

REPHY time series from other sites

July 17, 2017

1 Choices

1.1 Filtering in Quadriga

We queried the Quadriga 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?)** [4]. 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 (Mismar) 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.

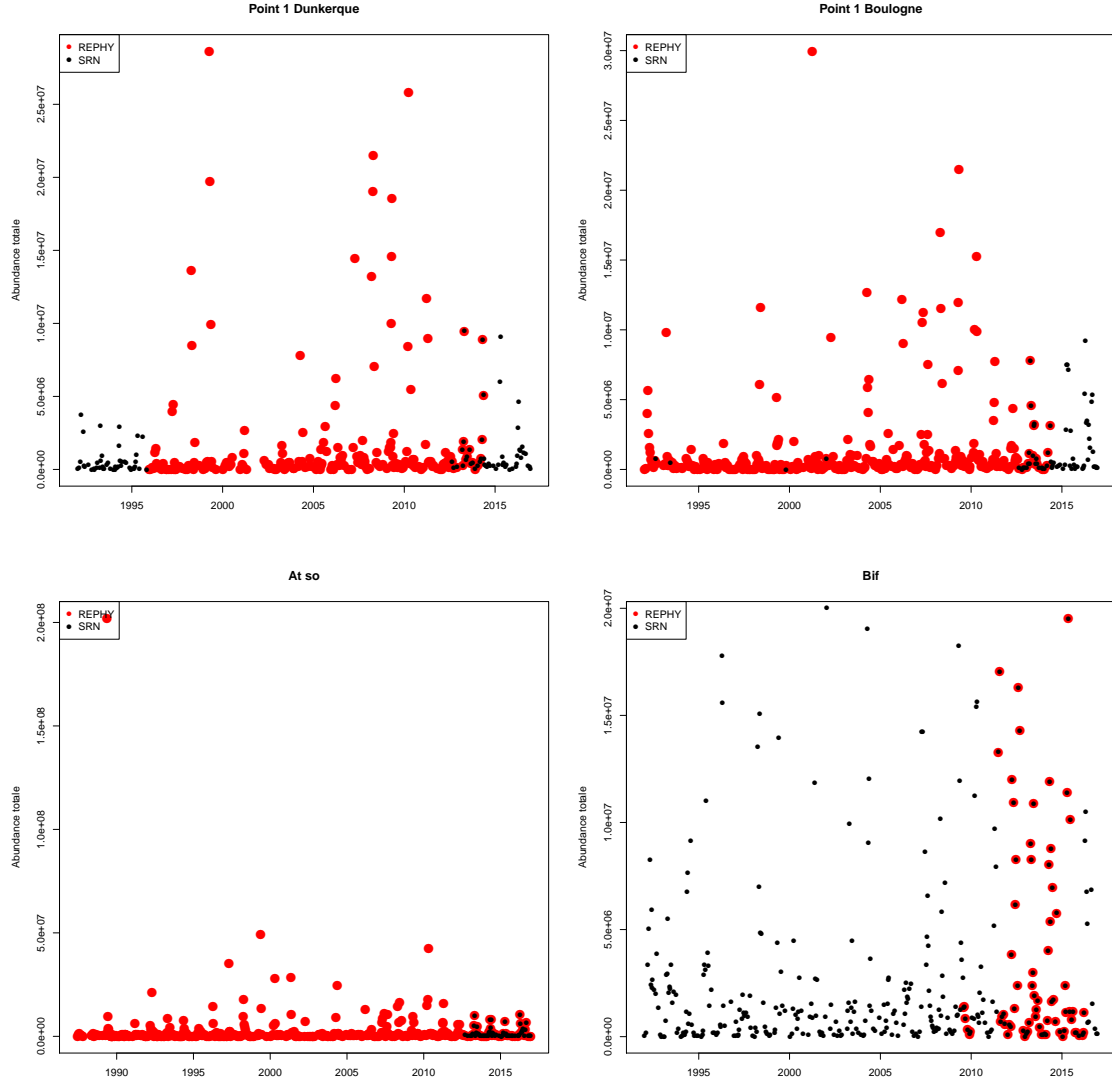


Figure 1: Comparison of total phytoplankton abundance for each site monitored by both REPHY and SRN.

In addition to the sites presented in (author?) [4], 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.

