planktonic sampling might be too big to model the environmental conditions of growth for planktonic growth. What is more, irradiance and temperature are correlated: it might not be useful to use the former instead of the later. Wind energy only had a small effect in previous estimations: we may be able to drop that variable for now.

2.1.2 Biotic variables

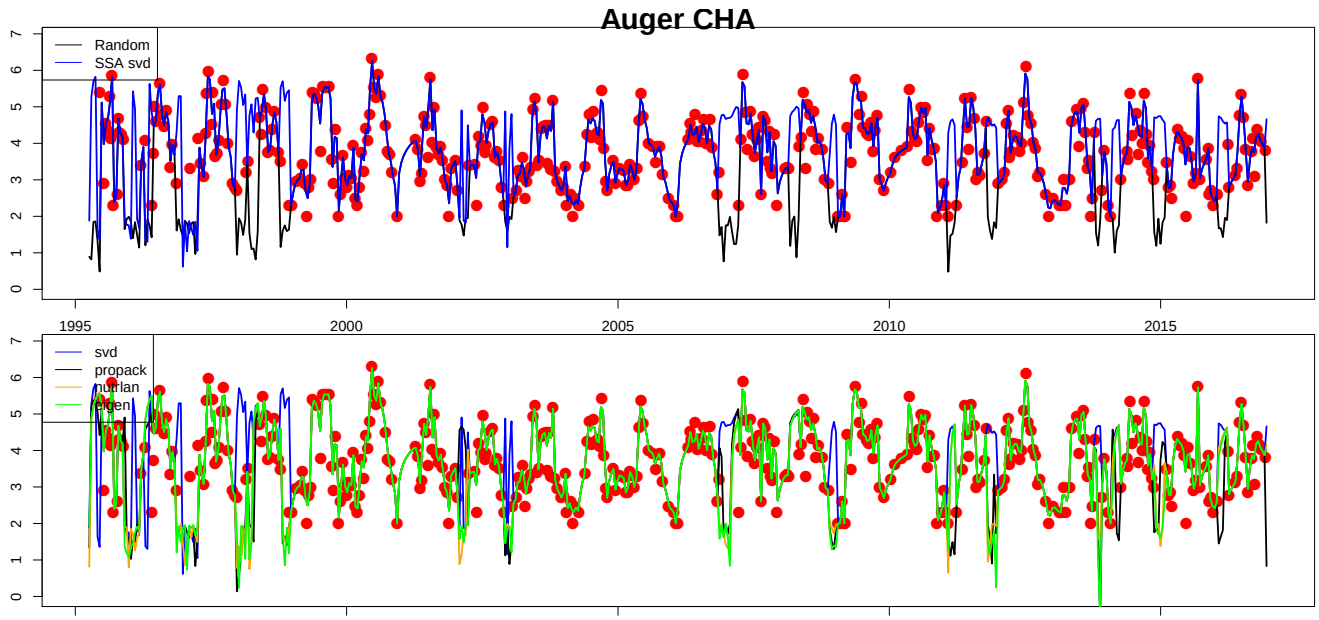
We considered planktonic time series which had less than 50% missing values[13] and used two different methods to reconstruct the dynamics of all species, by both handling missing values and irregular sampling. The first method is the one used in our first paper, based on **(author?)** [3]: missing values were replaced by a random value between 0 and the minimum value of the time series. The second method, Singular Spectral Analysis, can be described as a principal component analysis in the time domain, and is based on the spectral decomposition of the signal to fill gaps in the data while keeping its temporal structure (**(author?)** [7] ; see also **(author?)** [6][14] who compare SSA to other methods in time series analyses). We can see on Fig. 8 the effect of the different methods for two time series in Auger. For all sites and species, see files in Rapport/graphe/comparaison_SSA_random_*.pdf.

[13] Based on the time series which were chosen by Fred

[14] From EPOC :-)



(a) Comparison between observed data (red dots) and reconstructed times series (dashed lines) for a 'rare' species in Auger (more details below).



(b) Comparison between observed data (red dots) and reconstructed times series (solid lines) for a 'common' species in Auger (more details below)

Figure 8: Different methods were used for the reconstruction: the so-called 'Random' method (black lines) is based on **(author?)** [3]'s recommandation (replacing missing values by random values between 0 and the minimum value of the time series) while SSA is based on the temporal autocorrelation of the system to fill gaps. In the bottom panel, different algorithms of the SSA are compared.

I finally chose to keep the first 'random' method but this could be discussed later on. I also feel that we may be interested in the comparison of the different SSA decomposition (eigen values) of one species at different sites (for instance, this might be of interest for the analysis of synchrony for CHA).

This led me to remove some of the species for which gap filling seemed too flawed (these were mostly the species for which we weren't sure at first - dotted lines in the reconstructed plots).

