Rapport de stage de 2ème année

Master Spécialité Sciences de la Mer

Parcours Ecologie Marine et Halieutique

Biogeography of the historical Brittany rocky shore macrofauna, using citizen science

LERDOU Lucie

Grall Jacques - Institut Universitaire Européen de la Mer

Boyé Aurélien, Chevalier Mathieu - Institut Français pour la Recherche et l'Exploitation de la Mer

ANR-21-EXES-0011

Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du programme d'Investissements d'avenir portant la référence "ANR-21-EXES-00 11" dans le cadre de l'école graduée IFSEA

Acknowledgment

I would like to begin by expressing my sincere gratitude to my internship supervisor, Jacques Grall, research engineer at the IUEM (Institut Universitaire Européen de la Mer), and my internship tutors, Aurélien Boyé and Matthieu Chevalier, researchers at the IFREMER (Institut Français pour la Recherche et l'Exploitation de la Mer) for their support and kindness throughout my internship. I have learnt a lot under your mentorship, both in terms of ecological and statistical expertise, but also in terms of effective report writing and am deeply grateful for the knowledge and skills I have gained during this time.

Additionally, I extend my gratitude to the Bretagne Vivante association, especially Christian Hily for his support with the OBCE data and taxonomic expertise, Quentin Rochas and Pierre Le Dorze for their technical assistance, and Bernard Trebern for introducing me to OBCE's fieldwork.

Finally, I am grateful to the benthos team at IUEM and the Lebco-Dyneco team at IFREMER for their warm welcome, and particularly to Marion, Camille, Vincent, Lucas, and Guillaume for the thrilling board game breaks and Robin for the writing support.

Summary

Introduction

Biogeography was defined by Myers & Giller (1988) as the study of the geographical distribution of organisms. It aims to identify past, present and future distributions of species and their underlying drivers. This field has gained interest to monitor changes in species distributions due to climate change with a large number of studies showing that rising temperatures have altered species distribution patterns (Barry et al., 1995; Walther et al., 2002; Harley et al., 2006), with species often migrating polewards (Lathlean et al., 2015). Hence, studying biogeographic patterns can provide insights into the history and the drivers of distributional change. This is of utmost importance given that biodiversity is related to ecosystem health and that changes in biodiversity (e.g. species abundance decline or changes in species composition) can disturb critical ecological functions (Lotze et al., 2006; Hooper et al., 2012; Harley et al., 2006; Walther et al., 2010) and the services provided to human societies (Malone et al., 2013). For instance, studies have shown that biomass production increases with species richness in many environments, ensuring livelihoods (e.g. food, income) for many humans worldwide (Duffy et al., 2017).

Coastal ecosystems, known for their important taxa richness and productivity (Watanabe et al., 2018), are critical zones at the interface between marine and terrestrial ecosystems providing diverse habitats and heterogeneous living conditions (Spalding et al., 2007; Halpern et al., 2008). However, these habitats are facing important anthropogenic pressures due to cumulative impacts of climate change and other issues including overexploitation, introduction of invasive species or pollution (Halpern et al., 2019; Bowler et al., 2020) that can drive changes in species distribution and community composition (Walther et al., 2010). Such changes have been observed in several coastal benthic organisms such as crustaceans, bivalves, gastropods, and polychaetes (Pezy & Dauvin, 2016; Philippart et al., 2003; Mieszkowska et al., 2007; Wethey & Woodin, 2008).

Species Distribution Models (SDMs), based on the concept of the ecological niche, are often used to understand and predict changes in biodiversity. More specifically, SDMs are representing the species' realized niche, defined by Hutchinson (1957) as the range of environmental characteristics within which a species can maintain positive growth in the presence of other species and constrained by dispersal limitations (e.g. geographic barriers). In practice, SDMs are frequently used to predict the presence probability of species based on spatial predictors that characterize the conditions experienced by species across their distribution and which include climatic conditions (e.g. temperature, humidity, wind…) but also anthropogenic pressures (e.g. chemical load, fishing activities; Guisan et al., 2005; Elith et al., 2009; Guillera-Arroita et al., 2015).

One of the applications of SDMs consists in aggregating the predictions obtained for individual species to look at changes at higher levels of biological organization (e.g. community level). Several techniques, including group modeling and individual species modeling can be used to improve our understanding of the relationships between communities and their environment (Dubuis et al., 2011; Guisan & Rahbek, 2011; Calabrese et al., 2014). The two most frequently used methods to model communities are the “assemble first, predict later” and “predict first, assemble later” (D’Amen et al., 2017). In the “assemble first, predict later” approach, the data are first grouped using biological criteria or ordination analysis resulting in community-level groups that represent the object of inference. In the second method, “predict first, assemble later” the species are first modeled to produce a map of their repartition that is then aggregated and submitted to a classification to obtain community-level groups (D’Amen et al., 2017; Ferrier et al., 2006). These two methods are based on the assumption that community composition is stable in space and time (i.e. only the geography of the community changes) and view the communities as a set of co-occurring species (Ferrier et al., 2006; D’Amen et al., 2017).

Modeling species distributions requires spatial data on the distribution of species that can be collected using various approaches (Bled et al., 2013; Stockwell et al., 2002). Standardized fieldwork data collected by experts is the gold-standard approach but is expensive and time consuming (Peterson et al., 2008; Feldman et al., 2021). Another approach consists in using data from citizen science programmes which are sometimes designed following a standardized protocol where trained volunteers are collecting the data (Egar et al., 2016). Thanks to citizen science programs, we can acquire large amounts of data with a wide spatial and temporal coverage (Greenwood, 2007). Despite being generated using time-saving techniques and by volunteers without professional training, these databases often produce results comparable to those obtained from databases collected by specialists (Devictor et al., 2010). Indeed, contrary to the misconception that volunteer-based efforts produce imprecise outcomes, they actually maximize sample sizes, thereby increasing statistical power and robustness (Newman et al., 2003; Schmeller et al., 2009).

In historical Brittany (comprising the region of Brittany and the Loire-Atlantique department), the OBCE (Observatoire Breton des Changements sur l’Estran) established by the association Bretagne Vivante, is a citizen science project that started in 2017 with the aim of collecting occurrence data for intertidal benthic macrofaunal species. This program is particularly interesting as Brittany is considered a biodiversity hotspot for various species groups, including algae, plankton, and benthic macrofauna (VandenHoek & Donze, 1967; Kerswell, 2006; Santelices et al., 2009; Gallon et al., 2017; Kléparski et al., 2021; Fragkopoulou et al., 2022). This exceptional diversity can be explained by the location of Brittany between two biogeographic zones (the English Channel and the Bay of Biscay), that provides a high level of habitat heterogeneity (Tempera et al., 2019) favoring the establishment of species with various ecological characteristics. The geographical context of historical Brittany, characterized by the near absence of stratification in the north and increased stratification in the south, where tides and currents are less important (Spalding et al., 2007; Penard, 2009; Gallon et al., 2017), contributes to the unique environmental characteristics of the region (Spalding et al., 2007; Penard, 2009; Gallon et al., 2017).

To date, most large-scale research on benthic macrofauna were conducted across the north Atlantic ocean (e.g. Dinter et al., 2001; Chust et al., 2024) and often focused on a limited number of species or specific taxonomic groups (Melo-Merino et al., 2020). For instance, while it has been shown, in the United Kingdom, that the distribution of rocky intertidal species is changing rapidly as a result of climate change (Hawkins et al., 2009; Burrows et al., 2019; Mieszkowska et al., 2021), most studies focused on soft subtidal species (Gaudin et al., 2018; Hiddink et al., 2014; Hinz et al., 2011), with low attention to rocky shore communities. Therefore, we still have a poor understanding of the current distribution and drivers of spatial change in the biodiversity of benthic macrofauna in intertidal rocky shore habitats. The environmental differences between North and South and East and West Brittany are well known (Tempera et al., 2019), as well as the existence of the Ushant front, separating mixed coastal waters in the Iroise sea and the Manche from thermally stratified open Celtic Sea waters (Le Boyer et al., 2009; Gallon et al., 2014). However, our understanding of how and to which extent the biogeographical barrier represented by the Ushant front affects the distribution of shore macrofaunal species remains limited (Dinter et al., 2001). Using physico-chemical variables, Tempera et al. (2019) presented a high-resolution partitioning of Brittany's infralittoral seabed. They identified eight clusters characterizing the environmental conditions prevailing in bays, estuarine-influenced zones, coastal areas exposed to waves, and sectors influenced by currents. Although of importance, this study is only based on abiotic variables providing limited insights on the resulting biodiversity patterns at the scale of Brittany.

The aim of this study is to describe and spatially predict the biogeography of rocky shore macrofaunal communities along historical Brittany's coastlines, using SDMs. We further aim to identify the main environmental drivers influencing spatial variations in macrofaunal community composition. This will allow us to evaluate the OBCE program and its potential to be used in biogeography studies. At the scale of Brittanny, we expect to identify a latitudinal and longitudinal gradient owing to the presence of known biogeographic barriers (Le Boyer et al., 2009; Gallon et al., 2014), but we also expect more local patterns with communities specific to areas like bays and estuaries due to prominent influence of some specific environmental conditions such as exposition to waves and currents, river plumes, temperature or ocean stratification (as seen in Tempera et al., 2019 and Derrien-Courtel et al., 2013).

Materials and methods

1. Species occurrence data

Species occurrence data come from the OBCE, a citizen science project supported by the association Bretagne Vivante (https://www.estran-bretagne-vivante.org/). It covers approximately 800 km² of intertidal area exposed during high tide, and engages in a participatory project involving a network of over a hundred active observers, including former or still working professionals, capable of identifying intertidal species. This observatory also collaborates with taxonomic experts from various institutions such as the IUEM (Institut Universitaire Européen de la Mer), the Ifremer (Institut Français de Recherche pour l'exploitation de la Mer) and the CNRS (Centre National de la Recherche Scientifique). The data collection process is non-invasive, relying solely on observation techniques without specimen sampling. The approach, focusing exclusively on occurrence sampling, relies on a straightforward protocol (i.e. an inventory of all species present in the studied area) easily followed by volunteers.

The occurrence data were extracted from the Serena database and were collected between 2017 and 2023 in five French departments: Côtes d'Armor (22), Finistère (29), Ille-et-Vilaine (35), Loire-Atlantique (44) and Morbihan (56, a cartography of the inventories location is available in appendix 1). Species that are not characteristic of the rocky shore macrofauna such as plants, pelagic or benthic species characteristic of soft sediments were excluded based on expert knowledge (J. Grall). Species that are difficult to identify (e.g. small or look-alike species) or that are unevenly identified at the scale of Brittany (e.g. species only identified by some experts but not others) were also removed. Remaining species names were harmonized using WORMS (World Register of Marine Species). The final database contains 29,798 observations of 432 species, from 8 phyla (Annelida, Bryozoa, Chordata, Cnidaria, Crustacea, Echinoderma, Mollusca, Porifera) for a total of 649 inventories (i.e. species found in a site at a given time) with an average species richness of 54.6 species per inventory (statistics on species richness and number of inventories per year are available in appendix 2). Overall, the database contains information on the species identity, the taxonomic classification (order, family, phyla), the date of the observation, the geographical coordinates of the inventory, the sampling year and the name of the group observer. For further analysis, occurrence data were transformed to presence/absence by assuming that when a species was not present in an inventory, it was truly absent (i.e. we assumed no observation errors).

2. Environmental data

We used marine and terrestrial environmental data from different models, with spatial resolutions ranging from 1km² to 4000km² (details on environmental variables, comprising spatial and temporal resolution, the variables extracted, the sources and the models from which the data was extracted, are available in Appendix 3). Terrestrial data (cloud cover, humidity, air temperature, pressure and wind speed) were extracted from the Météo France climate model Arpège (Courtier et al., 1994). Seabed water characteristics (temperature, salinity, suspended particulate inorganic matter [Spim] and dissolved concentrations of ammonium, nitrate, phosphate, silicate and oxygen) were retrieved from the Ifremer model Ecomars (Ardhuin, 2013). For the physical variables, data on current velocity, tidal magnitude and amplitude were extracted from an assimilation of TOPEX/Poseidon and Jason altimetry data (Egbert et al. 2010). Wave height was extracted from the Coordinated Ocean Wave Climate Project (COWCLIP2.0, Morim et al., 2019). The fetch (i.e. a proxy of coastal exposure that indicates the distance over which the swell can propagate without encountering any obstacle) was extracted from Burrows (2020) and summed for every direction. Finally, the coastal type (e.g. Estuary, Erodible rock, Sand beach …) was extracted from the EMODnet Geology portal (Salman et al., 2004).

The Ifremer and météo france data were extracted using the Ifremer supercomputer Datarmor. Overall, we had 20 variables.

This environmental information comes in the form of a grid, with a given spatial resolution. In order to make the correspondence between environmental data and biotic data, we extracted environmental values using inventory coordinates. When extracting this data, we had to overcome the edge effect caused by the location of the inventories at the interface between terrestrial and marine environments since some environmental data such as Spim or salinity are only available in marine environments, and the information is often lacking in intertidal zones (Ge et al., 2005). To address this issue, two methods were tested: the nearest neighbor method (i.e. associating the inventory with the closest data) with a maximum distance of 4km, and the buffer method (i.e. associating the inventory to the average value of data within a buffer zone) with a 4km buffer. These two techniques were compared using a Wilcoxon test (Appendix 4), a non-parametric alternative to the t-test to compare two independent groups of samples with non-normal distribution. We found no significant difference between these two methods. Ultimately, environmental data was extracted using the nearest neighbor method. When temporal information was available, we computed the mean, the standard deviation, the maximum and the minimum of each variable over the sampling period (i.e. 2017 to 2023), bringing the number of variables to 59.

3. Statistical analyses

3.1. Identifying and characterizing benthic communities

3.1.1. Defining clusters of communities

In order to identify and model the distribution of the different community types, we used the approach “assemble first, predict later” (D’Amen et al., 2017; Ferrier et al., 2006). We made this choice relying on previous studies showing that the alternative approach (i.e. predict first, assemble later) provides less precise outcomes owing to propagation errors associated with each species model and also tends to predict unrealistic assemblages (Deschamps et al., 2023). Therefore, the assemble first method is to be privileged when mapping or predicting community-level indices (Deschamps et al., 2023).

We used the Kmeans clustering method to identify clusters of inventories presenting similar characteristics in terms of community composition (Hartigan & Wong, 1979). This non-hierarchical technique requires a pre-defined number of groups. The optimal number of clusters was determined by testing different numbers of clusters and comparing values of the Simple Structure Index (SSI; Legendre & Legendre, 2012). This was done with the *cascadeKM* function from the vegan R package (Oksanen et al., 2022). In all subsequent analyses, we used the *cascadeKM* to performs multiple kmeans considering different number of groups (from 2 to 20) with 1000 permutations each to evaluate the optimal number of clusters based on the SSI criterion.

To evaluate the temporal stability of the identified clusters, we performed a clustering analysis on the Hellinger transformed species-by-inventory matrix (Legendre & Gallagher., 2001; Legendre & Borcard, 2018). The clusters were then visualized for each year within a two-dimensional space defined by the first two-axis of a principal component analysis (PCA) performed on the above-mentioned matrix using the *rda* function from the vegan R package.

To model the spatial distribution of the clusters, a spatial correspondence must be made between biotic inventories that are defined with coordinates and environmental variables that come in the form of a grid. We used a 5km² grid to compromise with the different spatial resolution of the environmental predictors (varying between 1km² and 4000km²). We then aggregated inventories within environmental pixels, assuming that a species was present within a pixel as long as it is present in at least one inventory. This implies that the temporal dimension is ignored in this analysis (i.e. we assumed temporal stability).

Based on this species-by-pixel occurrence matrix summarizing species occurrence across the 2017-2023 period, we conducted a second clustering analysis with the same methodology as described above. To compare the clusters obtained with the two clustering realized, we drew an alluvial plot using the ggalluvial extension of the *ggplot2* R package.We then explored the extent to which the identified clusters represent actual ecological differences or methodological issues (e.g. inventorying bias), by exploring patterns of species richness across the clusters. A boxplot and a Wilcoxon rank sum test was performed to determine if there was a significant difference in species richness between the clusters (Appendix 5).

3.1.2. Characterizing community clusters

From the species-by-pixel presence-absence matrix, we used the Indval index to find out which species were the most representative of each cluster (Dufrêne & Legendre, 1997; De Cáceres & Legendre, 2009) using the *multipatt* function of the indicspecies R package (Caceres et al., 2016). This index combines specificity and fidelity by accounting for the frequency of occurrence of a species within a cluster relative to the other clusters (Dufrêne & Legendre, 1997). The significance of Indval values was tested using 1000 permutations.

In order to explore the drivers of cluster composition, we used three explanatory datasets : environmental predictors, predictors describing the observation process (hereafter referred to as *expertise*), and spatial predictors. The environmental predictor dataset corresponds to the environmental variables introduced above (see appendix 3). The expertise dataset was used to understand the importance of the observation process in the variability of benthic communities and in the identity and spatial distribution of the clusters. The variables used were : the average number of inventories per pixel, the number of observers per pixel and the average expertise per pixel. The average expertise was obtained by rating each of the 35 groups of observers on a scale ranging from 1 for the least experienced group to 5 for the most experienced group. This assessment was carried out by a professional taxonomist (J. Grall) who is familiar with the skills of the people involved in the sampling. Temporal information on the inventories constitutive of each cluster (average inventory year per pixel, percentage of inventories observed between 2017-2018, 2020-2021, and 2022-2023 per pixel) were also used to describe expertise variation, since the expertise was shown to neatly increase over time (appendix 6) as the observer gained knowledge about species taxonomy. The spatial predictor dataset is composed of a matrix of dbMEM (Distance-Based Moran’s Eigenvector Maps; Borcard & Legendre, 2002) and was constructed from the inventories coordinates using the *dbmem* function from the adespatial R package (Dray et al., 2018). The dbMEM method produces orthogonal (linearly independent) spatial descriptors. This method allows a much wider range of spatial scales to be covered and makes it possible to identify finer spatial patterns than just considering latitude and longitude (Dray et al., 2006).