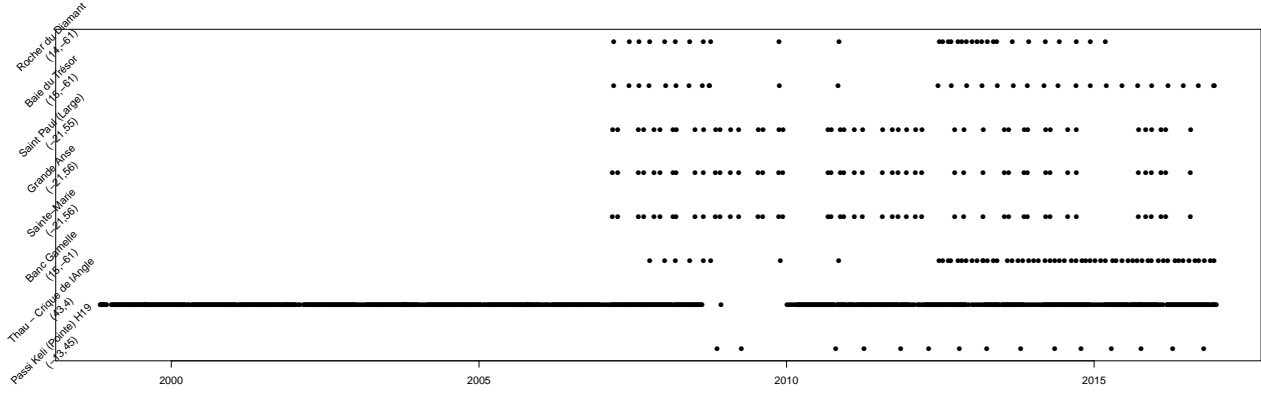
Several observation networks could complement the REPHY network. We also queried data from:

- 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, announced as a 'biological variable' monitoring, does not follow phytoplankton abundance
- AAMP: phytoplanktonic surveillance, leads to a void extraction

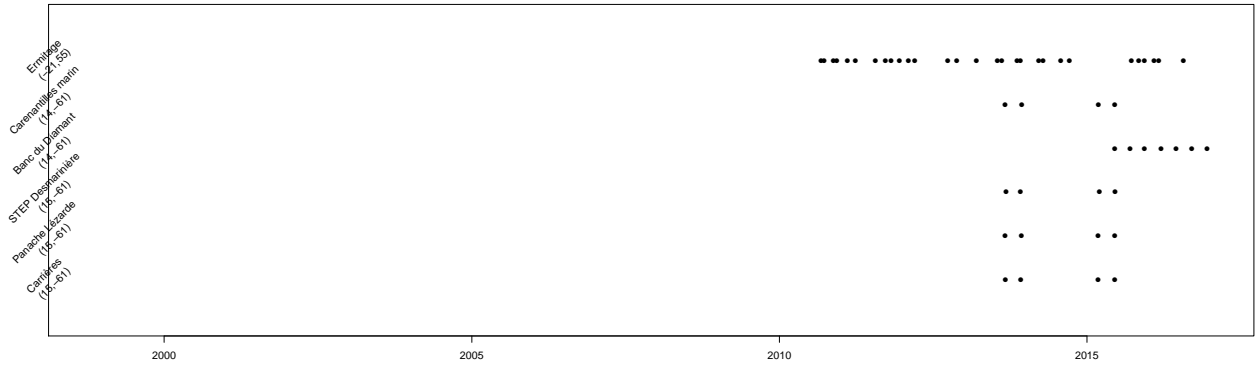
We only used data which were either qualified as 'good' or 'unqualified'. For certain sites and certain periods of times, plankton samples were collected from both the top and bottom of the water column. We only kept values from the top of the water column¹.

¹We 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

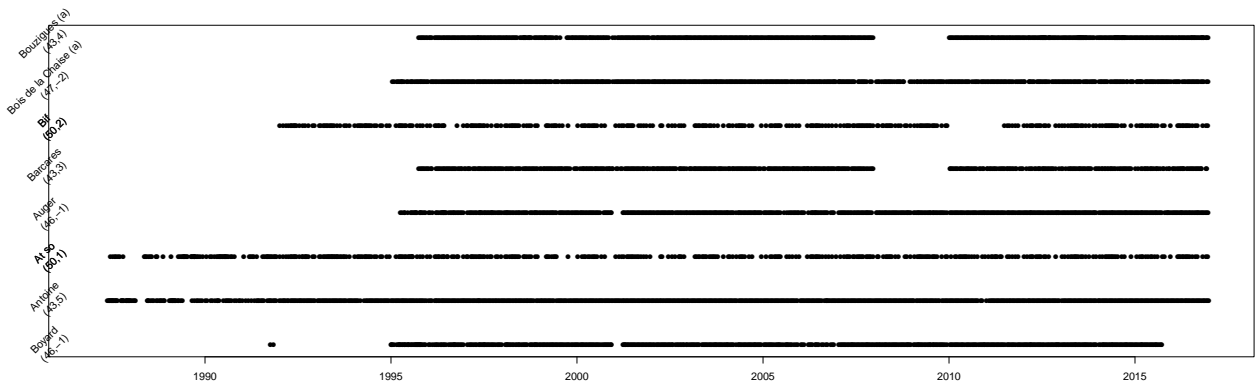
1.2 Filtering “best” time series to work with