Men er Roue	Loscolo	Croisic	LEperon	Cornard	Auger	Antoine	Lazaret
			AST				
CHA	CHA	CHA	CHA	CHA	CHA	CHA	CHA
COS	COS		COS		COS		
							DAC
DIT	DIT	DIT	DIT	DIT	DIT		
GUI	GUI	GUI	GUI	GUI	GUI		
LEP	LEP	LEP	LEP	LEP	LEP		LEP
						NAV	
NIT	NIT	NIT	NIT	NIT	NIT	NIT	
PLE	PLE	PLE	PLE	PLE	PLE	PLE	PLE
PSE	PSE	PSE	PSE	PSE	PSE	PSE	PSE
RHI	RHI	RHI		RHI	RHI		
SKE	SKE	SKE	SKE	SKE	SKE	SKE	SKE
THP	THP	THP	THP	THP	THP	THP	
THL	THL	THL					THL
GYM	GYM	GYM			GYM	GYM	
PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO
PRP	PRP	PRP	PRP	PRP	PRP	PRP	PRP
		SCR				SCR	SCR

Table 4: Groups we can study in further MAR analyses

2.2 MAR analyses[15]

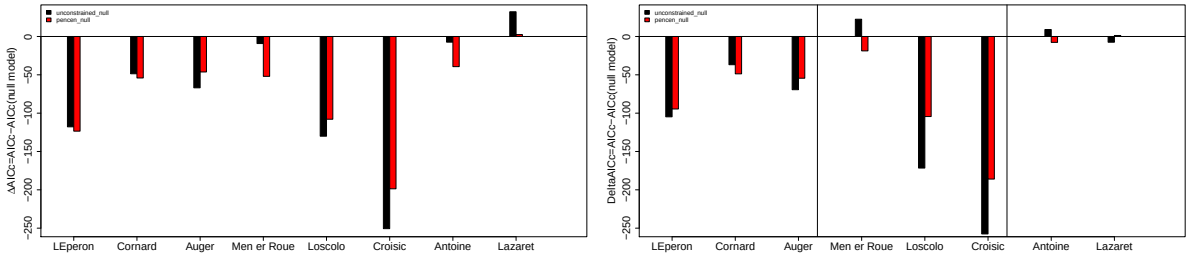
Methods were the same as in our first paper. We did not extract seasonality from temperature and salinity.

I first wondered how to use information contained in the NEI group: this can be used both as a biotic group with a temporal autocorrelation and interactions with other species, or as a covariate that can affect planktonic groups but cannot be affected by them (just like temperature and salinity). The diversity of this group from one sampling to another makes it less likely to be influenced by specific groups, or to have an homogeneous response to similar covariates: it seems more logical to see it as a covariate. However, we first completely removed it from further analyses, as was done in the previous paper.

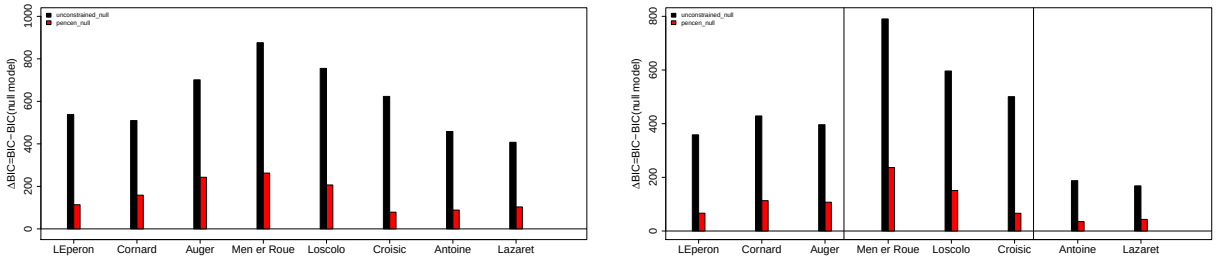
For now, no bootstrapping was used to test for significance of each parameter, this should be discussed later, when all methods are confirmed.

[15]This only presents a summary of the main results: all analyses are shown under the name Rapport/graphe/L_analyse_MAR_X_Y_Z_regular_S.pdf where L is the name of the site which is studied, when there is only one, X is the name of the model (null/unconstrained/pencen), Y is the way NEI is taken into account (as a species -sp ; as a covariate -cov; or not taken into account -null) ; Z is either 'single' or 'common' (when different sites are taken into account and we only consider common groups of species ; the group of sites is given by S, which can be eithr South Brittany -BZ, Marennes-Oléron -MR, Mediterranean Sea -SU ; or all sites -ALL).

2.2.1 Model criteria



(a) Difference of AICc for all sites and their specific groups of species (left) or keeping only common sub-groups (right)



(b) Difference of BIC for all sites and their specific groups of species or keeping only common sub-groups (right)

Figure 9: Comparison of AICc and BIC from different interaction matrices are compared. This figure shows the difference between the AICc/BIC of a given model and the corresponding criterium of the null (diagonal interaction matrix) model. A negative value indicates that the model performs better than the null model with NEI=cov. A positive value indicates that the null model is 'better'. As model structures (number of times series taken into account) are different, groups of bars should not be compared

Model criteria are not consistent. BIC is always more conservative than AICc and tend to select the null model, or the pennate/centric model, while AICc tend to select the unconstrained or pennate/centric model. It could be useful[16] to use the other interaction matrices (diatom/dino and inter-phylum) to test wether pennate-centric could still be seen as a consensus interaction matrix.

2.2.2 Coefficients

To test our observations regarding the sign and magnitude of interactions, we also pooled sites together as was done on B7 and Teychan. Results are shown below.[17]

As was found previously, there are mostly weak and positive interactions inter-groups. The only difference is in the Mediterranean Sea, where there are mostly negative interactions (this is interesting, as this might be the most polluted site we consider. This is confirmed in the 'single' analyses in which species groups differ in Antoine and Lazaret: there are only 48% positive interactions in Lazaret, whereas there are betwen 58% and 65% positive interactions elsewhere).