Given the high number of explanatory variables, we used a variable selection approach to identify the most relevant variables to explain the spatial distribution of the clusters at the pixel scale (appendix 7 shows the variables selected at each step of the selection). For this purpose, we first removed collinear variables within the explanatory datasets (this was realized only for the environmental and expertise dataset). Specifically, we identified collinear variables by computing Pearson’s correlation coefficient among all variables within each dataset. When a group of variables had a correlation greater than 0.6, we kept the most relevant variable and excluded the others from further analysis. Variance inflation factors (VIF), which measure the extent to which each variable in a dataset is collinear with the others (Borcard et al., 2018) were calculated to ensure that there was no further redundancy in the remaining variables. Finally, we used a forward selection approach on a redundancy analysis (RDA; Blanchet., 2008) using the function *forward.sel* from the adespatial R package to find the most relevant variables within each set of predictors (Borcard et al., 2018). The variables chosen for each set of explanatory data are, for the environment : tidal amplitude, fetch, average salinity, current velocity, average and maximum air temperature, minimum dissolved oxygen, phosphate and ammonium concentrations, minimum wind speed and average wave height (informations on the values of the variables used are available in appendix 8). The coast type variable was not included in the selection process as we decided to keep it in the final dataset because of its ecological importance for benthic species. For the expertise dataset, the selected variables are : average year the inventories and average expertise while for the spatial datasets, the selected variables are : the first four dbMEMs (i.e. those modeling large-scale variations; Borcard et al., 2018; Brind'Amour et al., 2018). The first two MEMs show a latitudinal and longitudinal gradient in the communities while the two others show a more complex spatial pattern (Appendix 9).We decided to realize the selection on each dataset instead of on all the variables at the same time to keep the shared portion of the variance between the dataset, that was to be examined in the following analysis. After this selection we had 18 variables, comprising 12 environmental variables, 4 spatial variables and 2 expertise variables.

We used a variance partitioning approach, based on an RDA, to decompose the variation of the species presence/absence matrix (Hellinger transformed) into fractions explained by each set of explanatory variables (environmental variables, spatial dbMEMs matrix and expertise variables), fractions shared between multiple sets of explanatory variables and residuals (Clappe et al., 2018). For that purpose, we used the *rdacca.hp* function (Lai et al., 2022) from the rdacca.hp R package (Lai et al., 2023). The results of this analysis were plotted using the UpSetVP package (Liu et al., 2022).

3.2. Modeling the spatial distribution of communities

To model the distribution of each identified cluster, we used species distribution models (SDM) using the package biomod2 (version 4.2-5; Thuiller et al., 2024). To simplify the models and for ease of comparison between communities, we did not make a variable selection for each model but chose to use the set of environmental variables selected in the previous analysis as the correlated variables were already excluded. Different modeling algorithms can be used for spatial modeling, each with its own set of strengths and weaknesses (Marmion et al. 2009). For instance, despite their tendency to overfit the data (Čengić et al., 2020), algorithms based on decision trees usually display good predictive performances, especially in the presence of outliers (Iverson et al. 2008; Čengić et al., 2020). In contrast, general linear models are less prone to overfit the data and can be better at extrapolating (Yates et al., 2018) but the constrained shape of the relationship, usually leads to lower predictive performance (Yee & Mitchell, 1991). Here, we used three modeling algorithms that have proved efficient in modeling species distributions (Friedman & Popescu, 2008; Seni & Elder, 2010; Hao et al., 2019) : Generalized linear models (GLM; Trevor et al., 1992), Generalized additive models (GAM; Trevor & Hastie., 1992), and Random Forest (RF; Breiman 2001). From these three models, we used a weighted average approach to combine predictions across models. The averaging was weighted by the predictive performance of each model (area under a receiver operating characteristic [ROC]; Hanley & McNeil, 1982), the models with an evaluation score under 0.7 with the ROC metric have been excluded (Manel et al., 2002; Meißner et al., 2014).

Ideally, a model should be tested with statically independent data (Thuiller et al., 2009). In this study such data are not available. We therefore used a data-splitting approach to test the explanatory and predictive performance of the models. Specifically, we randomly selected 80% of the data for model training and 20% of the data for model evaluation. This procedure was repeated 10 times. The evaluation was calculated with the TSS (True Skill Statistic, Allouche et al., 2006) and ROC (Hanley & McNeil, 1982) indices. The mean results of the evaluation are then calculated and we obtain a quasi-independent evaluation (Thuiller et al., 2009; appendix 10). The models were then used to make predictions, using the same data as for model calibration.

To assess the capability of the models to attribute a single cluster to each pixel, we computed the Pielou's evenness (corresponding to the value of the Shannon index of the pixel for all models divided by the maximum value of the Shannon index) of the clusters' probability of presence for each pixel. A pixel with an evenness value of 0 indicates that only one cluster is predicted to occur within that pixel whereas a value of 1 means that all clusters are predicted to occur in that pixel.

All analyses were conducted with the R programming language version 4.3.2 (2023-10-31).

Results

1. Identifying and characterizing the communities and their relationships with the environment

1.1. Spatial and temporal variation of the communities across inventories

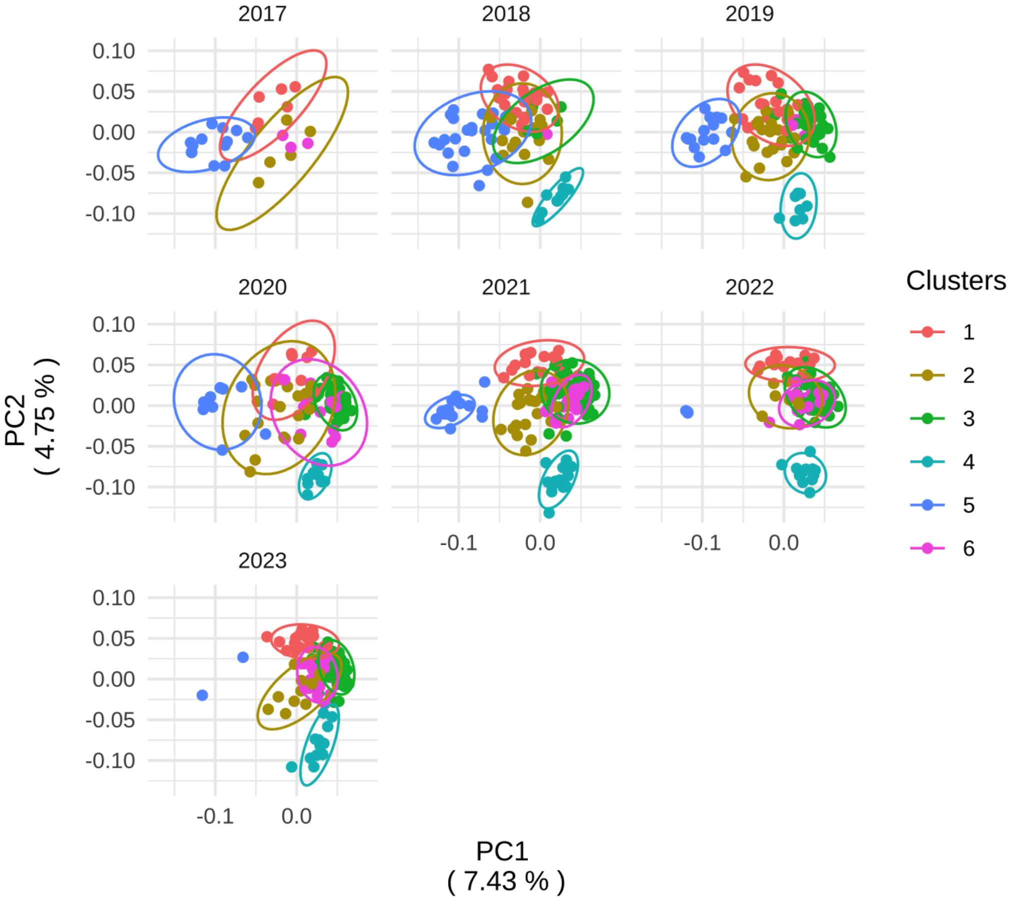


Fig 1 : PCA of the Hellinger-transformed species-by-inventories presence/absence matrix. Scaling 1. Dots represent inventories and are colored according to the 6 clusters defined from the first clustering. The ellipses correspond to each group 95% confidence interval (based on multi-normal distribution). To explore the temporal variability of community patterns, the output from this single PCA were represented separately for each year.

The first clustering, performed on the species-by-inventory matrix, gave us six clusters (Annexe 11). The cluster 2 roughly represents sheltered areas like bays (e.g Bay of Brest, Morbihan Gulf and Bay of Saint-Malo), the cluster 3 is composed of inventories from the Loire-Atlantique department, the cluster 4 represents the north of Brittany and the cluster 5 inventories from the south of Brittany. The clusters 1 and 6 do not show any clear geographical pattern. The community patterns and distinction among clusters were fairly consistent over time (Fig. 1), which support our choice to create a second dataset, excluding the temporal dimension.

1.2. Spatial variation of the communities across 5km² pixels and relationship with environmental conditions

1.2.1. Community clustering at the scale of 5km² and difference with the transect scale

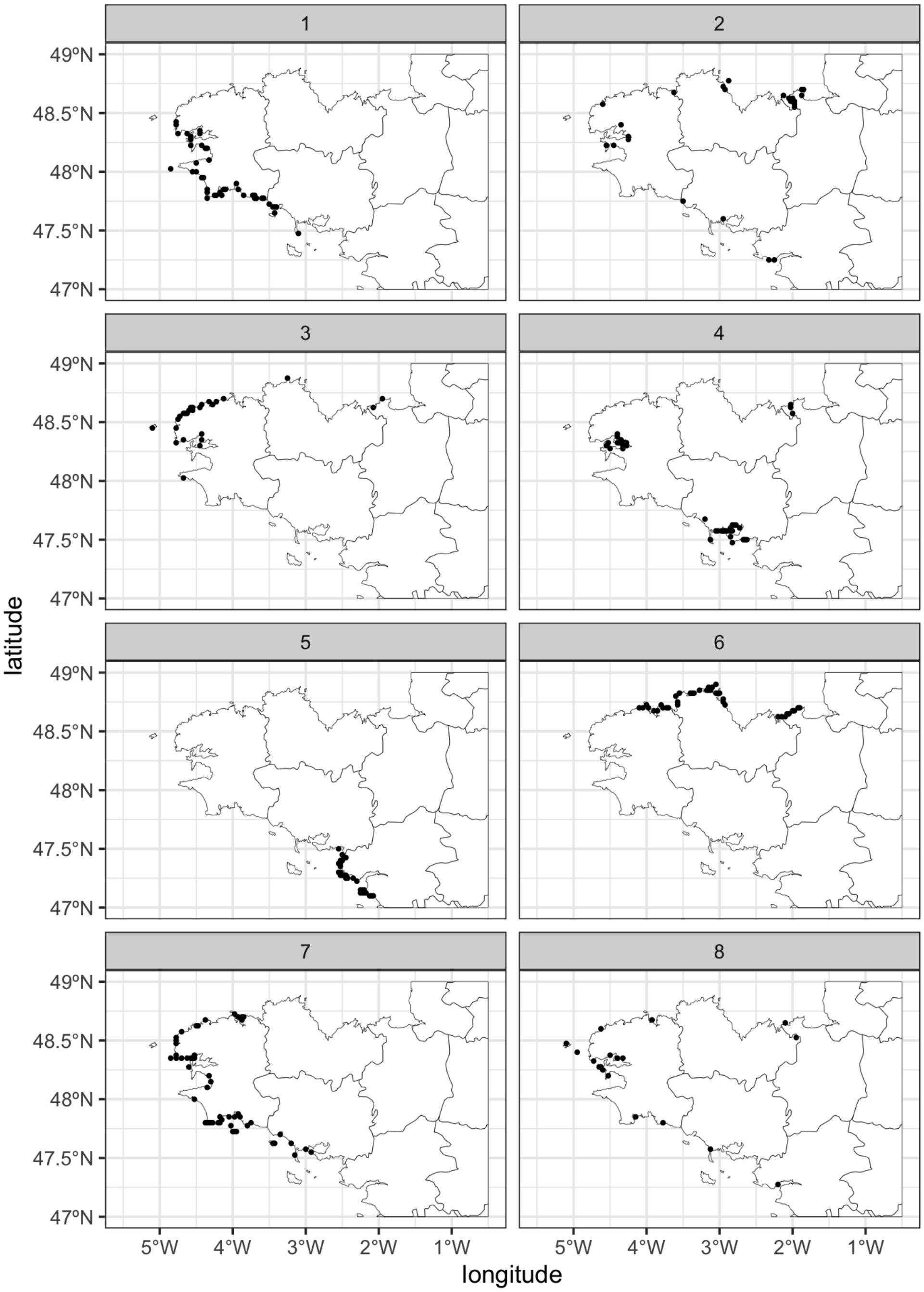


Fig. 2 : Cartography of the clusters' occurrence, from the second clusterization at the pixel scale. For the sake of clarity, the pixels were represented as points, with the coordinates representing the center of the cell.

From the second clustering, we obtained 8 clusters (Fig. 2). Cluster 1 is located in the south, cluster 3 is situated in the northwest, cluster 4 represents sheltered areas (e.g. bay of Brest, the Gulf of Morbihan and the bay of Saint-Malo), cluster 5 shows sites from the Loire-Atlantique department, cluster 6 displays northeast sites, cluster 7 is situated at the southwest and cluster 8 and 2 have an undefined geography. From now on, the clusters will be renamed as follows : 1 : South Brittany, 2 : Undefined, 3 : Northwest, 4 : Sheltered areas, 5 : Loire-Atlantique, 6 : North Brittany, 7 : Southwest, 8 : cluster 8.

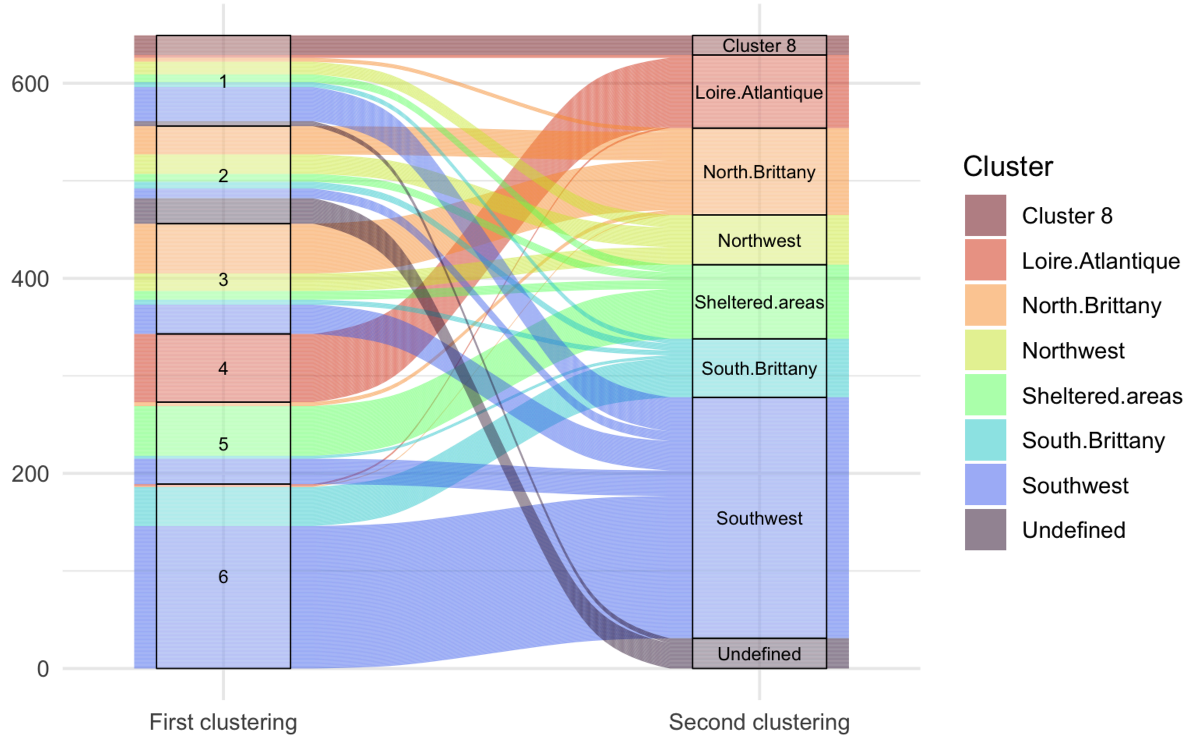


***Fig.* 3** : Alluvial plot showing the correspondence between the clusters of first clustering (at the inventory scale) and the clusters of the second clustering (at the pixel scale). The clusters represented by colors are those of the second clustering. The y-axis represents the number of inventories classified in each group. For the second clustering conducted at the pixel scale, all transects within a given pixel were attributed to its corresponding group.

Correspondence between the clusters obtained from the two clusterings were drawn and shows that the clustering at the inventories and pixel scales are fairly consistent (Fig. 3). Clusters at the pixel scale correspond mainly to one cluster at the inventory scale, except the Southwest, North Brittany and Northwest clusters that take origin in multiple inventory scale clusters.

1.2.2. Characterizing the pixel scale clusters

The southwest cluster is particularly rich with an average of almost 94 species per cell (appendix 2 and 12). It also contains more inventories than the other clusters with 52 pixels containing an average of 4.75 inventories including few older inventories and the highest mean expertise. The Loire-Atlantique cluster has a high species richness with an average of 57,1 species per pixel, as well as sheltered areas and South Brittany which show respectively 65,14 and 54,3 species per pixel on average. The north Brittany cluster shows a species richness of 42,31 on average, it also has few older inventories and the lowest average expertise (2,08). Cluster Northwest and Undefined show poor species richness with a mean of respectively 29 and 20,83 species per pixel. These clusters also represent few inventories from 2022 to 2023 and the Undefined group has a relatively low expertise (3.30). Cluster 8 is the poorest with only 5.61 species per pixel on average and represents older inventories with 60 inventories realized before 2019 and no inventories from 2022 to 2023. It is also the least represented cluster with only 18 pixels and just over 1 inventory per pixel. However, the cluster has a good average expertise (4.22). The results show that the species richness in cluster 8 is significantly lower (W ranging from 404.5 to 936.0; $p<0.0001$) than in the other clusters (appendix 5). Combined with the fact that most inventories constitutive of this cluster are from the beginning of the OBCE program, this cluster (and associated inventories) was excluded from further analysis since the low richness is not really representative of the communities.