(a) Overseas sites + Thau



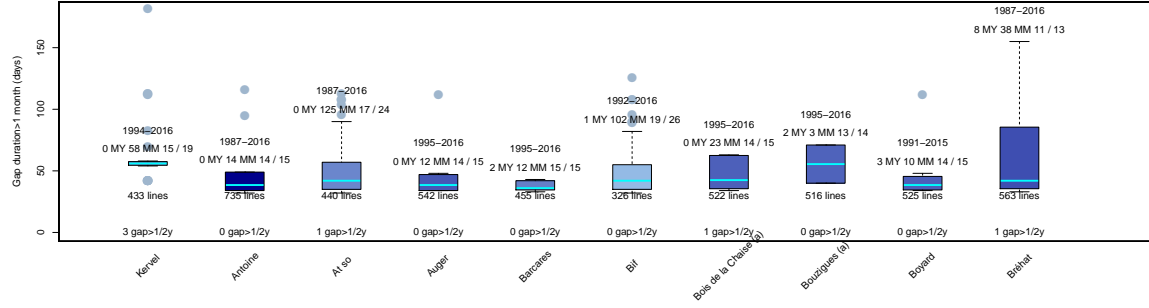
(b) Overseas sites + Thau



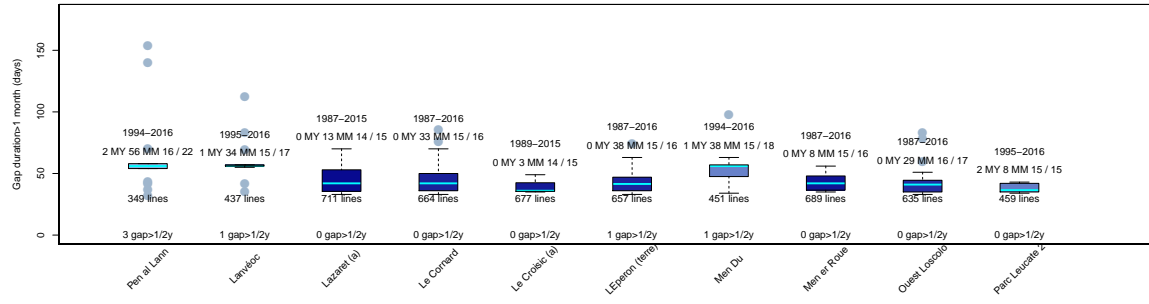
(c) REPHY + SRN (from Tania)

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

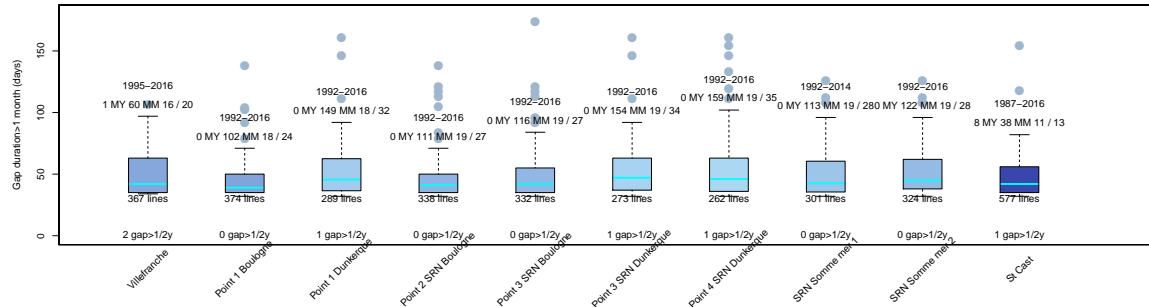
both bottom and top...