We also find the strong, positive influence of temperature on most groups of species, and a weak and negative effect and salinity.

[16] and will be done

[17] But all interaction matrices and AIC/BIC matrix for sites observed separately can also be found in Rapport/graphe/*. This report is only a summary of exploratory analyses

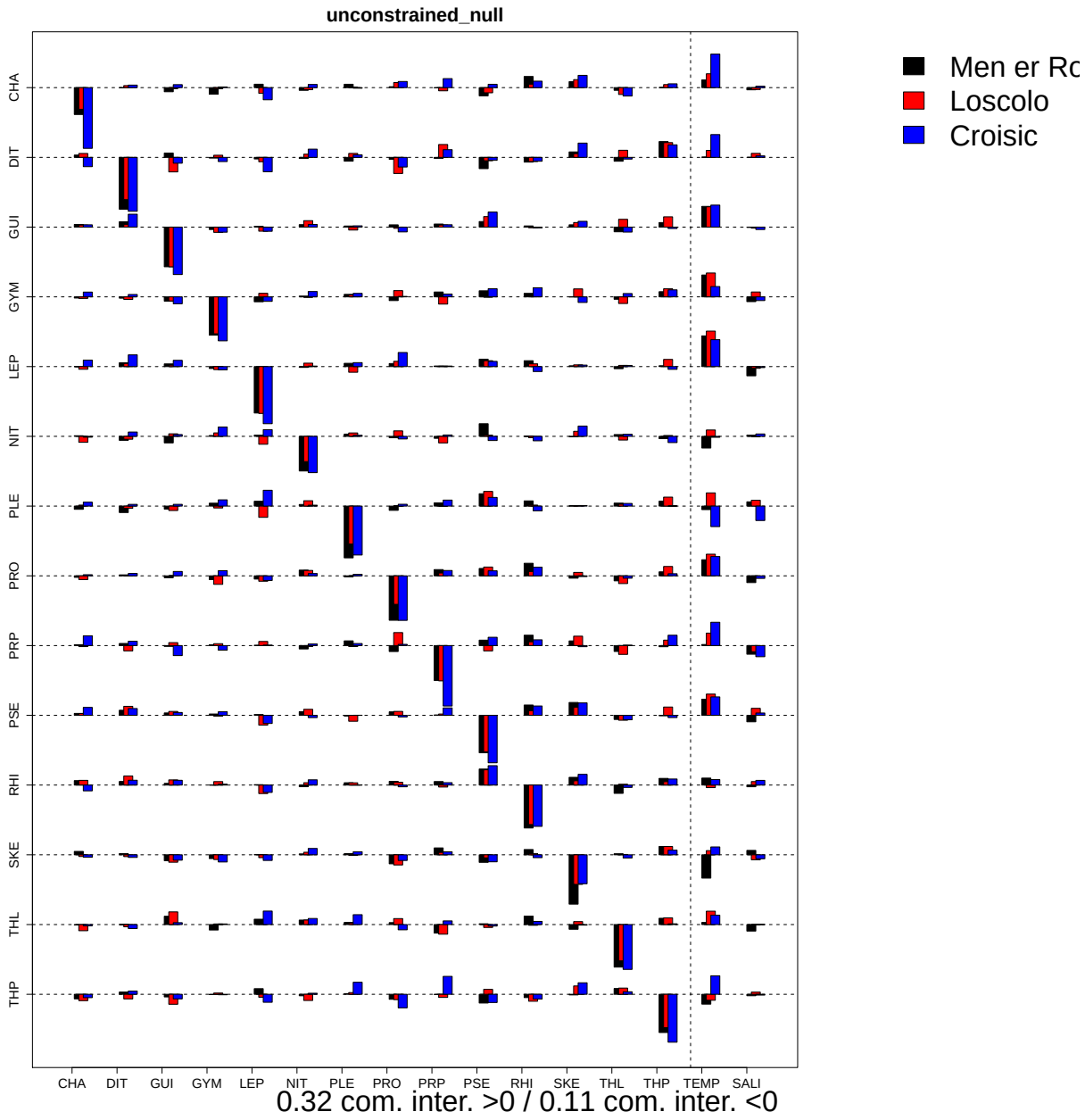


Figure 10: Coefficients of the unconstrained MAR model for species which are present at the three sampling sites in South Brittany, using Temperature and Salinity as covariates. There are 32% interactions positive at all sites and 11% interactions negative at all sites (not considering diagonal values, that is intragroup interactions).