Clusters strongly varied regarding community compositions (Appendix 13). The Loire-Atlantique community highly differed from the others, being mostly characterized by molluscs (e.g. *Magallana gigas, Modiolus barbatus, Heteranomia squamula, Tritia incrassata, Alvania lactea*), a few crustaceans (e.g. *Pachygrapsus marmoratus, Clibanarius erythropus*), echinoderms (e.g. *Asterias rubens, Psammechinus miliaris*) and polychaete species such as *Sabellaria alveolata*. South and sheltered communities were rather similar and mostly associated with sponges (e.g. *Aplysilla rosea, Aplysilla sulfurea, Ophelia spongia, Terpios gelatinosus*), gastropods (e.g. *Rissoa parva, Haliotis tuberculata*) and nudibranchs (e.g. *Berthella plumula*). The North and Northwest clusters were associated with crustaceans (e.g. *Galathea squamifera*) and gastropods (e.g. *Calliostoma zizyphinum*). The undefined cluster seems to be characterized only by very common species (e.g. *Littorina littorea*).

Four clusters (Southwest, Loire-Atlantique, Sheltered areas and North Brittany) were associated with indicator species with an Indval index above 0.35 (Appendix 14). The majority of the indicator species are found in the southwest cluster, with 27 species of various phyla being specific to this group. The loire-atlantique cluster has 12 indicator species, including 9 molluscs, the sheltered cluster shows 4 indicator species including 2 ascidia, 1 gastropod and a polychaete while the North Brittany cluster is characterized by one sponge.

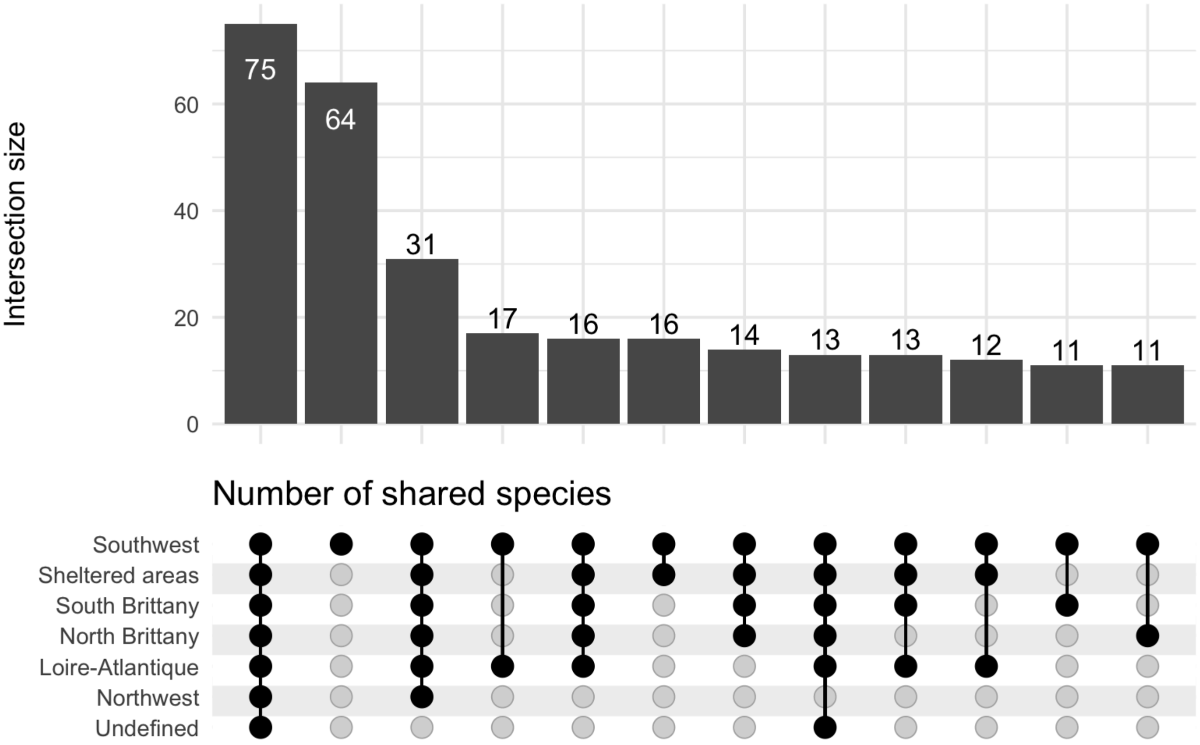


Fig. 4 : UpSet plot showing the number of unique and shared species for each cluster. The lower panel represents, with the black dots, the clusters included in each combination set. The upper panel displays the number of species shared by each combination. Only the combinations of clusters sharing at least 10 species have been represented.

Out of the 432 species, 75 species (17%) are shared by all the clusters (Fig. 4). Several species (64 species, 15%) are found exclusively in the Southwest cluster. It should also be noted that the Southwest cluster is in every combination of groups represented by at least 10 shared species. This analysis highlights the exceptional diversity of the Southwest community, in contrast to the Undefined community, which consists almost entirely of generalist species (i.e. species found in other clusters), only 0.6% (2 species) of its diversity being unique to this cluster.

1.2.3. Environmental variations among clusters and drivers of community variation

The environmental characteristics associated with each cluster (Appendix 15) highlight the unique environmental conditions of the Loire-Atlantique region, exhibiting low values for mean salinity, tidal amplitude, and fetch, but high values for mean and maximum air temperatures, current velocity, and minimum concentration of phosphates. In contrast, Northwest and North Brittany are characterized by low values of maximum air temperature, mean wave height, and mean air temperature, but high values of mean salinity. North Brittany also has low current velocity and high tidal amplitude. South Brittany and the Southwest region show low values of maximum air temperature and tidal amplitude. The Undefined and Sheltered areas are noted for their low values of fetch and minimum salinity, with higher average temperature in Sheltered area and lower current velocity and mean wave height in the Undefined cluster.

The coast type is relatively homogeneous among the clusters (appendix 16), the Loire-Atlantique cluster is the one with the least rocky and more sandy coastline. The Sheltered areas are those with a greater proportion of rocky coastline.

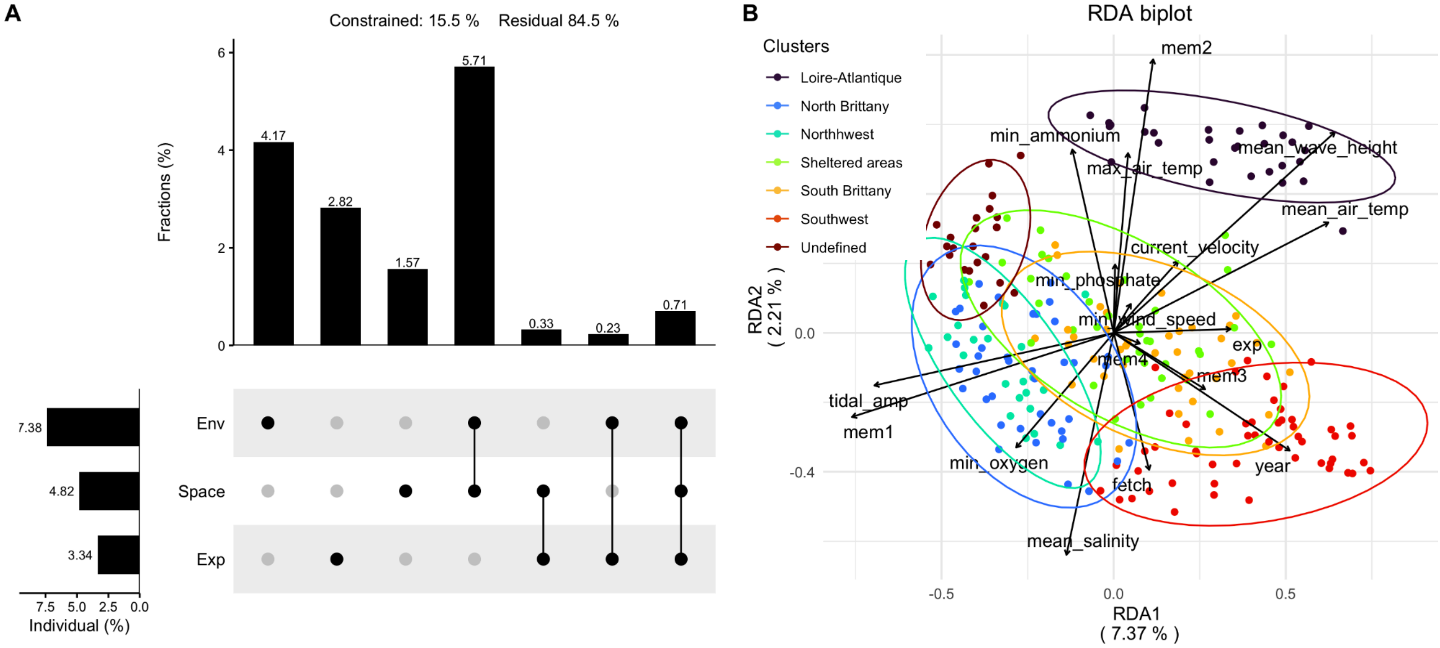


Fig. 5 : **A** : UpSet plot of a variation partitioning showing the percentage of variance in the Hellinger transformed species presence/absence matrix at the pixel scale explained by each group of explanatory variables : Env : environmental, Space : spatial, Exp : experience. The lower panel represents, with the black dots, the datasets included in each combination set. The upper panel displays the variance explained by each combination, and the left one displays the total variance explained by each single set of variables. **B** : RDA biplot on the species presence/absence matrix at the pixel scale. The percentage represents the variation explained by each axis, the arrows display the variables and the sites are grouped by clusters represented by a color and an ellipse corresponding to their 95% confidence interval (based on multi-normal distribution).

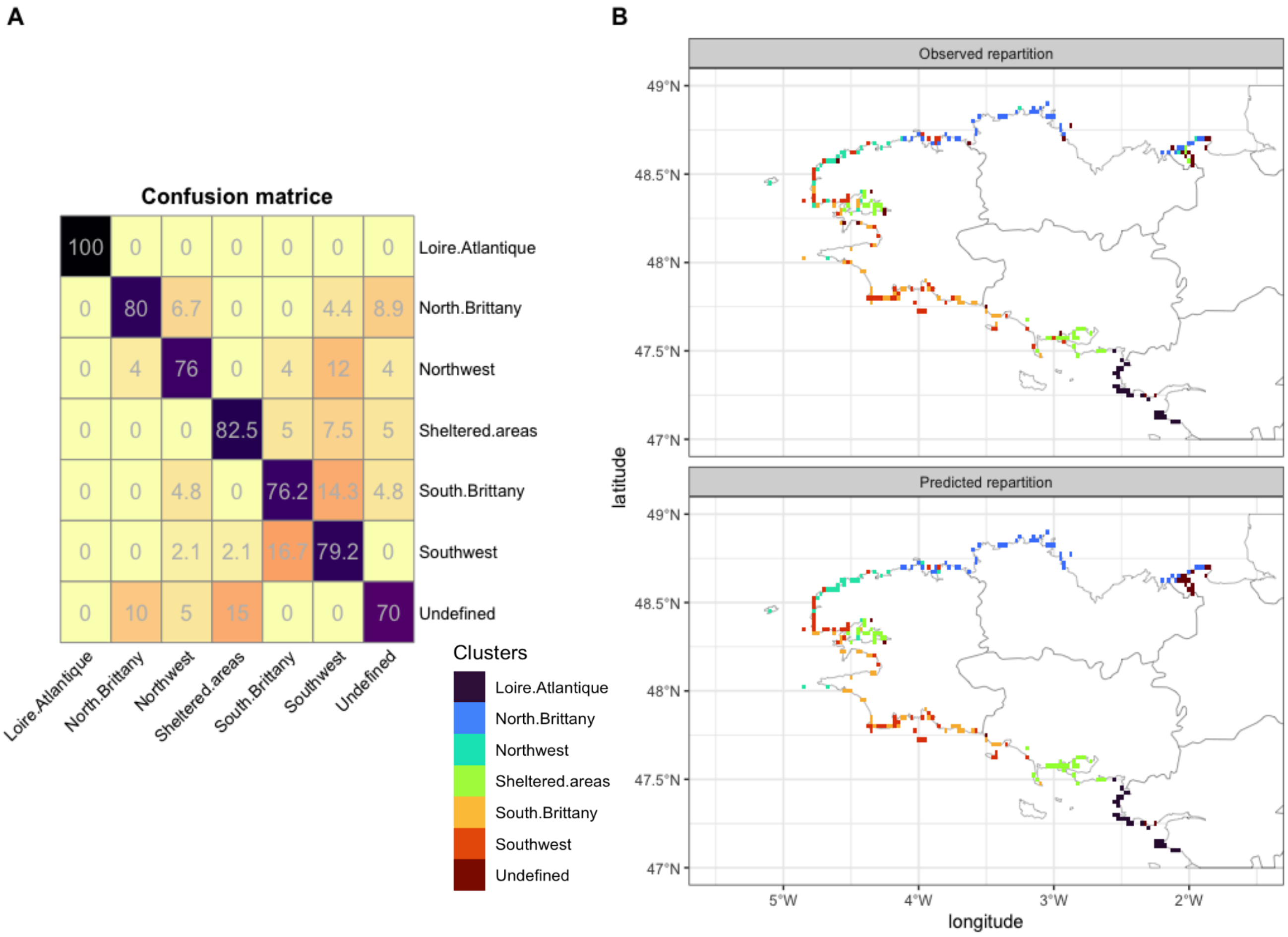
The largest proportion of variance in the species presence-absence matrix, 5.71%, is explained by the shared fraction between environment and space (Fig. 5A), followed by the effect of the environment alone (4.17%), of the observing process (2.82%) and of the spatial predictors (1.57%). In total, environmental variables accounted for 7.38% of the variance, spatial components for 4.82%, and the observer effect for 3.34%. Even if an important part of the variance explained by the environment is shared with large-scale spatial patterns, there is still 4.17% of the variance that can only be attributed to the environment.

The clusters show clear variation regarding geography and environmental conditions (Fig. 5B). The first axis represents the latitudinal gradient (MEM1) related to tidal amplitude and is positively associated with Southwest and Loire-Atlantique clusters and negatively associated with Undefined and north clusters. The second axis represents the longitudinal gradient (MEM2), related to temperature and salinity and is positively associated with Loire-Atlantique and Undefined and negatively associated with the Southwest cluster. The expertise (represented by the average inventories year) is positively associated with the Southwest cluster and negatively with the Undefined cluster.

2. Modeling of the cluster occurrences

The Loire-Atlantique model demonstrates the highest predictive power, with ROC values exceeding 0.9 for all three algorithms (GAM, GLM, RF; Appendix 10). The other clusters show values of ROC ranging from 0.8 to 0.9 in all 3 algorithms, except for Sheltered areas and North Brittany clusters exceeding 0.9 with RF. We observe similar results with the TSS metric, with lower values ranging from under 0.4 to over 0.8

The predicted probability of presence (Appendix 17, predicted occurrence for each cluster type is presented in the appendix 18) shows congruent spatial patterns with the observed occurrence of clusters.



	Loire.Atlantique	North.Brittany	Northwest	Sheltered.areas	South.Brittany	Southwest	Undefined
Loire.Atlantique	100	0	0	0	0	0	0
North.Brittany	0	80	6.7	0	0	4.4	8.9
Northwest	0	4	76	0	4	12	4
Sheltered.areas	0	0	0	82.5	5	7.5	5
South.Brittany	0	0	4.8	0	76.2	14.3	4.8
Southwest	0	0	2.1	2.1	16.7	79.2	0
Undefined	0	10	5	15	0	0	70

Fig. 6 : **A** : Confusion matrix showing the percentage of correspondence between the predicted occurrence of the clusters (in rows), based on the cluster with the highest probability of presence for each pixel, and the observed occurrence of the communities (in columns). **B** : Maps comparing the observed repartition of the clusters (top) to

their predicted repartition (bottom). In the predicted repartition map, the cluster represented for each cell is the one showing the higher predicted presence probability (maps of the predicted probability of each cluster can be found in Appendix 17).

The models all showed a good predictive performance with a rate of correct predictions ranging between 70% and 100% (Fig. 6A). The highest error rates occurred between the South Brittany and Southwest clusters and the Sheltered areas and Undefined clusters. Consequently, the comparison between predicted and observed communities presented similar spatial patterns (Fig. 6B).