(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 in the text below the boxplots. The text at the top indicates the starting and ending years (or the last available year on the Quadrigé 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 time 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. 3³. Finally, the following sites were chosen: Antoine, Lazaret (Mediterranean sea), Auger, L'Eperon, Le Cornard (Pertuis Charentais), Bois de la Chaise, Le Croisic, Men er Roue, Loscolo (South Brittany). They are located on Fig. 4.⁴

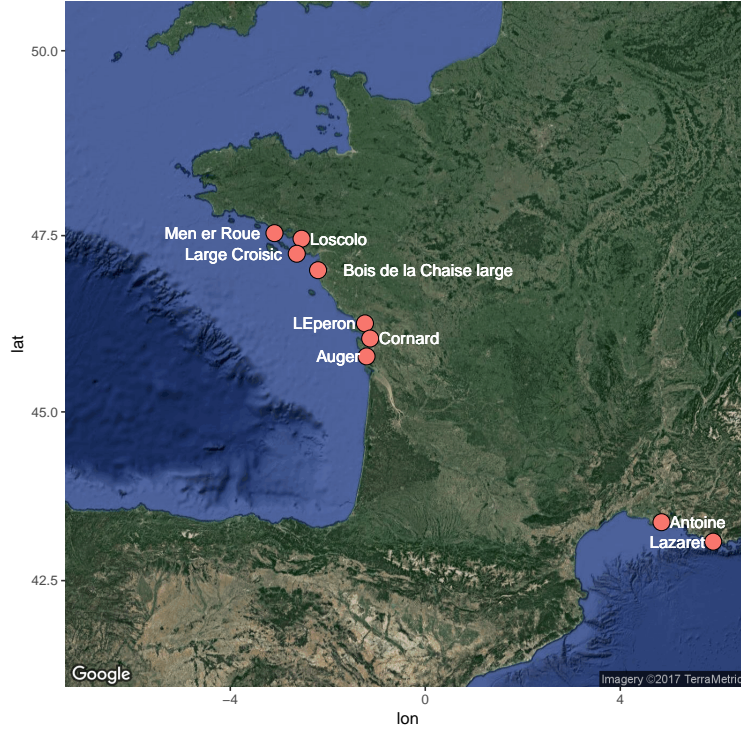


Figure 4: Location of sampling sites with 'seemingly' interesting time series

³Not 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

⁴We 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 site⁵, 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 that 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 analysis⁶. We used WORMS database⁷ 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 database⁸.

| | 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.

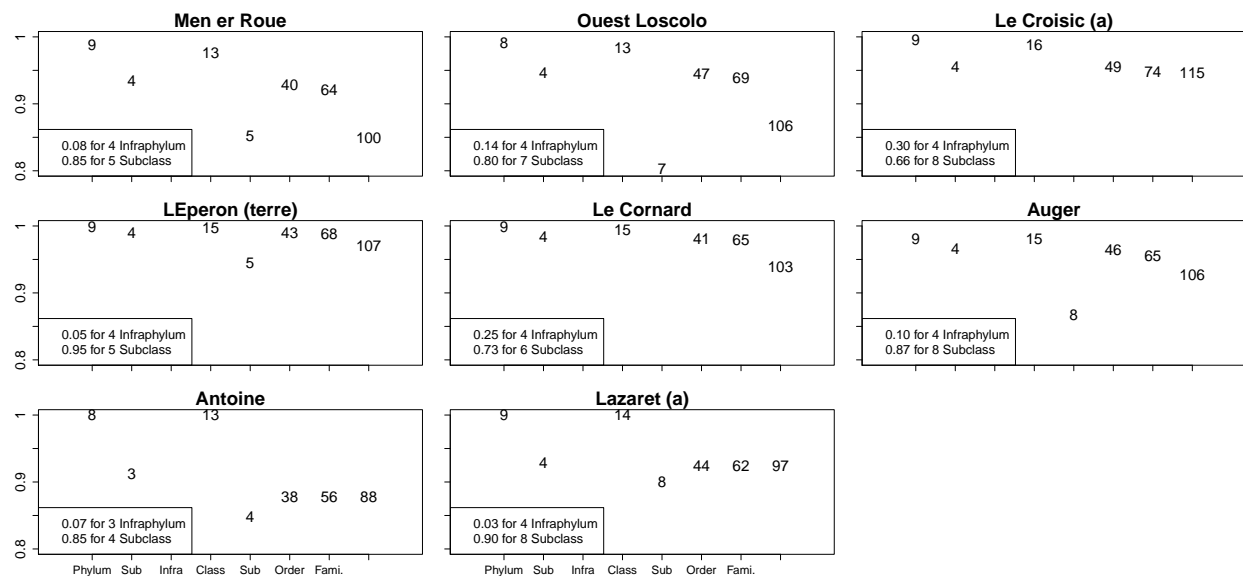
We can see in Fig. 5 that we can identify more than 85% of the observed cells at the genus level at all sites.