South Brittany

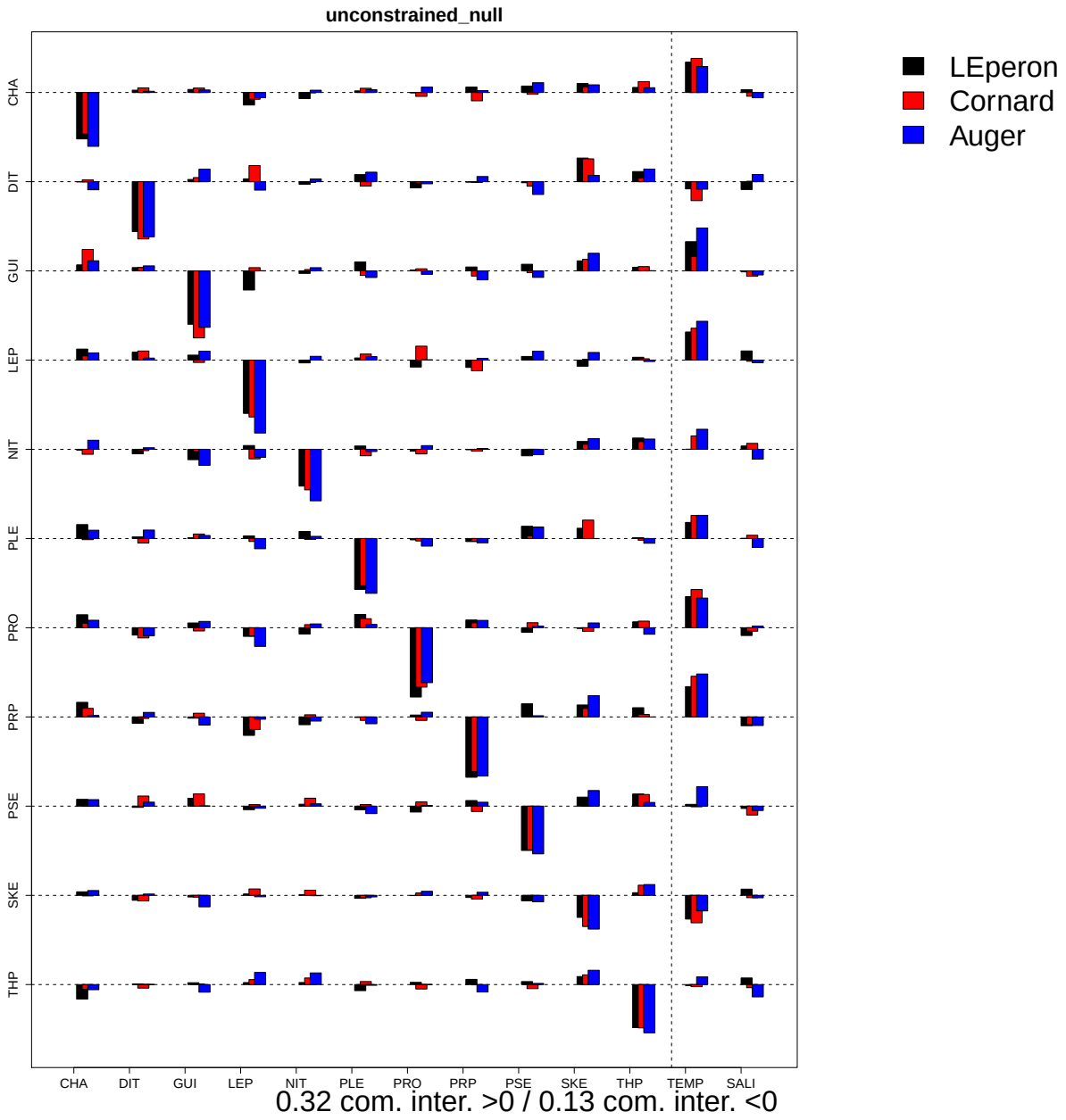


Figure 11: Coefficients of the unconstrained MAR model for species which are present at the three sampling sites in Marennes Oléron, using Temperature and Salinity as covariates. There are 32% interactions positive at all sites and 13% interactions negative at all sites (not considering diagonal values, that is intragroup interactions).