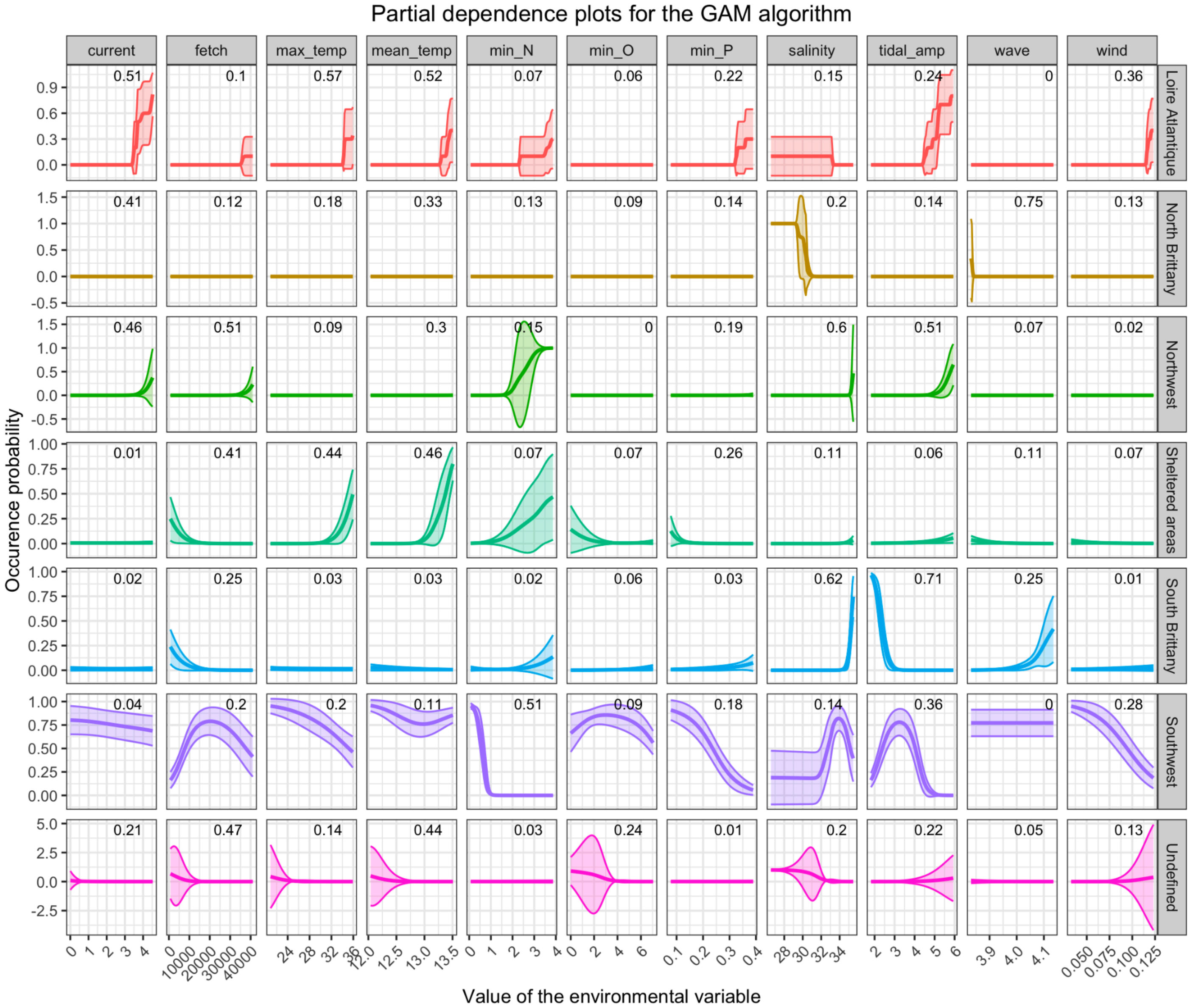


Fig. 7 : Partial dependence plots for the GAM algorithm showing the influence of each variables (current velocity, fetch, maximum and average ar temperature, minimum dissolved concentration of ammonium, oxygen, and phosphate, Salinity, tidal amplitude, mean wave height and wind speed), in columns, on the predicted probability of presence of each cluster, in rows and by color. The line represents the mean values of the 10 runs and the ribbon

represents the standard deviation of the values. The importance of the variables are printed in the top right corner of each facet.

For the Loire-Atlantique cluster, the most important variables were the average and maximum air temperatures (with importances of 0.82 and 0.58, respectively), current velocity (0.59), and tidal amplitude (0.32, Fig. 7). Each of these variables showed an increased probability of presence of the cluster as their values increased. In North Brittany, despite linear models only showing a negative correlation with average wave height (importance of the variable : 0.75), the RF model highlights tidal amplitude as an important variable for the model (0.67, appendix 19), with higher values increasing cluster presence probability. The Northwest cluster's presence is majoritarily influenced by salinity (0.51), current velocity (0.39), fetch (0.31), and tidal amplitude (0.31), all positively correlated with the presence of the cluster. Contrastingly, in Sheltered areas, the most important variables are average and maximum air temperatures (0.48 and 0.4) which increase with the cluster's presence probability, while fetch (0.32) is negatively correlated. South Brittany is mostly correlated to tidal amplitude (0.65), salinity (0.61), and average wave height (0.41). Salinity and wave height increase with increasing probability of cluster presence, contrary to tidal amplitudes. The Southwest cluster shows a more complex relationship: higher minimum ammonium levels (0.44) and maximum air temperature (0.30) decrease the likelihood of the cluster's presence, while tidal amplitude (0.35) appears optimal between values of 2 and 4. For the Undefined cluster, variables with the greatest importance are : fetch (0.47) and wave height (0.38) negatively impact presence. The cluster is mostly present in estuary coast types (importance of 0.61 in the GAM algorithm, appendix 21). The 3 algorithms consistently highlighted similar variable importance and trends (Appendix 19, 20 and 21).

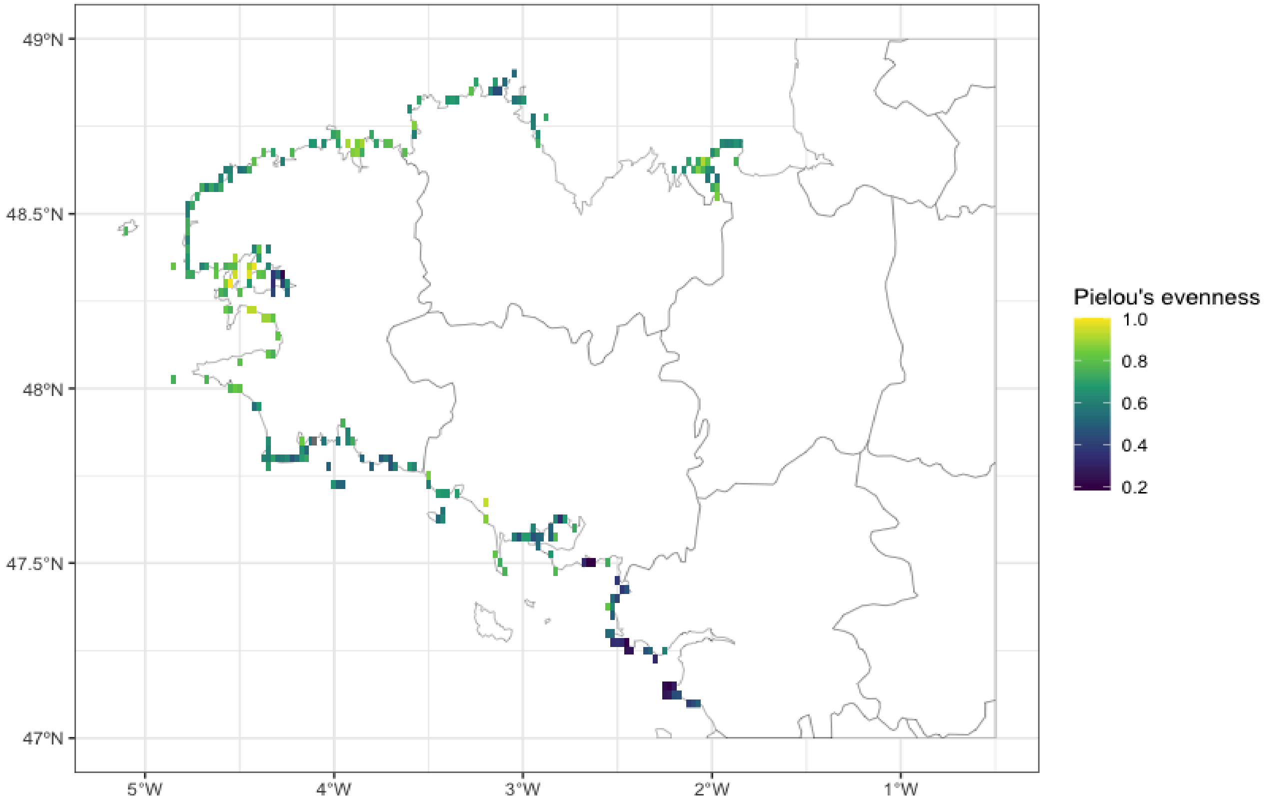


Fig. 8 : Cartography of Pielou's evenness from the predicted probability of presence of the clusters. An evenness value of 0 indicates that one cluster has a predicted probability of presence near 1 while others have near 0 probabilities, whereas an evenness value of 1 means that all clusters are predicted to have an equal probability of presence.

The Loire-Atlantique department exhibits the lowest evenness values between the probabilities of the different clusters (Fig. 8). In these pixels, the predicted probability of occurrence of the Loire-Atlantique cluster is considerably higher than the predicted probability of occurrence of the other clusters. Conversely, the highest evenness values are observed in West Brittany. In this area, the model makes less differences in the probabilities of the different clusters.

Discussion

In this study, we used the OBCE citizen science program to describe and model the biogeography of rocky shore macrofauna communities along Brittany's coastlines and to identify the drivers of spatial variations in macrofaunal communities. Considering data from all inventories (mixing spatial and temporal variations in communities), we were able to identify homogeneous clusters of communities that appeared to be fairly stable over time suggesting a form of robustness in the identified clusters. This analysis, coupled with data exploration, also allowed us to explore the effect of expertise on the temporal variability of communities and to limit its impact on the subsequent analysis. In a second analysis - conducted at the pixel scale, hence ignoring the temporal dimension - we retrieved the previously identified clusters. These clusters were characterized by different species assemblages and were associated with different environmental conditions. SDMs performed well in explaining and predicting their spatial distributions providing avenues for predicting the probability of occurrence of these clusters in other areas or time periods.

1. Characteristics of the clusters

We obtained both large scale and finer scale clusters (summary of the identity, environmental and biotic characteristics of each cluster is available in Table 1). Large scale clusters are divided in two biogeographic provinces : the Lusitanean province (comprising the Bay of Biscay) and the Boreal-Lusitanean province (comprising the English Channel; Dinter, 2001), separated by the Ushant front in the Iroise sea. These results are consistent with the biogeography of intertidal areas proposed in earlier studies, including studies on subtidal rocky shore communities (Derrien-Courtel et al., 2013).

The clusters present in the Bay of Biscay (Loire-Atlantique, South Brittany, Southwest) show higher species richness than those from the English Channel (Northwest and North Brittany). They represent species coming from a hot-temperate species group (Dinter, 2001). Some species that originated from the further south can also be found in this cluster probably due to distributional shifts owing to climate change. For exemples, we noted the presence of species newly-established in the area, particularly nudibranchs like *Spurilla neapolitana* (Grall et al., 2015), *Babakina anadoni*, *Felimida krohni* or *Berghia verrucicornis* (Droual & Bridier., 2020). We also observed a large diversity of fixed fauna (e.g. sponges, bryozoans) that can be explained by the foreshore heterogeneity (Archambault & Bourget, 1996). The two clusters present in the English Channel (Northwest and North Brittany) show lower species richness than the ones in the Bay of Biscay, as already observed in other studies on soft sediment communities (Gallon et al., 2017). These areas presented species characteristics from homogeneous rocky intertidal shores. Two finer scale clusters were further obtained : Sheltered areas and an Undefined cluster. The sheltered areas cluster shows a high species richness with species characteristic from protected conditions, such as *Raspailia hispida* (present exclusively in this cluster; Picton et al., 2011) or *Ascidiella aspersa* (which have been identified as an indicator species for this cluster; Lynch et al., 2016). This cluster also presents multiple introduced species like *Hemigrapsus sanguineus* (present exclusively in this cluster). The Undefined cluster shows poor species richness, predominantly represented by widespread species.

1.1. Drivers underlying biogeographical patterns

This variation in species richness could be associated with expertise bias. Indeed, with this study, we highlighted a non negligible bias due to i) difference in expertise among the volunteers, ii) temporal variation of expertise on the duration of the program, due to improvement of the knowledge of volunteers and iii) heterogeneous sampling effort with a larger amount of inventories in western Brittany. The species composition of some clusters were likely to be influenced by this bias, i.e. cluster 8 which was characterized by particularly poor species richness and predominantly consists of inventories sampled at the very beginning of the project. In contrast, despite having a small number of inventories, it showed a relatively good average level of expertise. This cluster was mostly represented by inventories done by experts focusing on identifying only one or a few species of particular interest, rather than conducting an exhaustive inventory (Christian Hily, personal communication). This selective sampling approach means that the data do not accurately reflect the true biodiversity of the area, skewing the results towards a narrower range of species. Therefore this cluster was removed from further analysis. The Undefined cluster shows low average expertise and a small number of inventories, which could explain that it exhibits the lowest species richness after cluster 8. In contrast, the Southwest cluster, which exhibits the highest species richness, displayed the highest number of inventories and the greatest average level of expertise among volunteers. The amount of recent inventories further enhances this group's total number of species, leading to an over-representation of rare species.

Although expertise seems to play a role in the observed species richness of the clusters, when quantified, its impact on the variation of communities' species composition appeared minor compared to that of the environment and space.

As mentioned above, the Northern Bay of Biscay exhibits higher species richness. This can be explained by the complex mosaic of habitats present in this area (Dinter, 2001; Le Loc'h et al., 2008), allowing the coexistence of a large number of species with different ecological affinities. The large input of freshwater, particularly from the Loire and Vilaine rivers, also induces a wide range of physicochemical conditions that could increase the region's biodiversity by being suitable for both cold and warm water species, as well as species adapted to brackish water (Dinter, 2001; Gallon et al., 2017). The hydrodynamic conditions of the area, characterized by a high exposure to waves and currents, can favor the presence of engineer species such as Mussels (*Mytilus edulis;* Hammond & Griffiths, 2003). These species form complex structures that facilitate the apparition of a large variety of species by providing habitat, food supply and reducing physical stress from current and waves (Hammond & Griffiths, 2003; Van der Zee et al., 2015).

The low species richness in the English Channel (Northwest and North Brittany) has already been observed in other studies (Gallon et al., 2017) and can be explained by ecological, geological and evolutionary processes. Indeed, the English channel opened quite recently (after the last glacial maximum, approximately 20,000 years ago). The recent colonization of the English Channel by marine species could explain the low species richness observed (Ellingsen & Gray, 2002).

The two biogeographical regions cited above are separated by the Iroise sea, corresponding to the Northwest and Southwest clusters. These two clusters represent an ecotone, a "sharp transition zone between two or more different ecological communities or regions, that often occur in areas of steep environmental transition, along environmental gradients" (Kark, 2007). The Southwest cluster represents the transition area between north and south Brittany, leading to a high variability of habitats and environmental conditions and thus higher species richness (Kark & Van Rensburg, 2006; Gallon et al., 2017). This high species richness has been observed in many transition zones, hosting the different species from the area it divides and sometimes species that are characteristic of the ecotone (Dinter, 2001; Kark, 2007). We observe this special ecotonal-species in the Southwest cluster as 15% of all the species in the study are only found in this region. The Northwest cluster represents the geographical barrier : the Ushant front. Being situated at the front, its low species richness can be explained by the lack of exchanges between species from this cluster and the other clusters due to the front, limiting larval dispersal (Ayata et al., 2010). In the summer, the Iroise sea hosts a thermohaline front (the Ushant front), and constitutes a cold-water enclave (Dinter, 2001; Penard, 2009; Gallon et al., 2014, Dauvin, 2015; gallon et al., 2017). Thermal and saline structures in these regions can disrupt the dispersal between the Eastern Channel and the Bay of Biscay (Pingree et al., 1982; Boyer et al., 2009; Ayata et al., 2010), potentially limiting the number of shared species between these two coasts. Indeed, the pelagic stages are highly sensitive to hydrological discontinuities—areas where water bodies differ in their physical and chemical properties—such as thermal and haline fronts (Ayata et al., 2010; Gallon et al., 2017). This cluster is composed of generalist species, found all around Brittany. It seems like only these eurythermal and euryhaline species can support the conditions of the front, leading to poor species richness in this area of the ecotone. Species at their northern limit have long fluctuated in their distribution. They have occasionally managed to cross the Ouessant front, but they did not establish themselves permanently, probably due to physical factors such as water temperature preventing reproduction. After decades of fluctuation, some species have finally managed to cross the barrier and establish themselves in the eastern English Channel and/or along the English coasts or in eastern Brittany (e.g. *Clibanarius erythropus*; Southward & Southward, 2009; Patterson et al., 2020, *Eriphia, Pachygrapsus marmoratus;* Ingle

& Clark, 2009, *Asthenognatus atlanticus*; Jourde et al., 2012, and *Gregariella semigranata*; Jacques Grall, personal communication). However, they may remain scarce along the northwest coasts.

The front represents a physico-chemical barrier for species but also for gene flows, as shown on studies on abalones (Chauvaud, 2022), polychaetes (Jolly et al., 2005) and *Ophiothrix fragilis* (Muths et al., 2009), a species found in the OBCE dataset. These studies showed differences in genetic structures within species between the two sides of the front. These studies place ecotones as important zones of evolution processes and lead to the idea that transition zones could be regions where speciation may take place (Jolly et al., 2005; Kark, 2007). In the ecotone, the evenness in the cluster probabilities predicted by the models is the highest, meaning that the models struggle to find which cluster is the most likely in this area. This suggests that the environmental conditions in west Brittany are suitable for different clusters.

The clustering also highlighted two finer scale clusters : the Sheltered areas and the Undefined cluster. The Undefined cluster when first defined did not show any obvious geographical pattern. Yet, the environmental characteristics revealed by the distribution models (low values of salinity, current velocity, wave height, fetch and wind speed with high values of Spim and nitrates) combined with its geographical distribution, lead us to believe that this cluster actually corresponds to sheltered regions of varying salinity. The model also showed the coast type variable to be the most important for this cluster, with estuaries presenting the highest probability of presence. Estuaries being subject to high anthropogenic pressure, including pollution and habitat modification, can explain the low biodiversity observed in estuarine communities (Vasconcelos et al., 2007; Wetz & Yoskowitz, 2013; Monteiro et al., 2016).

It should be noted that the assumption that any species not detected in an inventory was truly absent, implying no observation errors, while generally accurate, is not completely realistic as exhaustive inventories can still miss some species, introducing inevitable errors. Furthermore, discrepancies in data resolution can also introduce biases. For instance, finer spatial resolution improves prediction accuracy for fixed or locally mobile organisms (Guisan & Thuiller, 2005), whereas lower resolution data can reduce predictive performance.

Overall, our results are consistent with the biogeographic classification found in the litterature (Dinter, 2001) and similar to the results of Tempera et al. (2019) obtained from a clustering of Brittany's coasts based on environmental abiotic variables, with northern and southern clusters, bays and estuaries. These similarities with our study suggest that community composition shows similar spatial patterns as environmental conditions. Some differences can still be observed as we did not show clusters representing areas exposed to currents and waves. The biogeographic regions identified, as well as the placement of the barrier between these two regions are congruent with what have been found in Hayden et al. (1984) and Golikov et al. (1990) but show a more precise spatial delimitation. We showed that the influence of the expertise does not overtake the influence of space and environmental conditions to explain the variation in species composition and in the biogeography of the communities.