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

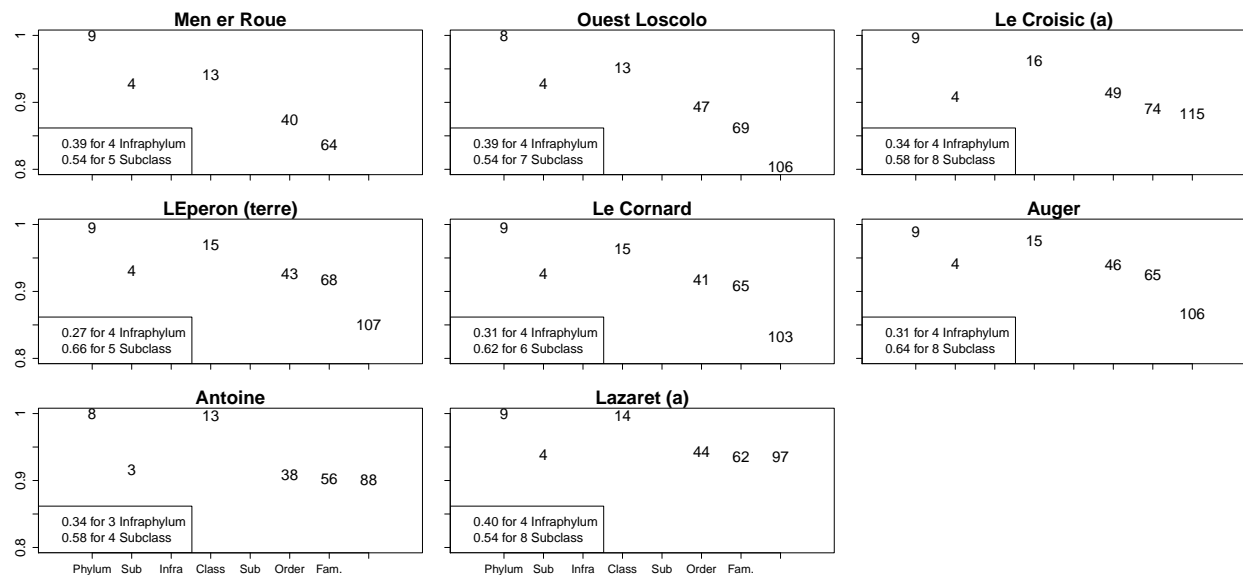
⁶We could also try functional groups... if we had such information

⁷<http://www.marinespecies.org/>

⁸This 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.

⁹I 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?) [3] 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 kept¹⁰.

| 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+Archaeoperidinium+Peridinium |
| SCR | Scrippsiella+Enciculifera+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. **We can discuss it together with the names of the groups.**