Marennes-Oléron

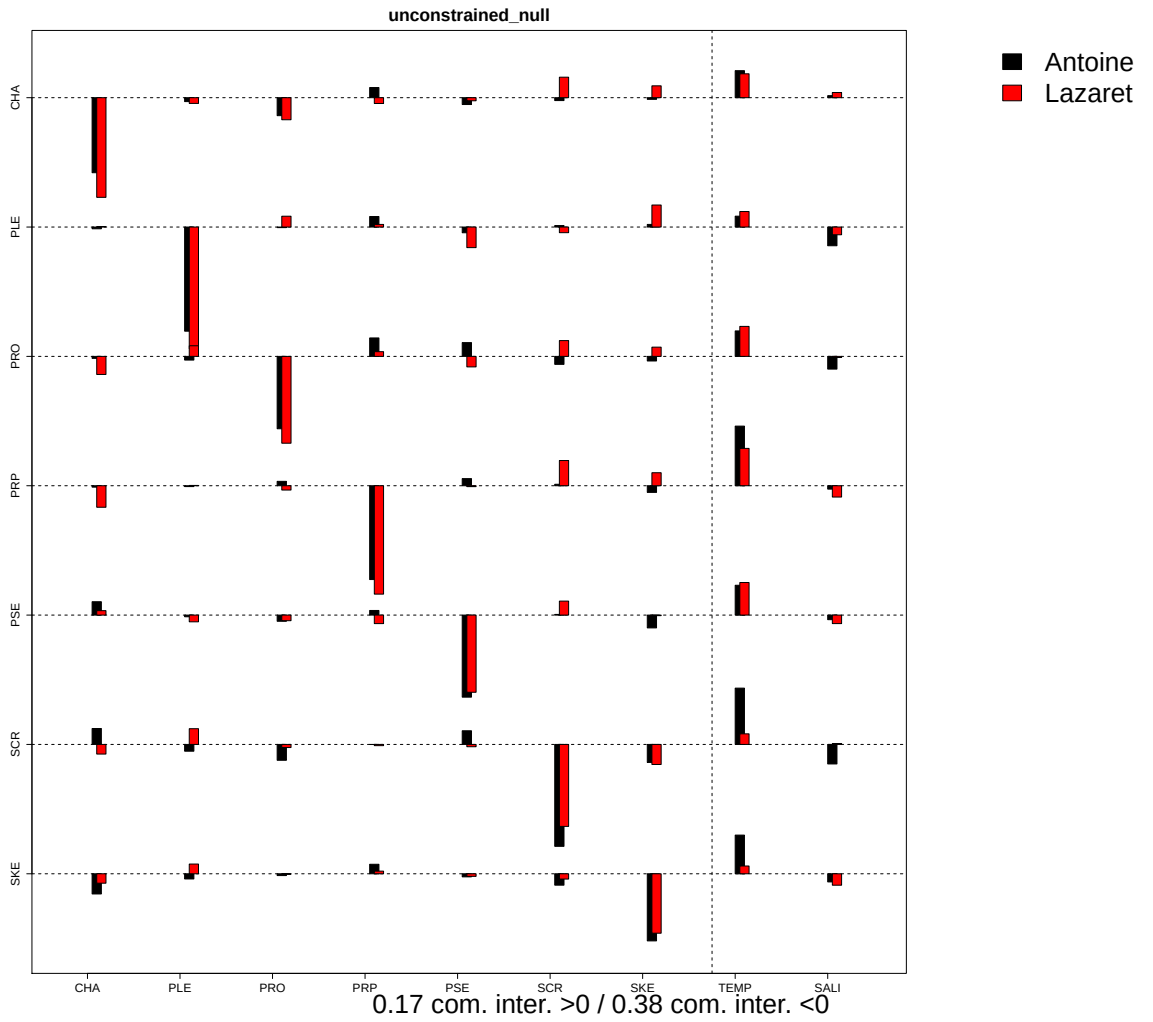


Figure 12: Coefficients of the unconstrained MAR model for species which are present at the two sampling sites in the Mediterranean Sea, using Temperature and Salinity as covariates. There are 17% interactions positive at all sites and 38% interactions negative at all sites (not considering diagonal values, that is intragroup interactions).

Mediterranean Sea

3 What do we do now?

We could focus on toxic species only, which might be monitored more carefully than others, and check empirically the results from model-based papers [1, 2]. We could also use FLORPAR and FLORIND to follow the synchrony of toxic groups on more sites than the ones that are only followed with FLORTOT (see Rapport/graphe/time_series_florparind.pdf: we can confirm this by studying the species which are monitored at each site, and complete with FLORTOT but we have to keep in mind that methods differ: in FLORPAR/FLORIND, zeroes are 'real' because species are searched for).

We can check the absence of competition between main groups (pennate, centric, dinoflagellates), as we observed in Teychan. Not sure we can do it for all sites.

We can also look at the interaction intra-group when they are well-monitored (but in this case, we can only use a subset of the previously defined subset)/ We can check for compensation intra-site.

We can have a look at the spatial synchrony

4 What do the others do?

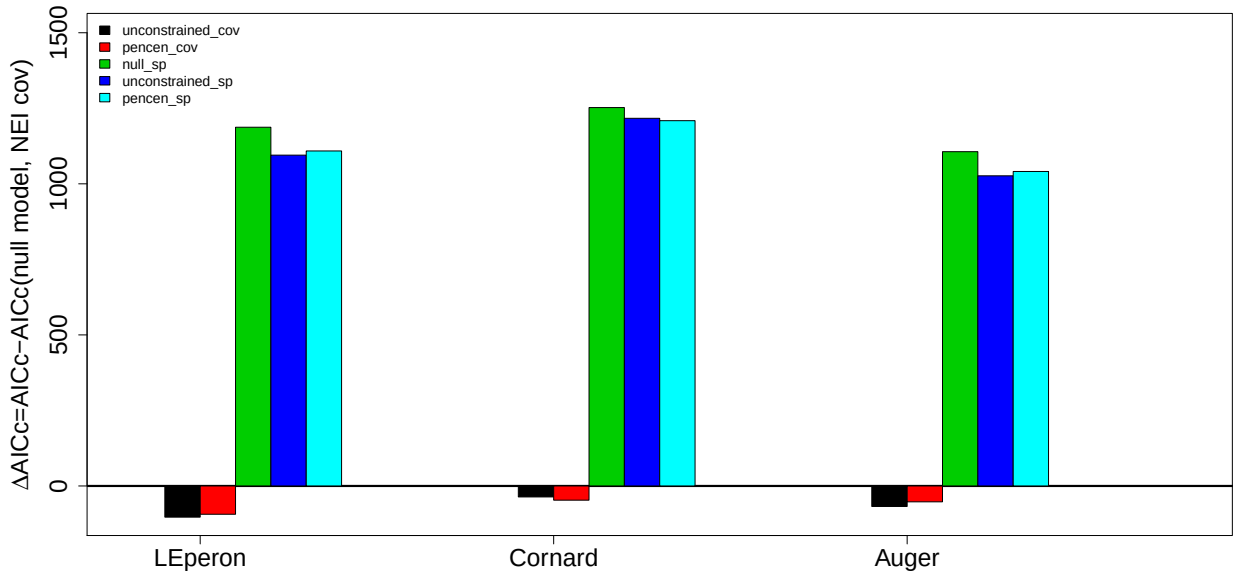
From previous report[18]