Further studies and data would be needed to confirm or refute the above hypotheses describing the west clusters as an ecotone and putting the geographical barrier in the region of the Northwest cluster. In the context of climate change, special attention should be paid to these clusters and their specific evolution, particularly the weakening of the Ouessant front, as well as the homogenization of the biodiversity of the shorelines at the tip of Brittany. These evolutions could be predicted by the models generated in this study. Macroalgae should in the future be included in the analysis, considering not only species but perhaps also biomasses as a driving factor as well.

2. Evaluation of a cityzen science program : the OBCE

Working with citizen science data is highly practical for researchers as they can access a large amount of cost-effective data (i.e. by optimizing time, effort and resources). However, despite its convenience, it still has its challenges. Working with non-professional volunteers can result in lower quality data. A wide range of data quality can be observed (Feldman et al., 2021), making it difficult to compare observations. These discrepancies can stem from differences in observer expertise (Cooper et al., 2007; Yu et al., 2010; Kelling et al., 2015). In particular, species richness might have artificially increased from 2017 to 2023 due to increase in volunteers' expertise. Additionally, some volunteers can have a high experience but be specialized in a specific taxa and lead to an over-representation of this taxa in the region where observations are made. Environmental conditions during the day of the inventory can also play a role in data quality. For instance, depending on the tidal coefficient, lower parts of the shore can emerge or not. In this case, some species that are only found in the lowest part of the intertidal region can be accessible or not. The time spent doing the inventory, the search area and the number of observers involved can also play a role as less common species can be harder and take more time to find (Kelling et al., 2015). Another bias that can be introduced with citizen science is the heterogeneity in spatial coverage. Indeed, we observed significantly more inventories in west Brittany than in east Brittany, which may lead to an apparent increase in species richness in the over-represented regions.

To enhance data quality, it is crucial to provide standardized observation methods and train volunteers (Mackechnie et al., 2011). Communication between volunteers and experts can also improve data quality by enhancing volunteers' knowledge (Sullivan et al., 2014). Additionally, paying attention to details like sampling effort, identification uncertainty, and species absences can help understanding differences in observer expertise (Amano et al., 2016).

The expertise bias observed in this study was taken into account by incorporating expertise variables to assess the quality of the data. The species richness was also taken into account in the study as a measure of expertise, leading to the exclusion of the inventories belonging to cluster 8.

Furthermore, even if the expertise dataset was the one that better explained variations in community composition, after the environmental dataset, the effect of space and environment appears to be mostly independent from expertise.

Citizen science shows many challenges but is increasingly used in research and particularly in biogeography such as when using SDM as it allows access to a considerable amount of data with high spatial and temporal resolution. Furthermore, even if data quality is usually lower than with professional inventories, the large amount of data offers more statistical power and in the end can yield results comparable to those collected by professionals (Greenwood, 2007; Schmeller et al., 2009). Remarkably, over half of the species discovered between 1998 and 2007 were identified by amateur taxonomists (Fontaine et al., 2010; Silvertown, 2010), highlighting the significant contribution of non-professionals to biodiversity science (Bœuf et al., 2012) that can also serve as a tool to monitor invasive species (Kleitou et al., 2019).

Another advantage of citizen science is that it actively involves local communities in ecological conservation efforts and raises awareness about environmental issues and conservation initiatives (Devictor et al., 2010).

Despite the presence of a significant expertise bias, the identity of the clusters remains stable over time, reinforcing the confidence regarding cluster identification and associated biogeographic patterns. This might not have been observed at the species scale, and supports the chosen community scale approach as it allows us to smooth the noise coming from this bias. The OBCE program's data quality has seen continuous improvement over the years and is gaining stability as volunteers benefit from ongoing training, including regular taxonomic workshops led by professionals and post-inventory meetings to discuss observed species. Although the program is still young, entering its eighth year, it shows considerable promise and potential for the future by providing extensive coverage of the rocky shores of historical Brittany, enabling fine scale studies of a wide variety of species.

In the future, it would be valuable to harmonize sampling efforts. This could be achieved by revising the sampling protocol toward something more stringent and informative. For instance, additional information such as the tide coefficient, the time dedicated to the inventory, the number of volunteers involved, and the approximate size of the area surveyed could be used to better account for sampling conditions in analyses. Additionally, establishing a few consistently monitored sites across Brittany would be beneficial for studying temporal variations more effectively. These sites could be strategically located within each of the seven clusters identified by this study. This would also allow us to acquire information on the true species richness (e.g. by doing rarefaction curves or species accumulation curves, as done in Kelling et al., 2015) and quantify the level of expertise.

3. Modelisation of Brittany rocky shores

Species Distribution Models (SDMs) is also based on the assumption that the community types modeled are in pseudo-equilibrium with their environment (Guisan & Theurillat, 2000; Guisan & Thuiller, 2005; Ferrier et al., 2006; D'Amen et al., 2017; Deschamps et al., 2023). Although our findings indicated that the identity of the clusters remains relatively stable over time, confirming their equilibrium with the environment is challenging.

Despite this assumption whose validity cannot be fully proven, models demonstrated good predictive power, particularly within the Loire-Atlantique cluster. The partial dependence plots mostly confirmed the findings from earlier analyses and is congruent with the existing literature. These plots reveal non-linear relationships between clusters' probability of presence and environmental drivers, offering deeper insights into the factors influencing the spatial patterns of these clusters. Notably, they confirmed the hypothesis that the Undefined cluster is associated with estuaries, providing a clearer understanding of its environmental characteristics.

One possible improvement in future models could be to better account for the different scales of community spatial patterns. The use of positive dbMEM in the models allowed us to take into account broad scale spatial patterns, which is already an improvement compared to the sole use of the inventories coordinates (that would have only allowed us to study the longitudinal and latitudinal gradients). However, many parameters are still excluded, such as the geographical barriers or the currents. For example, in this study, we believe that the Ushant front prevents the interactions between some, yet close, communities. In order to include these parameters and have a model that is closer to reality, we could take into account the distance from the shoreline and integrate current and larval dispersion models. This improvement to the model would allow to better understand how connectivity affects the spatial distribution of the clusters. At the end, these models could be used to generate continuous spatial predictions and future projections in order to study the influence of climate change in the repartition of the clusters.

Table 1 : Summary of the results from this study

Initial interpretation	Final interpretation	Environmental characteristics	Biotic characteristics
Loire - Atlantique	Loire - Atlantique	Most sandy region, Influence by the Loire river : low salinity, high concentration of chla, Spim, nitrates, phosphates and silicates. Southern East region : high air temperature, dryer climate High exposition to ocean swells and currents	High species richness, dominated by molluscs species and south affinity species (*Eriphia, Clybanarius, Pachygrapsus...*). High presence of exotic species (*Crepidula fornicata,*

			Magalana gigas, Hemigrapsus sanguineus)
South Brittany	South Brittany	Diverse range of environmental conditions and habitat.	High species richness, species composition representative of exposed conditions and dominated by *Mytilus edulis* and its associated predators (e.g. *Marthasterias, Asterias*)
Southwest	West Brittany / Ecotone area	Diverse range of environmental conditions and habitat.	Greatest species richness with 15% of species exclusive to this cluster. Include a large part of fixed species (e.g. sponges and bryozoans) and some species novelty found in Brittany (e.g. *Spurilla neapolitana, Babakina anadoni, Felimida krohni* or *Berghia verrucicornis*)
Northwest	Ushant front	Low air and water temperature and high salinity, humidity and nebulosity	Low species richness, representing generalist (i.e. eurythermal and euryhaline) species.
North Brittany	North Brittany	Low values of maximum air temperature, mean wave height, current velocity and mean air temperature, but high values of mean salinity and tidal amplitude.	Low species richness, dominate by species of rocky shore gastropods (*Gibbula, Patella, Calliostoma, Haliotis*)
Sheltered areas	Sheltered areas	Protected zones : small fetch, tidal amplitude and wind speed Anthropized zones : high concentration of chla, nitrates and ammonium	High species richness, presence of sheltered and invasive species only found in this cluster
Undefined	Estuaries	Low values of salinity, current velocity, wave height, fetch and wind speed with high values of Spim and nitrates	Poor species richness, representing mostly common species
Cluster 8	Excluded cluster	Cluster driven by methodological bias (non-exhaustive inventories) and not environmental ones	Very low species richness

Conclusion

With this study, we characterized and modeled the biodiversity of rocky shore fauna at a regional scale (i.e. across Brittany) using data from a citizen science program, the OBCE. We observed an expertise bias with temporal variations in the database and included expertise as an explanatory variable in our analyses. Overall, while the influence of expertise bias was significant, its effect can be disentangled from environmental and spatial conditions that showed a stronger influence on the spatial variations observed in the clusters' repartition. Ultimately, the clusters we identified showed patterns consistent with previous descriptions of the biogeography along the coast of Brittany and our primary hypothesis. These results provide new insights into the biogeography of rocky shore macrofauna at a community level with high spatial resolution. This allowed us to precisely describe the transition zone in terms of spatial distribution, ecological characteristics, and species composition. The models developed for each cluster demonstrated good predictive performance, enabling us to confirm and detail the major environmental characteristics structuring the observed clusters. Further, these models could allow us to build up continuous maps of cluster distribution and predict future changes in a context of climate change.

Bibliography

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models : Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*(6), 1223‑1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Amano, T., Lamming, J. D. L., & Sutherland, W. J. (2016). Spatial Gaps in Global Biodiversity Information and the Role of Citizen Science. *BioScience*, *66*(5), 393‑400. https://doi.org/10.1093/biosci/biw022

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, *22*(1), 42‑47. https://doi.org/10.1016/j.tree.2006.09.010

Archambault, P., & Bourget, E. (1996). Scales of coastal heterogeneity and benthic intertidal species richness, diversity and abundance. Marine Ecology Progress Series, 136, 111‑121. https://doi.org/10.3354/meps136111

Ayata, S.-D., Lazure, P., & Thiébaut, É. (2010). How does the connectivity between populations mediate range limits of marine invertebrates? A case study of larval dispersal between the Bay of Biscay and the English Channel (North-East Atlantic). *Progress in Oceanography*, *87*(1), 18‑36. https://doi.org/10.1016/j.pocean.2010.09.022

Barry, J. P., Baxter, C. H., Sagarin, R. D., & Gilman, S. E. (1995). Climate-Related, Long-Term Faunal Changes in a California Rocky Intertidal Community. *Science*, *267*(5198), 672‑675. https://doi.org/10.1126/science.267.5198.672

Blanchet, G., Legendre, P., Borcard, D., Forward Selection of Explanatory Variables. *Ecological Society of America*, 89 (9), 2623-2632. https://doi.org/10.1890/07-0986.1

Bled, F., Nichols, J. D., & Altwegg, R. (2013). Dynamic occupancy models for analyzing species' range dynamics across large geographic scales. *Ecology and Evolution*, *3*(15), 4896‑4909. https://doi.org/10.1002/ece3.858

Bœuf, G., Allain, Y.-M., & Bouvier, M. (2012). L'apport des sciences participatives à la connaissance de la biodiversité en France. *La Lettre de l'OCIM. Musées, Patrimoine et Culture scientifiques et techniques*, *144*, Article 144. https://doi.org/10.4000/ocim.1119

Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical Ecology with R*. Springer. https://doi.org/10.1007/978-1-4419-7976-6

Bowler, D. E., Bjorkman, A. D., Dornelas, M., Myers-Smith, I. H., Navarro, L. M., Niamir, A., Supp, S. R., Waldock, C., Winter, M., Vellend, M., Blowes, S. A., Böhning-Gaese, K., Bruelheide, H., Elahi, R., Antão, L. H., Hines, J., Isbell, F., Jones, H. P., Magurran, A. E., … Bates, A. E. (2020). Mapping human pressures on biodiversity across the planet uncovers anthropogenic threat complexes. *People and Nature*, *2*(2), 380‑394. https://doi.org/10.1002/pan3.10071

Boyer, K. E., Kertesz, J. S., & Bruno, J. F. (2009). Biodiversity effects on productivity and stability of marine macroalgal communities : The role of environmental context. *Oikos*, *118*(7), 1062‑1072. https://doi.org/10.1111/j.1600-0706.2009.17252.x

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5‑32. https://doi.org/10.1023/A:1010933404324

Brind'Amour, A., Mahévas, S., Legendre, P., & Bellanger, L. (2018). Application of Moran Eigenvector Maps (MEM) to irregular sampling designs. *Spatial Statistics*, *26*, 56‑68. https://doi.org/10.1016/j.spasta.2018.05.004

Brito-Morales, I., Molinos, J. G., Schoeman, D. S., Burrows, M. T., Poloczanska, E. S., Brown, C. J., Ferrier, S., Harwood, T. D., Klein, C. J., McDonald-Madden, E., Moore, P. J., Pandolfi, J. M., Watson, J. E. M., Wenger, A. S., & Richardson, A. J. (2018). Climate Velocity Can Inform Conservation in a Warming World. *Trends in Ecology & Evolution*, *33*(6), 441‑457. https://doi.org/10.1016/j.tree.2018.03.009

Burrows, M. T., Schoeman, D. S., Buckley, L. B., Moore, P., Poloczanska, E. S., Brander, K. M., Brown, C., Bruno, J. F., Duarte, C. M., Halpern, B. S., Holding, J., Kappel, C. V., Kiessling, W., O'Connor, M. I., Pandolfi, J. M., Parmesan, C., Schwing, F. B., Sydeman, W. J., & Richardson, A. J. (2011). The Pace of Shifting Climate in Marine and Terrestrial Ecosystems. *Science*, *334*(6056), 652‑655. https://doi.org/10.1126/science.1210288

Burrows, M. Wave fetch GIS layers for Europe at 100m scale. (2020). [jeu de données]. figshare. https://doi.org/10.6084/m9.figshare.8668127.v1

Burrows, M. T., Hawkins, S. J., Moore, J. J., Adams, L., Sugden, H., Firth, L., & Mieszkowska, N. (2020). Global-scale species distributions predict temperature-related changes in species composition of rocky shore communities in Britain. *Global Change Biology*, *26*(4), 2093‑2105. https://doi.org/10.1111/gcb.14968

Cáceres, M. D., & Legendre, P. (2009). Associations between species and groups of sites : Indices and statistical inference. *Ecology*, *90*(12), 3566‑3574. https://doi.org/10.1890/08-1823.1

Čengić, M., Rost, J., Remenska, D., Janse, J. H., Huijbregts, M. A. J., & Schipper, A. M. (2020). On the importance of predictor choice, modelling technique, and number of pseudo-absences for bioclimatic envelope model performance. *Ecology and Evolution*, *10*(21), 12307‑12317. https://doi.org/10.1002/ece3.6859

Chauvaud, P. (2022). Semer l'ormeau dans le milieu naturel : Rôles du milieu et des caractéristiques biologiques d'Haliotis tuberculata dans la réussite du repeuplement [Phd thesis, Université de Bretagne occidentale - Brest]. https://theses.hal.science/tel-04112034 « En ligne ».

Christopher D. G. Harley, S. L. W. (2006). The impacts of climate change in coastal marine systems. *Ecology Letters 9(*2), 228-241. https://doi.org/10.1111/j.1461-0248.2005.00871.x

Chust, G., Villarino, E., McLean, M., Mieszkowska, N., Benedetti-Cecchi, L., Bulleri, F., Ravaglioli, C., Borja, A., Muxika, I., Fernandes-Salvador, J. A., Ibaibarriaga, L., Uriarte, A., Revilla, M., Villate, F., Iriarte, A., Uriarte, I., Zervoudaki, S., Carstensen, J., Somerfield, P. J., … Lindegren, M. (2024). Cross-basin and cross-taxa patterns of marine community tropicalization and deborealization in warming European seas. *Nature Communications*, *15*(1), 2126. https://doi.org/10.1038/s41467-024-46526-y

Clappe, S., Dray, S., & Peres-Neto, P. R. (2018). Beyond neutrality : Disentangling the effects of species sorting and spurious correlations in community analysis. *Ecology*, *99*(8), 1737‑1747. https://doi.org/10.1002/ecy.2376

Contrier, P. (1991). The Arpege project at Meteo-France. Proc. of ECMWF seminar on Numerical Methods in Atmospheric Models. 9-13 September 1991. https://cir.nii.ac.jp/crid/1572543023982199168 « En ligne ».

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR : An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938‑2940. https://doi.org/10.1093/bioinformatics/btx364

Cooper, C. B., Dickinson, J., Phillips, T., & Bonney, R. (2007). Citizen Science as a Tool for Conservation in Residential Ecosystems. *Ecology and Society*, *12*(2). https://www.jstor.org/stable/26267884 « En ligne ».

Courtier, P., C. Freydier, J.-F.Geleyn, F. Rabier and M. Rochas (1994). - The ARPEGE project at Météo-France. in ECMWF Seminar Proceedings: Numerical methods in atmospheric models, II:193-231. https://cir.nii.ac.jp/crid/1572543023982199168 « En ligne ».

Dalrymple, R. W., Mackay, D. A., Ichaso, A. A., & Choi, K. S. (2012). Processes, Morphodynamics, and Facies of Tide-Dominated Estuaries. In R. A. Davis Jr. & R. W. Dalrymple (Éds.), *Principles of Tidal Sedimentology* (p. 79‑107). Springer Netherlands. https://doi.org/10.1007/978-94-007-0123-6_5

D'Amen, M., Rahbek, C., Zimmermann, N. E., & Guisan, A. (2017). Spatial predictions at the community level : From current approaches to future frameworks. *Biological Reviews*, *92*(1), 169‑187. https://doi.org/10.1111/brv.12222

Dauvin, J.-C. (2015). History of benthic research in the English Channel : From general patterns of communities to habitat mosaic description. *Journal of Sea Research*, *100*, 32‑45. https://doi.org/10.1016/j.seares.2014.11.005

De Caceres, M., Jansen, F., & De Caceres, M. M. (2016). Package 'indicspecies'. Indicators, 8(1). ftp://r-project.org/pub/R/web/packages/indicspecies/indicspecies.pdf « En ligne ».