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

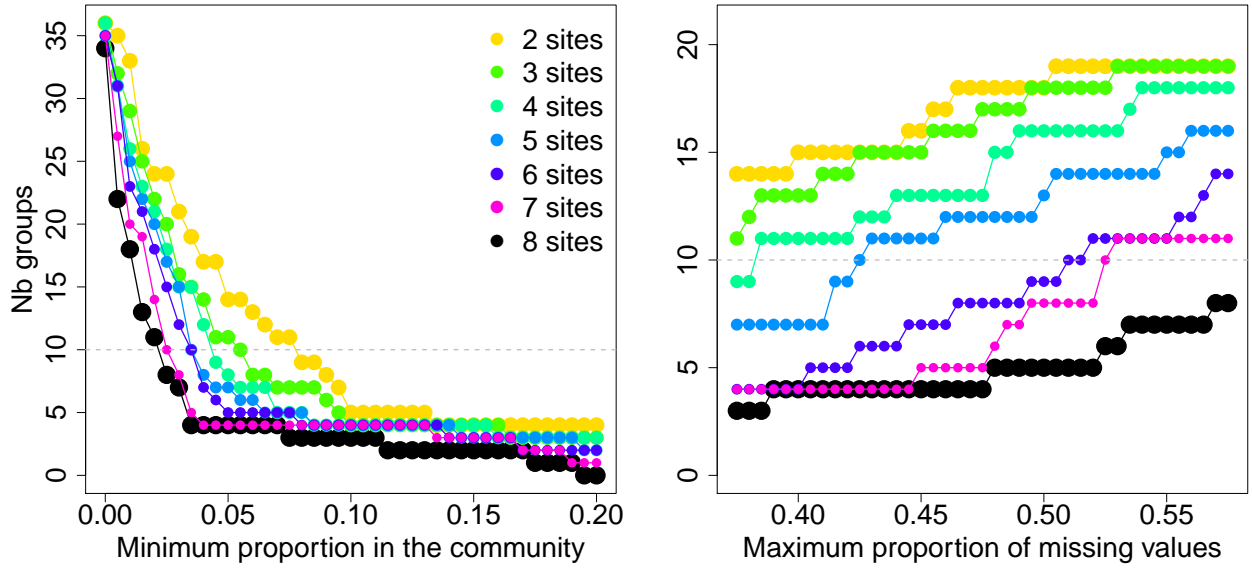


Figure 6: Number of planktonic groups whose average proportion in the community is above a threshold given on the x-axis (left) or whose maximum proportion of missing values is below a threshold on the x-axis (right) for 2 to 8 out of the 8 sites we consider. For example, black dots show that a minimum 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 organisms¹¹. 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' [5], and to the difference we found in the first paper. Times series can be seen in graphe/Subclass **Discuss this**.

¹¹See time series in the graphe/Subclass folder

2 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 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

From previous report¹²

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 4, 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 ¹³ | |
|---|-------|---|------|
| <i>Skeletonema</i> (C) | 0.242 | Not Elsewhere Identified | 0.87 |
| <i>Leptocylindrus</i> (C) | 0.170 | <i>Nitzschia</i> / <i>Ceratoneis</i> /... (P) | 0.78 |
| <i>Chaetoceros</i> (C) | 0.168 | <i>Gymnodinium</i> (D) | 0.75 |
| Not Elsewhere Identified | 0.095 | <i>Chaetoceros</i> (C) | 0.72 |
| <i>Pseudo-Nitzschia</i> (P) | 0.069 | <i>Naviculaceae</i> (P) | 0.71 |
| <i>Gymnodinium</i> (D) | 0.036 | <i>Protoperidinium</i> (D) | 0.68 |
| <i>Cryptophytes</i> | 0.029 | <i>Prorocentrum</i> (D) | 0.64 |
| <i>Thalassiosira.Porosira</i> (C) | 0.021 | <i>Thalassiosira.Porosira</i> (C) | 0.63 |
| <i>Asterionellopsis</i> (P) | 0.020 | <i>Leptocylindrus</i> (C) | 0.63 |
| <i>Nitzschia</i> / <i>Ceratoneis</i> /... (P) | 0.016 | <i>Pleurosigma</i> / <i>Gyrosigma</i> (P) | 0.62 |
| <i>Dactyliosolen</i> (C) | 0.016 | <i>Skeletonema</i> (C) | 0.57 |
| <i>Prorocentrum</i> (D) | 0.015 | <i>Pseudo-Nitzschia</i> (P) | 0.56 |
| <i>Guinardia</i> (C) | 0.012 | <i>Scrippsiella</i> (D) | 0.55 |
| <i>Ceratoneis</i> (P) | 0.012 | <i>Coscinodiscus</i> (C) | 0.49 |
| <i>Rhizosolenia</i> (C) | 0.009 | <i>Thalassionema</i> / ... (P) | 0.48 |
| <i>Thalassionema</i> / ... (P) | 0.007 | <i>Guinardia</i> (C) | 0.47 |
| <i>Scrippsiella</i> / ... (D) | 0.007 | <i>Euglenophytes</i> | 0.44 |
| <i>Naviculaceae</i> (P) | 0.007 | <i>Ditylum</i> (C) | 0.41 |
| <i>Pennate</i> (P) | 0.005 | <i>Pennate</i> (P) | 0.41 |
| <i>Protoperidinium</i> (D) | 0.005 | <i>Rhizosolenia</i> (C) | 0.29 |

Table 4: 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)

¹²This first analysis was naive: I used the same classification as for Arcachon Bay, while it is dependent upon the local analysts' knowledge.

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] 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.
- [4] 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.
- [5] Péter Török, Enikő 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.