There are two ways to define the species which form the majority of the community. We can look at the overall relative abundance (ratio of cumulated abundance of a group over the total phytoplankton abundance through the all time series), or the presence of each group in the samples (number of sampling dates for which the group was not deemed as missing, over the total number of sampling dates). For Table 5, we first ranked each groups according to its abundance or presence at each site and kept only the first 20 groups over 58. Then, we considered all groups and averaged abundance and presence values among all sites. It should be noted that for each computation, there were at least 10 species ranking in the most important species at all sites.

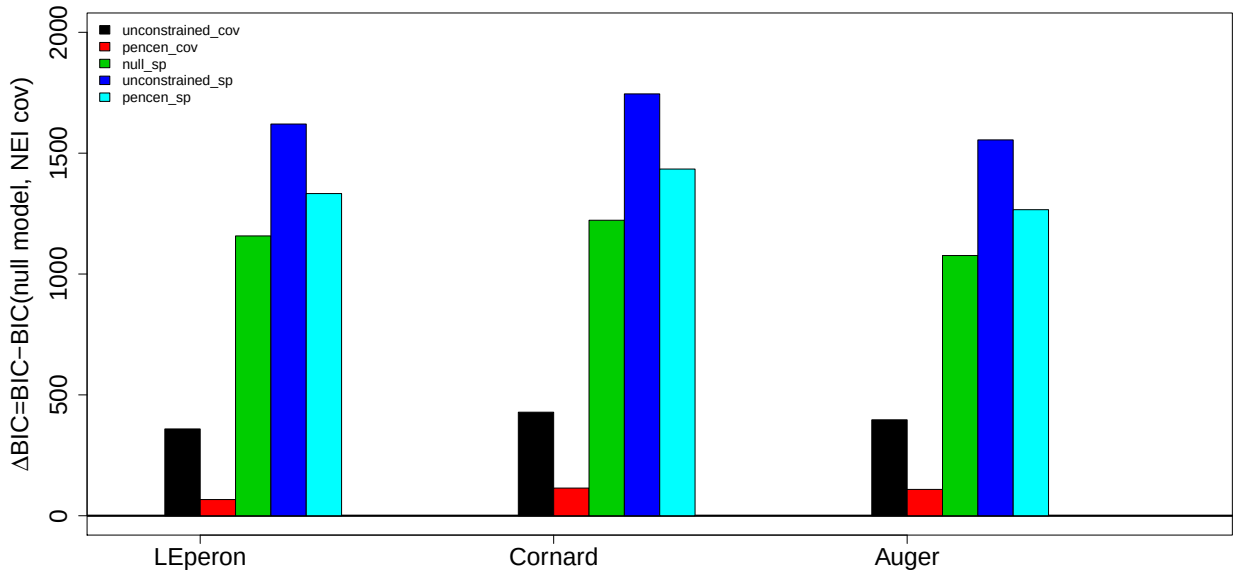
Abundance		*Presence*[19]	
Skeletonema (C)	*0.242*	*Not Elsewhere Identified*	*0.87*
Leptocylindrus (C)	*0.170*	*Nitzschia / Ceratoneis /... (P)*	*0.78*
Chaetoceros (C)	*0.168*	*Gymnodinium (D)*	*0.75*
Not Elsewhere Identified	*0.095*	*Chaetoceros (C)*	*0.72*
Pseudo-Nitzschia (P)	*0.069*	*Naviculaceae (P)*	*0.71*
Gymnodinium (D)	*0.036*	*Protoperidinium (D)*	*0.68*
Cryptophytes	*0.029*	*Prorocentrum (D)*	*0.64*
Thalassiosira.Porosira (C)	*0.021*	*Thalassiosira.Porosira (C)*	*0.63*
Asterionellopsis (P)	*0.020*	*Leptocylindrus (C)*	*0.63*
Nitzschia / Ceratoneis /... (P)	*0.016*	*Pleurosigma / Gyrosigma (P)*	*0.62*
Dactyliosolen (C)	*0.016*	*Skeletonema (C)*	*0.57*
Prorocentrum (D)	*0.015*	*Pseudo-Nitzschia (P)*	*0.56*
Guinardia (C)	*0.012*	*Scrippsiella (D)*	*0.55*
Ceratoneis (P)	*0.012*	*Coscinodiscus (C)*	*0.49*
Rhizosolenia (C)	*0.009*	*Thalassionema / ... (P)*	*0.48*
Thalassionema / ... (P)	*0.007*	*Guinardia (C)*	*0.47*
Scrippsiella / ... (D)	*0.007*	*Euglenophytes*	*0.44*
Naviculaceae (P)	*0.007*	*Ditylum (C)*	*0.41*
Pennate (P)	*0.005*	*Pennate (P)*	*0.41*
Protoperidinium (D)	*0.005*	*Rhizosolenia (C)*	*0.29*

Table 5: Average relative abundance and proportion of presence throughout the time series for the 20 phytoplankton groups which can be found in eight sites along the French coast. Letters in parenthesis correspond to centric (C), pennate (P) diatoms or dinoflagellates (D)

[18] *This first analysis was naive: I used the same classification as for Arcachon Bay, while it is dependent upon the local analysts' knowledge.*



(a) Difference of AICc for three sites and common groups of species



(b) Difference of BIC for three sites and common groups of species

Figure 13: Comparison of AICc and BIC from different models: NEI is either used as a covariate (cov) or a group of species (sp) like CHA; and different interaction matrices are compared. This figure shows the difference between the AICc/BIC of a given model and the corresponding criterium of the null (diagonal interaction matrix) model with NEI as a covariate. A negative value indicates that the model performs better than the null model with NEI=cov. A positive value indicates that the null model is 'better'.