Derrien-Courtel, S., Grall, J., Thiebaut, E., Houbin, C., Gentil, F., Ar Gall, E., Hily, C., Hassini, S., (2010). Faune et Flore benthiques du littoral breton. Listes d'habitats déterminants pour la réalisation des fiches ZNIEFF-Mer proposées en CSRPN Bretagne le 26 novembre 2015, 28. https://cotedegranitrose-septiles.n2000.fr/sites/cotedegranitrose-septiles.n2000.fr/files/2023-10/Liste_znieff_%20habitats_d%C3%A9terminants_marins_benthiques2015.pdf « En ligne ».

Deschamps, G., Poggiato, G., Brun, P., Galiez, C., & Thuiller, W. (2023). Predict first–assemble later versus assemble first–predict later : Revisiting the dilemma for functional biogeography. *Methods in Ecology and Evolution*, *14*(10), 2680‑2696. https://doi.org/10.1111/2041-210X.14203

Devictor, V., Whittaker, R. J., & Beltrame, C. (2010). Beyond scarcity : Citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions*, *16*(3), 354‑362. https://doi.org/10.1111/j.1472-4642.2009.00615.x

Dickinson, J. L., & Bonney, R. (2012). Citizen Science : Public Participation in Environmental Research. *Cornell University Press*. https://www.degruyter.com/document/doi/10.7591/9780801463952/html « En ligne ».

Dinter, W. P. (2001). Biogeography of the OSPAR maritime area. Bonn, Germany: Federal Agency for Nature Conservation, 167. https://www.vliz.be/imisdocs/publications/ocrd/259329.pdf « En ligne ».

Dohet, A., Dolisy, D., Hoffmann, L., & Dufrêne, M. (2002). Identification of bioindicator species among Ephemeroptera, Plecoptera and Trichoptera in a survey of streams belonging to the rhithral classification in the Grand Duchy of Luxembourg. *Internationale Vereinigung Für Theoretische Und Angewandte Limnologie: Verhandlungen*. https://www.tandfonline.com/doi/abs/10.1080/03680770.2001.11902609

Données issues du modèle *ECO-MARS3D*. (s. d.). Sextant. https://sextant.ifremer.fr/geonetwork/srv/api/records/7ec6aa6e-2406-42d1-a798-5aa2984e6311 « En ligne ».

Dormann, C. F., Bobrowski, M., Dehling, D. M., Harris, D. J., Hartig, F., Lischke, H., Moretti, M. D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S. I., Sheppard, C. S., Steinbauer, M. J., Zeuss, D., & Kraan, C. (2018). Biotic interactions in species distribution modelling : 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, *27*(9), 1004‑1016. https://doi.org/10.1111/geb.12759

Dray, S., Pélissier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P. R., Bellier, E., Bivand, R., Blanchet, F. G., De Cáceres, M., Dufour, A.-B., Heegaard, E., Jombart, T., Munoz, F., Oksanen, J., Thioulouse, J., & Wagner, H. H. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, *82*(3), 257‑275. https://doi.org/10.1890/11-1183.1

Droual, G., & Bridier, G. (2020). One-step beyond : Northernmost record of Berghia verrucicornis (Costa, 1867). *An Aod - Les cahiers naturalistes de l'Observatoire marin*. https://hal.science/hal-02986674 « En ligne ».

Duffy, J. E., Godwin, C. M., & Cardinale, B. J. (2017). Biodiversity effects in the wild are common and as strong as key drivers of productivity. *Nature*, *549*(7671), Article 7671. https://doi.org/10.1038/nature23886

Dufrêne, M., & Legendre, P. (1997). Species Assemblages and Indicator Species:the Need for a Flexible Asymmetrical Approach. *Ecological Monographs*, *67*(3), 345‑366. https://doi.org/10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2

Dukes, J. S., Mooney, H. A., Dukes, J. S., Mooney, H. A., Dukes, J. S., & Mooney, H. A. (1999). Does global change increase the success of biological invaders? *Trends in Ecology & Evolution*, *14*(4), 135‑139. https://doi.org/10.1016/S0169-5347(98)01554-7

Edgar, G. J., Bates, A. E., Bird, T. J., Jones, A. H., Kininmonth, S., Stuart-Smith, R. D., & Webb, T. J. (2016). New Approaches to Marine Conservation Through the Scaling Up of Ecological Data. *Annual Review of Marine Science*, *8*(8), 435‑461. https://doi.org/10.1146/annurev-marine-122414-033921

Egbert, G. D., Erofeeva, S. Y., & Ray, R. D. (2010). Assimilation of altimetry data for nonlinear shallow-water tides : Quarter-diurnal tides of the Northwest European Shelf. *Continental Shelf Research*, *30*(6), 668‑679. https://doi.org/10.1016/j.csr.2009.10.011

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models : Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677‑697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Ellingsen, Karie., & Gray, J. s. (2002). Spatial patterns of benthic diversity : Is there a latitudinal gradient along the Norwegian continental shelf? *Journal of Animal Ecology*, *71*(3), 373‑389. https://doi.org/10.1046/j.1365-2656.2002.00606.x

Feebarani, J., Joydas, T. V., Damodaran, R., & Borja, A. (2016). Benthic quality assessment in a naturally- and human-stressed tropical estuary. *Ecological Indicators*, *67*, 380‑390. https://doi.org/10.1016/j.ecolind.2016.03.005

Feldman, M. J., Imbeau, L., Marchand, P., Mazerolle, M. J., Darveau, M., & Fenton, N. J. (2021). Trends and gaps in the use of citizen science derived data as input for species distribution models : A quantitative review. *PLOS ONE*, *16*(3), e0234587. https://doi.org/10.1371/journal.pone.0234587

Ferrier, S., & Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, *43*(3), 393‑404. https://doi.org/10.1111/j.1365-2664.2006.01149.x

Fontaine, A., Simard, A., Brunet, N., & Elliott, K. H. (2022). Scientific contributions of citizen science applied to rare or threatened animals. *Conservation Biology*, *36*(6), e13976. https://doi.org/10.1111/cobi.13976

Fragkopoulou, E., Serrão, E. A., De Clerck, O., Costello, M. J., Araújo, M. B., Duarte, C. M., Krause-Jensen, D., & Assis, J. (2022). Global biodiversity patterns of marine forests of brown macroalgae. *Global Ecology and Biogeography*, *31*(4), 636‑648. https://doi.org/10.1111/geb.13450

François Gaudin, N. D. (2018). Marine sublittoral benthos fails to track temperature in response to climate change in a biogeographical transition zone | ICES Journal of Marine Science | Oxford Academic. *Ices journal of marine science*. https://doi.org/10.1093/icesjms/fsy095

Friedman, J. H., & Popescu, B. E. (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics*, *2*(3), 916‑954. https://doi.org/10.1214/07-AOAS148

Gallon, R. K., Lavesque, N., Grall, J., Labrune, C., Gremare, A., Bachelet, G., Blanchet, H., Bonifácio, P., Bouchet, V. M. P., Dauvin, J.-C., Desroy, N., Gentil, F., Guerin, L., Houbin, C., Jourde, J., Laurand, S., Le Duff, M., Le Garrec, V., de Montaudouin, X., … Gauthier, O. (2017). Regional and latitudinal patterns of soft-bottom macrobenthic invertebrates along French coasts : Results from the RESOMAR database. *Journal of Sea Research*, *130*, 96‑106. https://doi.org/10.1016/j.seares.2017.03.011

Ge, Y., Cheng, Q., & Zhang, S. (2005). Reduction of edge effects in spatial information extraction from regional geochemical data : A case study based on multifractal filtering technique. *Computers & Geosciences*, *31*(5), 545‑554. https://doi.org/10.1016/j.cageo.2004.11.005

Grall, J., Serre-Arnoldy, D., Serre, S., & Quillien, N. (2015). Première signalisation du nudibranche aeolidien Spurilla neapolitana en mer d'Iroise (Bretagne ouest). *An aod-les cahiers naturalistes de l'Observatoire marin*, 4 (2),29-33. https://www-iuem.univ-brest.fr/observatoire/l-observatoire/ressources/cahiers-naturalistes/copy_of_AnAod_2015_IV_2_pp_29_33.pdf « En ligne ».

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, *24*(3), 276‑292. https://doi.org/10.1111/geb.12268

Guisan, A., & Thuiller, W. (2005). Predicting species distribution : Offering more than simple habitat models. *Ecology Letters*. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Halpern, B. S., Frazier, M., Afflerbach, J., Lowndes, J. S., Micheli, F., O'Hara, C., Scarborough, C., & Selkoe, K. A. (2019). Recent pace of change in human impact on the world's ocean. *Scientific Reports*, *9*(1), 11609. https://doi.org/10.1038/s41598-019-47201-9

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., Bruno, J. F., Casey, K. S., Ebert, C., Fox, H. E., Fujita, R., Heinemann, D., Lenihan, H. S., Madin, E. M. P., Perry, M. T., Selig, E. R., Spalding, M., Steneck, R., & Watson, R. (2008). A Global Map of Human Impact on Marine Ecosystems. *Science*, *319*(5865), 948‑952. https://doi.org/10.1126/science.1149345

Hammond, W., & Griffiths, C. L. (2004). Influence of wave exposure on South African mussel beds and their associated infaunal communities. *Marine Biology*, *144*(3), 547‑552. https://doi.org/10.1007/s00227-003-1210-9

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29‑36. https://doi.org/10.1148/radiology.143.1.7063747

Hao, T., Elith, J., Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, *25*(5), 839‑852. https://doi.org/10.1111/ddi.12892

Hartigan, J. A., & Wong, M. A. (1979). A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *28*(1), 100‑108. https://doi.org/10.2307/2346830

Hastie, T. J. (1992). Generalized Additive Models. In *Statistical Models in S*. Routledge. https://www.taylorfrancis.com/chapters/edit/10.1201/9780203738535-7/generalized-additive-models-trevor-hastie « En ligne ».

Hastie, T. J., & Pregibon, D. (1992). Generalized Linear Models. In *Statistical Models in S*. Routledge.https://www.taylorfrancis.com/chapters/edit/10.1201/9780203738535-6/generalized-linear-models-trevor-hastie-daryl-pregibon « En ligne ».

Hawkins, S. J., Sugden, H. E., Mieszkowska, N., Moore, P. J., Poloczanska, E., Leaper, R., Herbert, R. J. H., Genner, M. J., Moschella, P. S., Thompson, R. C., Jenkins, S. R., Southward, A. J., & Burrows, M. T. (2009). Consequences of climate-driven biodiversity changes for ecosystem functioning of North European rocky shores. *Marine Ecology Progress Series*, *396*, 245‑259. https://doi.org/10.3354/meps08378

Helmuth, B., Broitman, B. R., Blanchette, C. A., Gilman, S., Halpin, P., Harley, C. D. G., O'Donnell, M. J., Hofmann, G. E., Menge, B., & Strickland, D. (2006). Mosaic Patterns of Thermal Stress in the Rocky Intertidal Zone : Implications for Climate Change. *Ecological Monographs*, *76*(4), 461‑479. https://doi.org/10.1890/0012-9615(2006)076[0461:MPOTSI]2.0.CO;2

Hiddink, J. G., Burrows, M. T., & García Molinos, J. (2015). Temperature tracking by North Sea benthic invertebrates in response to climate change. *Global Change Biology*, *21*(1), 117‑129. https://doi.org/10.1111/gcb.12726

Hinz, H., Capasso, E., Lilley, M., Frost, M., & Jenkins, S. R. (2011). Temporal differences across a bio-geographical boundary reveal slow response of sub-littoral benthos to climate change. *Marine Ecology Progress Series*, *423*, 69‑82. https://doi.org/10.3354/meps08963

Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E. K., Hungate, B. A., Matulich, K. L., Gonzalez, A., Duffy, J. E., Gamfeldt, L., & O'Connor, M. I. (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, *486*(7401), 105‑108. https://doi.org/10.1038/nature11118

Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, *22*, 415‑427. https://doi.org/10.1101/SQB.1957.022.01.039

Ingle, R. W., & Clark, P. F. (2008). First reported occurrences of the marbled crab, Pachygrapsus marmoratus (Crustacea : Brachyura: Grapsoidea) in southern coastal waters of the British Isles. *Marine Biodiversity Records,* 1, 26. https://doi.org/10.1017/S1755267206002454

Iverson, L. R., Prasad, A. M., Matthews, S. N., & Peters, M. (2008). Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecology and Management*, *254*(3), 390‑406. https://doi.org/10.1016/j.foreco.2007.07.023

Jolly, M. T., Jollivet, D., Gentil, F., Thiébaut, E., & Viard, F. (2005). Sharp genetic break between Atlantic and English Channel populations of the polychaete Pectinaria koreni, along the North coast of France. *Heredity*, 94(1), 23‑32. https://doi.org/10.1038/sj.hdy.6800543

Jolly, M. T., Guyard, P., Ellien, C., Gentil, F., Viard, F., Thiébaut, E., & Jollivet, D. (2009). Population genetics and hydrodynamic modeling of larval dispersal dissociate contemporary patterns of connectivity from historical expansion into European shelf seas in the polychaete Pectinaria koreni. *Limnology and Oceanography*, *54*(6), 2089‑2106. https://doi.org/10.4319/lo.2009.54.6.2089

Jourde, J., Alizier, S., Dancie, C., Dauvin, J.-C., Desroy, N., Dubut, S., Gentil, F., Grall, J., Hanin, C., Lanshere, J., & Thiébaut, É. (2012). First and repeated records of the tropical-temperate crab Asthenognathus atlanticus Monod, 1932 (Decapoda : Brachyura) in the eastern part of the Bay of Seine (eastern English Channel, France). *Cahiers de Biologie Marine*. https://hal.science/hal-01245433

Kark, S., & van Rensburg, B. J. (2006). Ecotones : Marginal or Central Areas of Transition? *Israel Journal of Ecology & Evolution*, *52*(1), 29‑53. https://doi.org/10.1560/IJEE.52.1.29

Kark, S. (2017). Effects of ecotones on biodiversity: reference module in life sciences. https://enviro2.doe.gov.my/ekmc/wp-content/uploads/2016/08/1385472293-3-s2.0-B9780122268656005730-main.pdf « En ligne ».

Kelling, S., Johnston, A., Hochachka, W. M., Iliff, M., Fink, D., Gerbracht, J., Lagoze, C., Sorte, F. A. L., Moore, T., Wiggins, A., Wong, W.-K., Wood, C., & Yu, J. (2015). Can Observation Skills of Citizen Scientists Be Estimated Using Species Accumulation Curves? *PLOS ONE*, *10*(10), e0139600. https://doi.org/10.1371/journal.pone.0139600

Kerswell, A. P. (2006). Global Biodiversity Patterns of Benthic Marine Algae. *Ecology*, 87(10), 2479‑2488. https://doi.org/10.1890/0012-9658(2006)87[2479:GBPOBM]2.0.CO;2

Kleitou, P., Giovos, I., Wolf, W., & Crocetta, F. (2019). On the importance of citizen-science : The first record of Goniobranchus obsoletus (Rüppell & Leuckart, 1830) from Cyprus (Mollusca: Gastropoda: Nudibranchia). https://doi.org/10.3391/bir.2019.8.2.06

Kléparski, L., Beaugrand, G., & Edwards, M. (2021). Plankton biogeography in the North Atlantic Ocean and its adjacent seas : Species assemblages and environmental signatures. *Ecology and Evolution*, *11*(10), 5135‑5149. https://doi.org/10.1002/ece3.7406

Kröncke, I., Reiss, H., Eggleton, J. D., Aldridge, J., Bergman, M. J. N., Cochrane, S., Craeymeersch, J. A., Degraer, S., Desroy, N., Dewarumez, J.-M., Duineveld, G. C. A., Essink, K., Hillewaert, H., Lavaleye, M. S. S., Moll, A., Nehring, S., Newell, R., Oug, E., Pohlmann, T., … Rees, H. L.

(2011). Changes in North Sea macrofauna communities and species distribution between 1986 and 2000. *Estuarine, Coastal and Shelf Science*, *94*(1), 1‑15. https://doi.org/10.1016/j.ecss.2011.04.008

Lai, J., Zou, Y., Zhang, J., & Peres-Neto, P. R. (2022). Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package. *Methods in Ecology and Evolution*, *13*(4), 782‑788. https://doi.org/10.1111/2041-210X.13800

Lathlean, J. A., McWilliam, R. A., Ayre, D. J., & Minchinton, T. E. (2015). Biogeographical patterns of rocky shore community structure in south-east Australia : Effects of oceanographic conditions and heat stress. *Journal of Biogeography*, *42*(8), 1538‑1552. https://doi.org/10.1111/jbi.12511

Lazure, P., & Dumas, F. (2008). An external–internal mode coupling for a 3D hydrodynamical model for applications at regional scale (MARS). *Advances in Water Resources*, *31*(2), 233‑250. https://doi.org/10.1016/j.advwatres.2007.06.010

Le Boyer, A., Cambon, G., Daniault, N., Herbette, S., Le Cann, B., Marié, L., & Morin, P. (2009). Observations of the Ushant tidal front in September 2007. *Continental Shelf Research*, *29*(8), 1026‑1037. https://doi.org/10.1016/j.csr.2008.12.020

Le Loc'h, F., Hily, C., & Grall, J. (2008). Benthic community and food web structure on the continental shelf of the Bay of Biscay (North Eastern Atlantic) revealed by stable isotopes analysis. *Journal of Marine Systems*, *72*(1), 17‑34. https://doi.org/10.1016/j.jmarsys.2007.05.011

Legendre, P., & Borcard, D. (2018). Box–Cox‑chord transformations for community composition data prior to beta diversity analysis. *Ecography*, *41*(11), 1820‑1824. https://doi.org/10.1111/ecog.03498

Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, *129*(2), 271‑280. https://doi.org/10.1007/s004420100716

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet : Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1983‑1992. https://doi.org/10.1109/TVCG.2014.2346248

Li, X., & Wang, Y. (2013). Applying various algorithms for species distribution modelling. *Integrative Zoology*, *8*(2), 124‑135. https://doi.org/10.1111/1749-4877.12000

Lotze, H. K., Lenihan, H. S., Bourque, B. J., Bradbury, R. H., Cooke, R. G., Kay, M. C., Kidwell, S. M., Kirby, M. X., Peterson, C. H., & Jackson, J. B. C. (2006). Depletion, Degradation, and Recovery Potential of Estuaries and Coastal Seas. *Science*, *312*(5781), 1806‑1809. https://doi.org/10.1126/science.1128035

Lynch, S. A., Darmody, G., O'Dwyer, K., Gallagher, M. C., Nolan, S., McAllen, R., & Culloty, S. C. (2016). Biology of the invasive ascidian Ascidiella aspersa in its native habitat : Reproductive patterns and parasite load. *Estuarine, Coastal and Shelf Science*, 181, 249‑255. https://doi.org/10.1016/j.ecss.2016.08.048

Mackechnie, C., Maskell, L., Norton, L., & Roy, D. (2011). The role of 'Big Society' in monitoring the state of the natural environment. *Journal of Environmental Monitoring*, *13*(10), 2687‑2691. https://doi.org/10.1039/C1EM10615E

Malone, T. C., DiGiacomo, P. M., Gonçalves, E., Knap, A. H., Talaue-McManus, L., de Mora, S., & Muelbert, J. (2014). Enhancing the Global Ocean Observing System to meet evidence based needs for the ecosystem-based management of coastal ecosystem services. *Natural Resources Forum*, *38*(3), 168‑181. https://doi.org/10.1111/1477-8947.12045

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, *15*(1), 59‑69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

McClintock, B. T., Bailey, L. L., Pollock, K. H., & Simons, T. R. (2010). Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, *91*(8), 2446‑2454. https://doi.org/10.1890/09-1287.1

Meißner, K., Fiorentino, D., Schnurr, S., Martinez Arbizu, P., Huettmann, F., Holst, S., Brix, S., & Svavarsson, J. (2014). Distribution of benthic marine invertebrates at northern latitudes ― An evaluation applying multi-algorithm species distribution models. *Journal of Sea Research*, *85*, 241‑254. https://doi.org/10.1016/j.seares.2013.05.007

Melo-Merino, S. M., Reyes-Bonilla, H., & Lira-Noriega, A. (2020). Ecological niche models and species distribution models in marine environments : A literature review and spatial analysis of evidence. *Ecological Modelling*, *415*, 108837. https://doi.org/10.1016/j.ecolmodel.2019.108837

Mieszkowska, N., Burrows, M. T., Hawkins, S. J., & Sugden, H. (2021). Impacts of Pervasive Climate Change and Extreme Events on Rocky Intertidal Communities : Evidence From Long-Term Data. *Frontiers in Marine Science*, *8*. https://doi.org/10.3389/fmars.2021.642764

Monteiro, M. C., Pereira, L. C. C., & Jiménez, J. A. (2016). The Trophic Status of an Amazonian Estuary Under Anthropogenic Pressure (Brazil). *Journal of Coastal Research*, *75 (10075)*, 98‑102. https://doi.org/10.2112/SI75-020.1

Morim, J., Trenham, C., Hemer, M., Wang, X. L., Mori, N., Casas-Prat, M., Semedo, A., Shimura, T., Timmermans, B., Camus, P., Bricheno, L., Mentaschi, L., Dobrynin, M., Feng, Y., & Erikson, L. (2020). A global ensemble of ocean wave climate projections from CMIP5-driven models. *Scientific Data*, *7*(1), 105. https://doi.org/10.1038/s41597-020-0446-2

Muths, D., Jollivet, D., Gentil, F., & Davoult, D. (2009). Large-scale genetic patchiness among NE Atlantic populations of the brittle star Ophiothrix fragilis. *Aquatic Biology*, 5, 117‑132. https://doi.org/10.3354/ab00138

Myers, A. A., & Giller, P. S. (1988). Process, pattern and scale in biogeography. *Analytical Biogeography : An Integrated Approach to the Study of Animal and Plant Distributions* (p. 3‑12). Springer Netherlands. https://doi.org/10.1007/978-94-009-1199-4_1

Mieszkowska, N., Hawkins, S. J., Burrows, M. T., & Kendall, M. A. (2007). Long-term changes in the geographic distribution and population structures of Osilinus lineatus (Gastropoda : Trochidae) in Britain and Ireland. Journal of the Marine Biological Association of the United Kingdom, 87(2), 537‑545. https://doi.org/10.1017/S0025315407053799

Newman, G., Chandler, M., Clyde, M., McGreavy, B., Haklay, M., Ballard, H., Gray, S., Scarpino, R., Hauptfeld, R., Mellor, D., & Gallo, J. (2017). Leveraging the power of place in citizen science for effective conservation decision making. *Biological Conservation*, *208*, 55‑64. https://doi.org/10.1016/j.biocon.2016.07.019

Patterson, C., Slater, M., Early, R., & Laing, C. (2020). The status of Clibanarius erythropus after a recent range expansion to Great Britain, with the highest latitude recording of a gravid individual. *Marine Biodiversity Records,* 13(1), 3. https://doi.org/10.1186/s41200-020-00186-1

Penard, C. (2009). Détection satellitaire et modélisation opérationnelle de la production végétale non-fixée dans la bande côtière bretonne, Remote sensing and operationnal modelling of non-fixed algal production on the French Brittany coastal shelf [Université de Bretagne Occidentale]. https://archimer.ifremer.fr/doc/00015/12603/ « En ligne ».

Philippart, C. J. M., van Aken, H. M., Beukema, J. J., Bos, O. G., Cadée, G. C., & Dekker, R. (2003). Climate-related changes in recruitment of the bivalve Macoma balthica. *Limnology and Oceanography*, *48*(6), 2171‑2185. https://doi.org/10.4319/lo.2003.48.6.2171

Picton, B.E., Morrow, C.C. & van Soest, R.W.B. (2011). Sponges of Britain and Ireland. https://www.habitas.org.uk/marinelife/sponge_guide/sponges.asp?item=C4250 « En ligne ».

Pingree, R. D. (1988). Internal Tidal Oscillations and Water Column Instability in the Upper Slope Region of the Bay of Biscay. In J. C. J. Nihoul & B. M. Jamart (Éds.), *Elsevier Oceanography* 46, 387‑ 404). Elsevier. https://doi.org/10.1016/S0422-9894(08)70560-9

Ramírez, F., Afán, I., Davis, L. S., & Chiaradia, A. (2017). Climate impacts on global hot spots of marine biodiversity. *Science Advances*, *3*(2). https://doi.org/10.1126/sciadv.1601198

Reise, K., & van Beusekom, J. E. E. (2008). Interactive effects of global and regional change on a coastal ecosystem. *Helgoland Marine Research*, *62*(1), 85‑91. https://doi.org/10.1007/s10152-007-0102-7

Salman, A., Lombardo, S., & Doody, P. (2004). Living with coastal erosion in Europe : Sediment and Space for Sustainability. *Eurosion Project Reports*. https://repository.tudelft.nl/islandora/object/uuid%3A483327a3-dcf7-4bd0-a986-21d9c8ec274e « En ligne ».

Schlund, E., Basuyaux, O., Lecornu, B., Pezy, J.-P., Baffreau, A., & Dauvin, J.-C. (2016). Macrofauna associated with temporary Sabellaria alveolata reefs on the west coast of Cotentin (France). *SpringerPlus*, *5*(1), 1260. https://doi.org/10.1186/s40064-016-2885-y

Schmeller, D. S., Henry, P.-Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F., Lengyel, S., Nowicki, P., Déri, E., Budrys, E., Kull, T., Tali, K., Bauch, B., Settele, J., Van Swaay, C., Kobler, A., Babij, V., Papastergiadou, E., & Henle, K. (2009). Advantages of Volunteer-Based Biodiversity Monitoring in Europe. *Conservation Biology*, *23*(2), 307-316. https://doi.org/10.1111/j.1523-1739.2008.01125.x

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., & Masson, V. (2011). The AROME-France Convective-Scale Operational Model. *Monthly Weather Review*, *139*(3), 976-991. https://doi.org/10.1175/2010MWR3425.1

Seni, G., & Elder, J. F. (2010). Model Complexity, Model Selection and Regularization. In G. Seni & J. F. Elder (Éds.), *Ensemble Methods in Data Mining : Improving Accuracy Through Combining Predictions* (p. 28-52). Springer International Publishing. https://doi.org/10.1007/978-3-031-01899-2_3

Silvertown, J., Buesching, C. D., Jacobson, S. K., & Rebelo, T. (2013). Citizen science and nature conservation. In *Key Topics in Conservation Biology 2* (p. 127-142). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118520178.ch8

Southward, A. J., Hawkins, S. J., & Burrows, M. T. (1995). Seventy years' observations of changes in distribution and abundance of zooplankton and intertidal organisms in the western English Channel in relation to rising sea temperature. *Journal of Thermal Biology*, 20(1), 127-155. https://doi.org/10.1016/0306-4565(94)00043-I

Spalding, M. D., Fox, H. E., Allen, G. R., Davidson, N., Ferdaña, Z. A., Finlayson, M., Halpern, B. S., Jorge, M. A., Lombana, A., Lourie, S. A., Martin, K. D., McManus, E., Molnar, J., Recchia, C. A., & Robertson, J. (2007). Marine Ecoregions of the World : A Bioregionalization of Coastal and Shelf Areas. *BioScience*, *57*(7), 573-583. https://doi.org/10.1641/B570707

Stockwell, S., & Gallo, S. (2017). Citizen Science and Wildlife Conservation : Lessons from 34 Years of the Maine Loon Count. *Maine Policy Review*, *26*(2), 25-32. https://doi.org/10.53558/XECR1627

Tempera, F., Vasquez, M., & Caillaud, M. (2019). Brittany's infralittoral seabed : An objective partitioning into marine ecological units. https://archimer.ifremer.fr/doc/00586/69762/ « En ligne ».

Thompson, R. C., Crowe, T. P., & Hawkins, S. J. (2002). Rocky intertidal communities : Past environmental changes, present status and predictions for the next 25 years. *Environmental Conservation*, *29*(2), 168-191. https://doi.org/10.1017/S0376892902000115

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, *32*(3), 369-373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Thuiller, W., Georges, D., Engler, R., Breiner, F., Georges, M. D., & Thuiller, C. W. (2016). Package 'biomod2'. ftp://137.208.57.37/pub/R/web/packages/biomod2/biomod2.pdf « En ligne ».

Travis, J. M. J. (2003). Climate change and habitat destruction : A deadly anthropogenic cocktail. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1514), 467-473. https://doi.org/10.1098/rspb.2002.2246

Van den Hoek, C. (1975). Phytogeographic provinces along the coasts of the northern Atlantic Ocean. *Phycologia*, *14*(4), 317-330. https://doi.org/10.2216/i0031-8884-14-4-317.1

Van der Zee, E. M., Tielens, E., Holthuijsen, S., Donadi, S., Eriksson, B. K., van der Veer, H. W., Piersma, T., Olff, H., & van der Heide, T. (2015). Habitat modification drives benthic trophic diversity in an intertidal soft-bottom ecosystem. *Journal of Experimental Marine Biology and Ecology*, *465*, 41‑48. https://doi.org/10.1016/j.jembe.2015.01.001

Vasconcelos, R. P., Reis-Santos, P., Fonseca, V., Maia, A., Ruano, M., França, S., Vinagre, C., Costa, M. J., & Cabral, H. (2007). Assessing anthropogenic pressures on estuarine fish nurseries along the Portuguese coast : A multi-metric index and conceptual approach. *Science of The Total Environment*, *374*(2), 199‑215. https://doi.org/10.1016/j.scitotenv.2006.12.048

Walther, G.-R. (2010). Community and ecosystem responses to recent climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1549), 2019‑2024. https://doi.org/10.1098/rstb.2010.0021

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J.-M., Hoegh-Guldberg, O., & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, *416*(6879), Article 6879. https://doi.org/10.1038/416389a

Walther, G.-R., Roques, A., Hulme, P. E., Sykes, M. T., Pyšek, P., Kühn, I., Zobel, M., Bacher, S., Botta-Dukát, Z., Bugmann, H., Czúcz, B., Dauber, J., Hickler, T., Jarošík, V., Kenis, M., Klotz, S., Minchin, D., Moora, M., Nentwig, W., … Settele, J. (2009). Alien species in a warmer world : Risks and opportunities. *Trends in Ecology & Evolution*, *24*(12), 686‑693. https://doi.org/10.1016/j.tree.2009.06.008

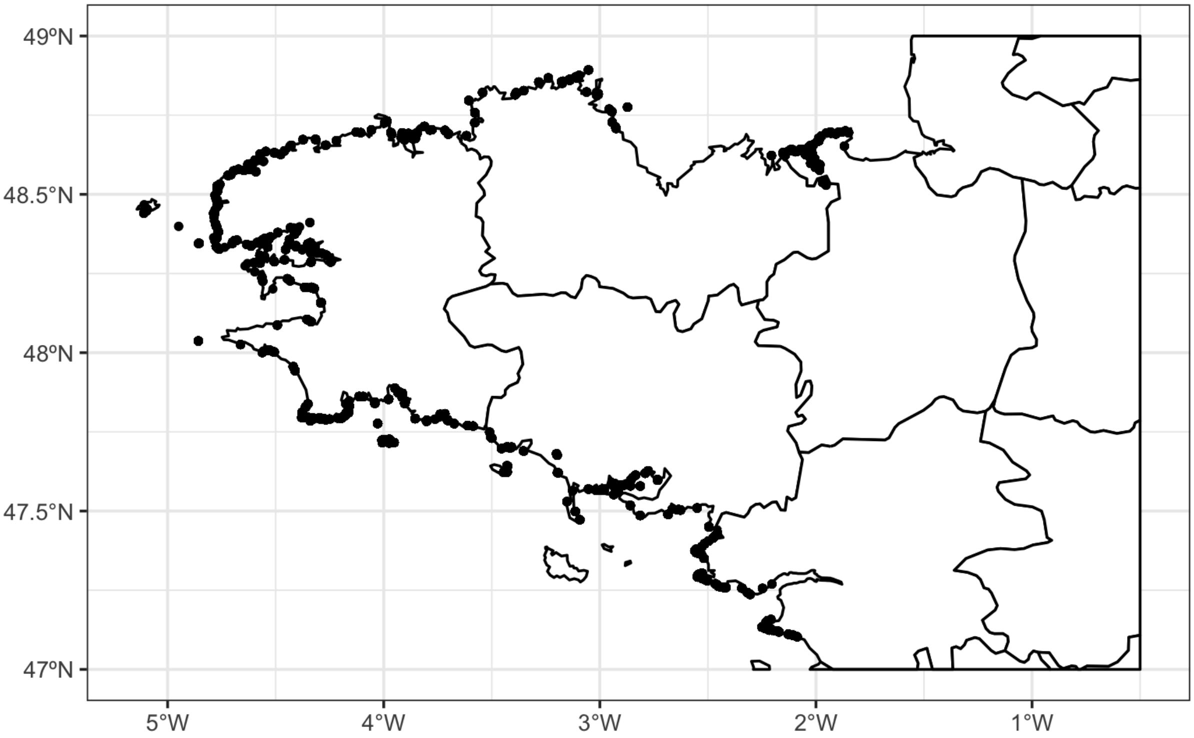
Wethey, D. S., & Woodin, S. A. (2008). Ecological hindcasting of biogeographic responses to climate change in the European intertidal zone. In J. Davenport, G. M. Burnell, T. Cross, M. Emmerson, R. McAllen, R. Ramsay, & E. Rogan (Éds.), *Challenges to Marine Ecosystems* (p. 139‑151). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8808-7_13

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., … Sequeira, A. M. M. (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution*, *33*(10), 790‑802. https://doi.org/10.1016/j.tree.2018.08.001

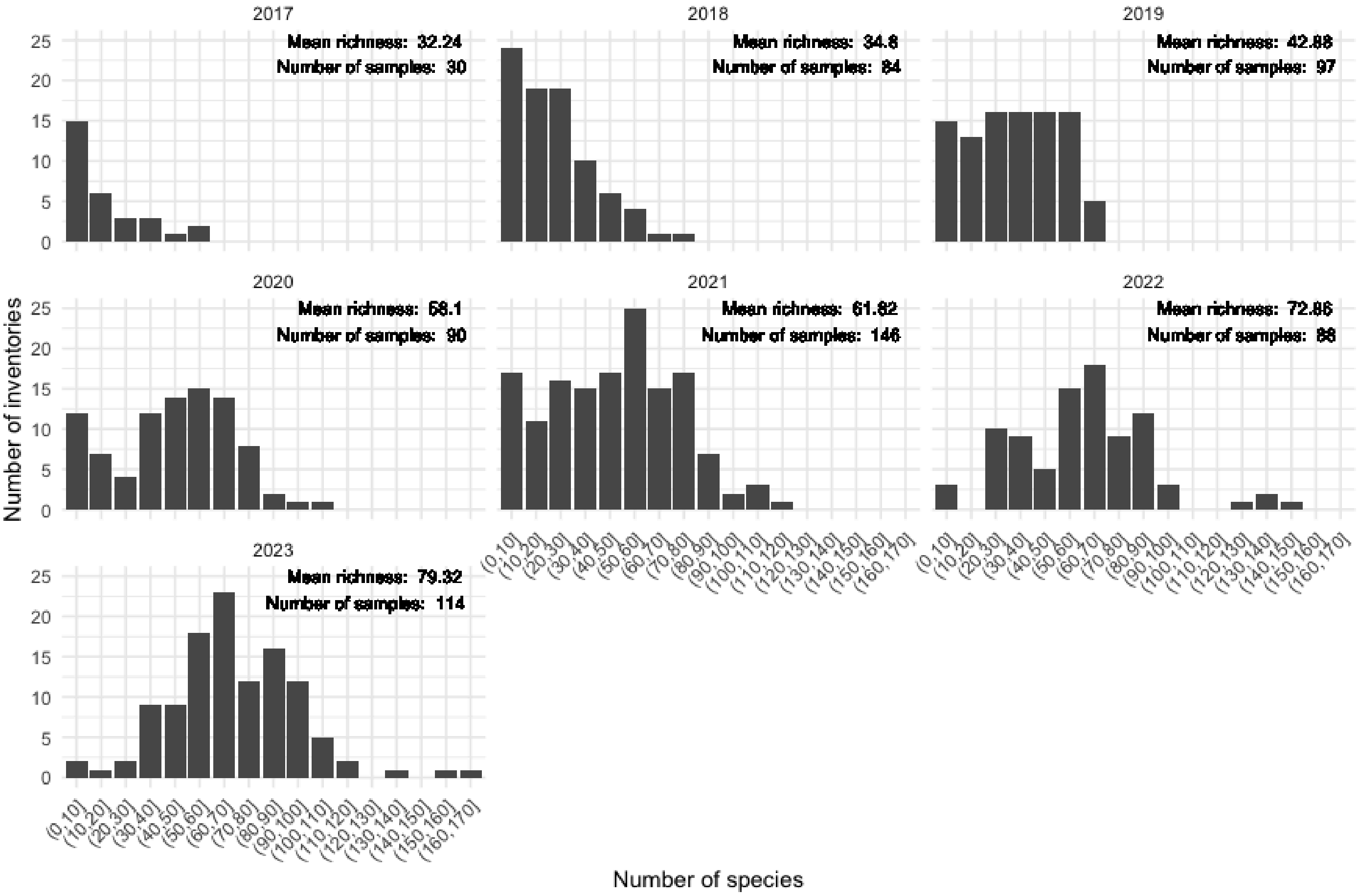
Yee, T. W., & Mitchell, N. D. (1991). Generalized additive models in plant ecology. *Journal of Vegetation Science*, *2*(5), 587‑602. https://doi.org/10.2307/3236170