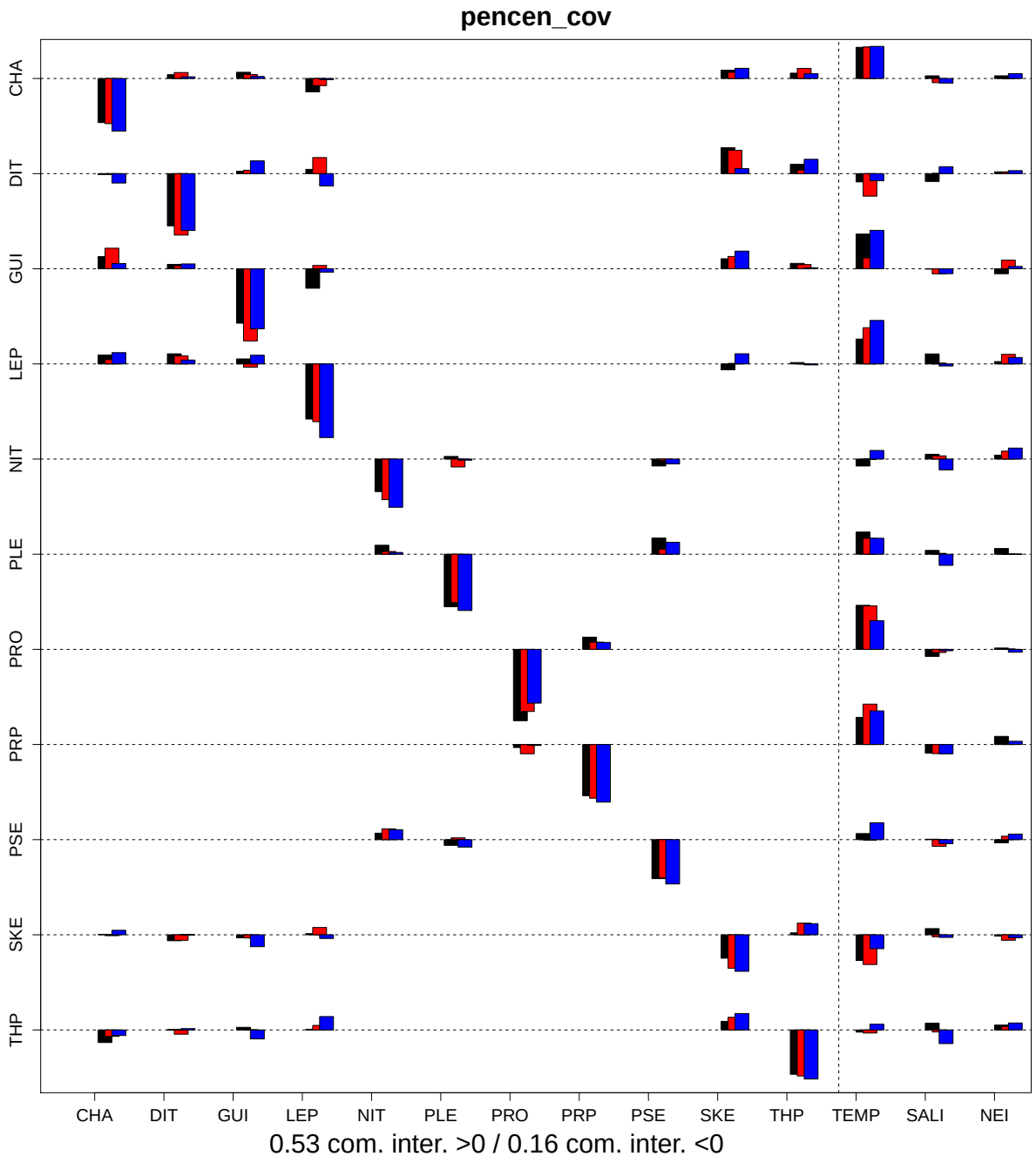


Figure 14: Coefficients of the pennate-centric MAR model for the three time series, using Temperature, Salinity and NEI as a covariate, for common species only. Black bars are for LEperon, red is for Cornard and blue bare are for Auger. There are 53% interactions that are positive at all sites; 16% are negative at all sites.

References

[1] Patricio Díaz, Beatriz Reguera, Manuel Ruiz-Villarreal, Yolanda Pazos, Lourdes Velo-Suárez, Henrick Berger, and Marc Sourisseau. Climate Variability and Oceanographic Settings Associated with Interannual Variability in the Initiation of Dinophysis acuminata Blooms. *Marine Drugs*, 11(8):2964–2981, August 2013.

[2] Christopher J. Gobler, Owen M. Doherty, Theresa K. Hattenrath-Lehmann, Andrew W. Griffith, Yoonja Kang, and R. Wayne Litaker. Ocean warming since 1982 has expanded the niche of toxic algal blooms in the North

Atlantic and North Pacific oceans. *Proceedings of the National Academy of Sciences*, 114(19):4975–4980, May 2017.

[3] Stephanie E. Hampton and Daniel E. Schindler. Empirical evaluation of observation scale effects in community time series. *Oikos*, 113(3):424–439, 2006.

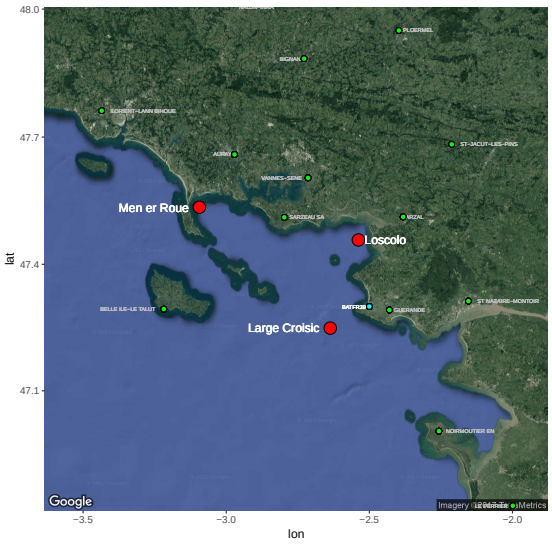
[4] T. Hernández Fariñas, C. Bacher, D. Soudant, C. Belin, and L. Barillé. Assessing phytoplankton realized niches using a French national phytoplankton monitoring network. *Estuarine, Coastal and Shelf Science*, 159:15–27, June 2015.

[5] Tania Hernández Fariñas. *Analyse et modélisation des évolutions à long terme de la biodiversité phytoplanctonique dans les zones côtières sous l'effet des pressions environnementales et anthropiques*. PhD thesis, Université de Nantes, Faculté des Sciences et des Techniques, 2015.

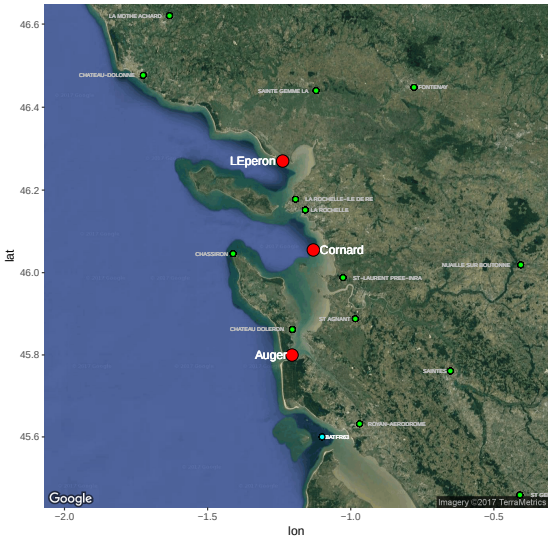
[6] Isabel Jalón-Rojas, Sabine Schmidt, and Aldo Sottolichio. Evaluation of spectral methods for high-frequency multiannual time series in coastal transitional waters: advantages of combined analyses: Tests of spectral analysis on coastal time series. *Limnology and Oceanography: Methods*, 14(6):381–396, June 2016.

[7] Dmitri Kondrashov and Michael Ghil. Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Processes in Geophysics*, 13(2):151–159, 2006.

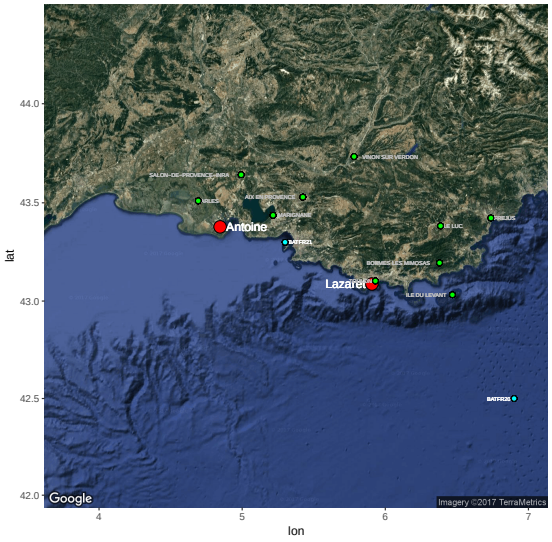
[8] Péter Török, Enikñ T-Krasznai, Viktória B-Béres, István Bácsi, Gábor Borics, and Bála Tóthmérász. Functional diversity supports the biomass-diversity humped-back relationship in phytoplankton assemblages. *Functional Ecology*, 30(9):1593–1602, September 2016.



(a) Location of sampling sites in South Brittany



(b) Location of sampling sites close to Marennes-Oléron



(c) Location of sampling sites on the Mediterranean Sea. Note that Antoine is located in the "Golfe de Fos", which is known to be quite polluted

Figure 4: Location of sampling sites (red dots) with interesting time series. Green points indicate the terrestrial observation site for Météo France while blue points correspond to observation buoy.