Yu, J., Wong, W.-K., & Hutchinson, R. A. (2010). Modeling Experts and Novices in Citizen Science Data for Species Distribution Modeling. *2010 IEEE International Conference on Data Mining*, 1157‑1162. https://doi.org/10.1109/ICDM.2010.103

Appendices



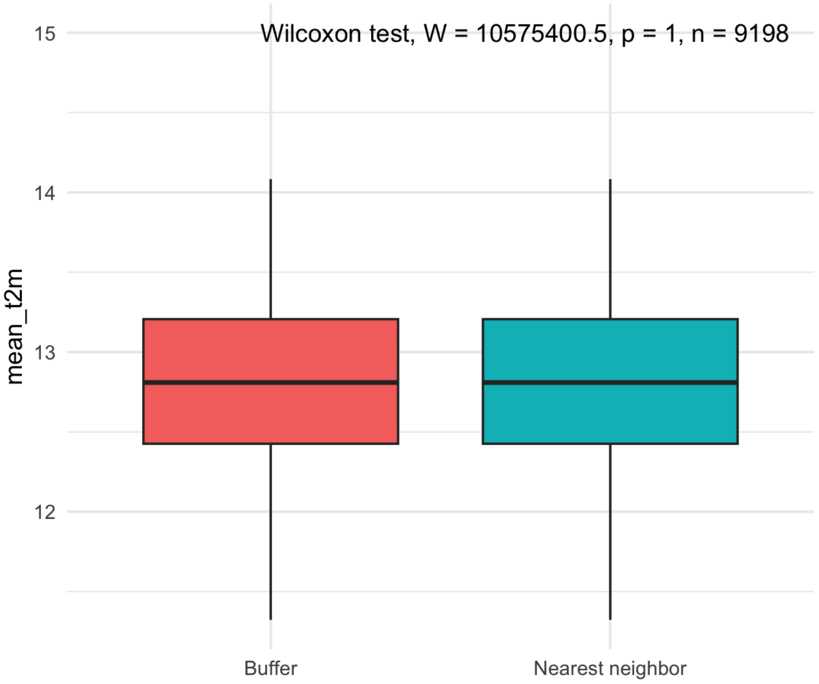
Appendix 1 : Cartography of the location of the inventories represented as black dots.



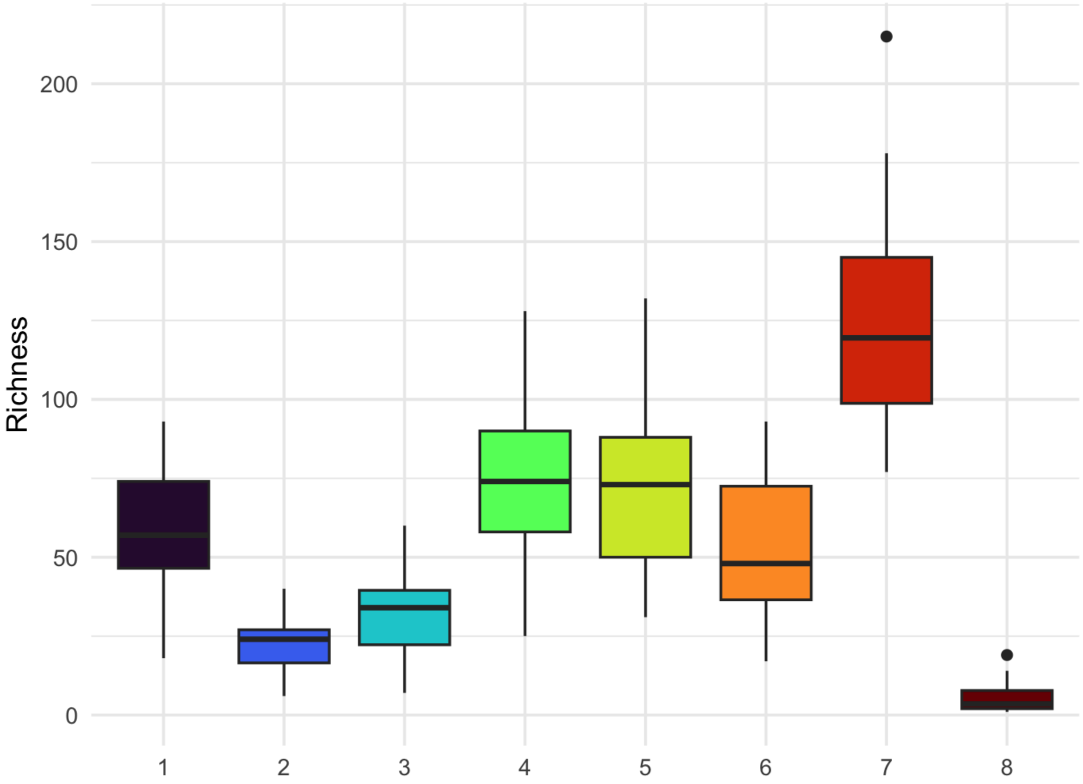
Appendix 2 : Barplot of the inventories' species richness, mean species richness and total number of inventories by year are represented in the top right corner.

Appendix 3 : Table of information regarding the environmental data tested

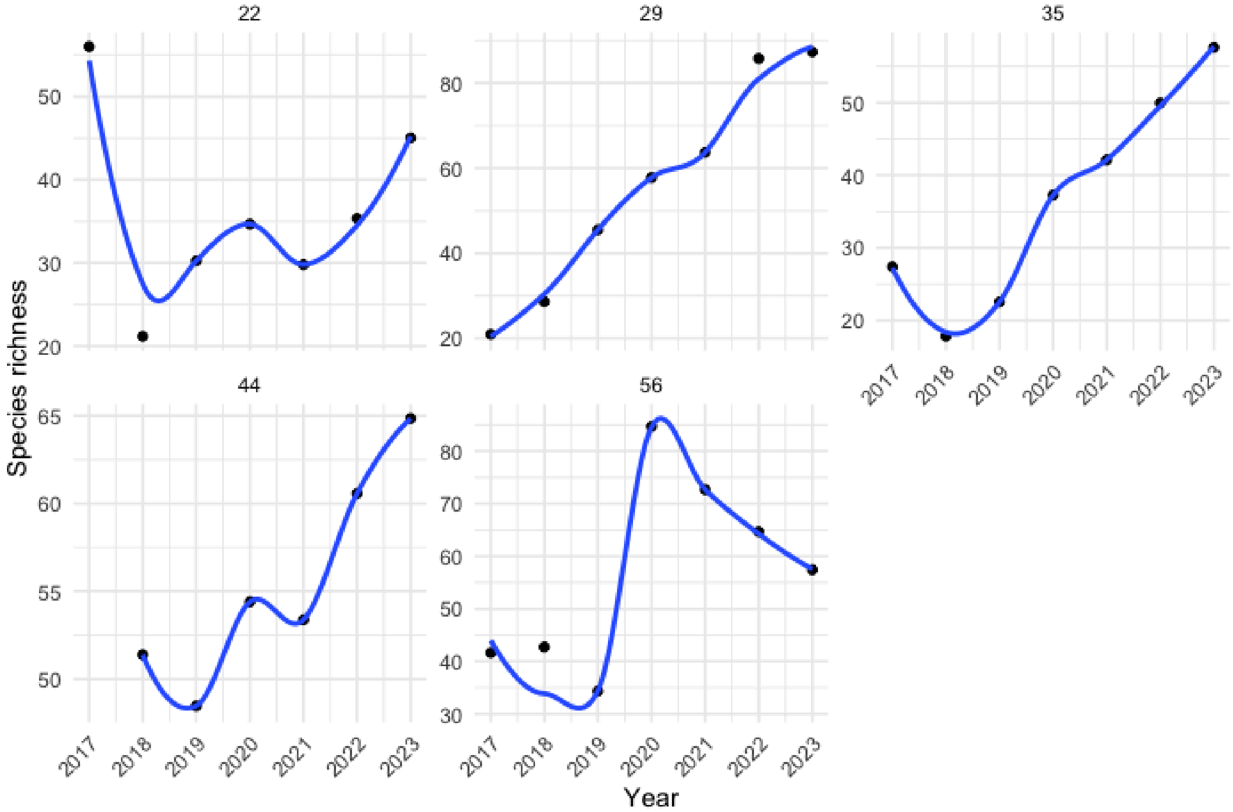
Extracted from	Temporal resolution	Spatial resolution	Variables	Source
Ecomars3D	Daily	16km²	Ammonium, nitrate, phosphate, silicate and dissolved oxygen concentration, water temperature, salinity, Spim (Seabed)	Ardhuin, 2013 (Ifremer)
Arpège	Hourly	10km²	Cloud cover, Humidity, air temperature, pressure, wind speed	Courtier et al., 1994 (météo france)
Wave fetch GIS layers for Europe at 100m scale	None	1km²	Global averaged fetch	Burrows, 2020
Assimilation of TOPEX/Poseidon and Jason altimetry data	None	10km²	Current velocity, tide amplitude, bed shear stress based on the current magnitude	Egbert et al., 2010
Coordinated Ocean Wave Climate Project (COWCLIP2.0)	None	0.703° longitude, 0.469° latitude (approximately 80km longitude, 50km latitude, 4000km²)	Maximum and average wave height	Morim et al., 2019
Coastal type	None	1km²	Coastal type	Salman et al., 2004 EMODnet Geology portal



Appendix 4 : Boxplot showing the values of the average air temperature (mean_t2m), for the 2 techniques of extraction tested (Buffer and Nearest neighbor) with the results of the Wilcoxon test used to compare the 2 methods.



Appendix 5 : Boxplot of the species richness of each cluster (from the second clustering performed at the scale of the pixels).

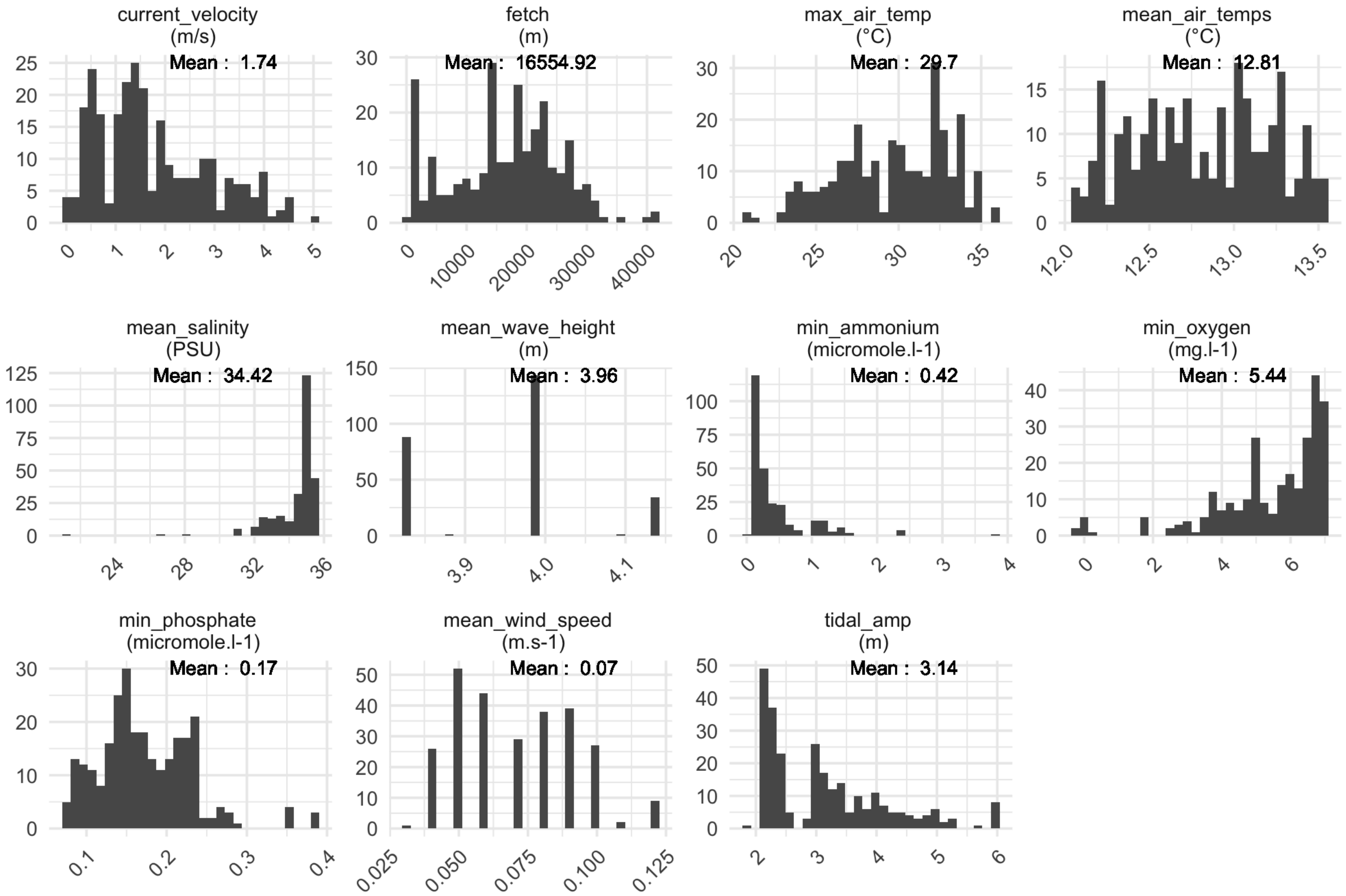


Appendix 6 : Temporal evolution of the species richness from 2017 to 2023, in the 5 departments studied : Côtes d’Armor (22), Finistère (29), Ille-et-Vilaine (35), Loire-Atlantique (44) and Morbihan (56). The blue line represents a loess regression to illustrate the trends.

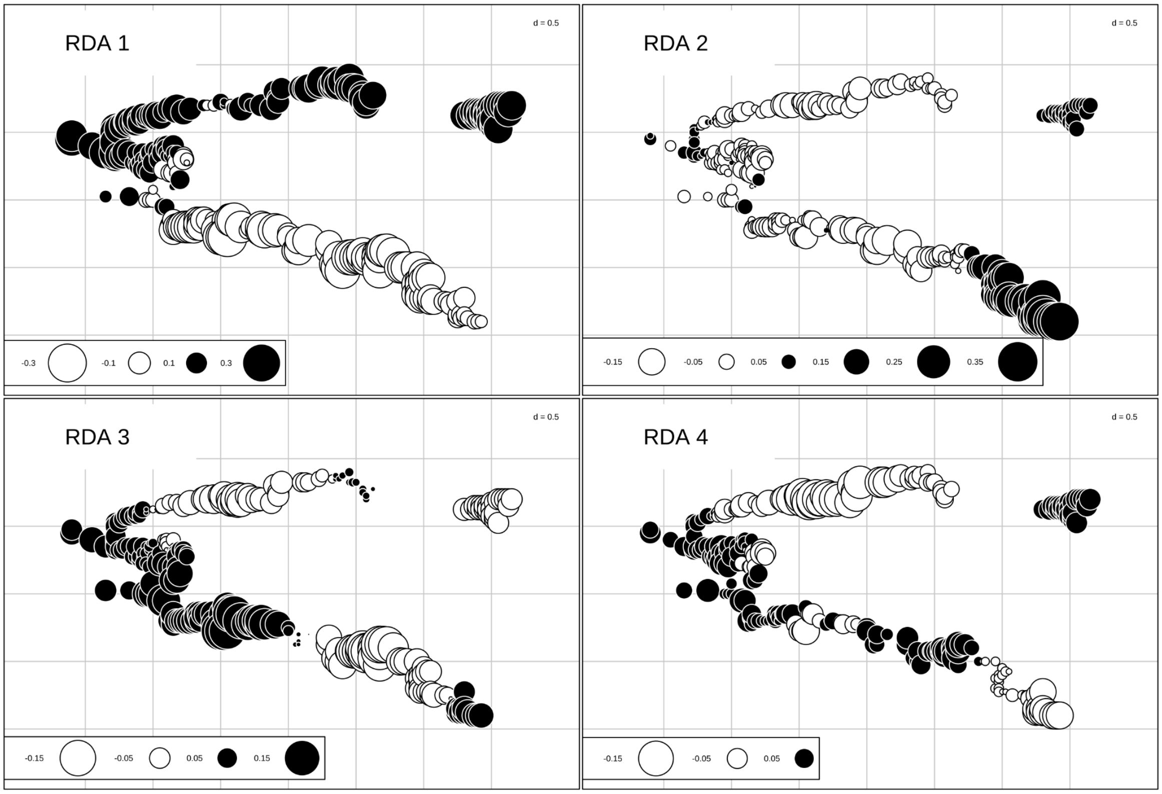
Appendix 7 : Variables chosen at each step of the variable selection for the 3 datasets used (Environment, Space and Expertise)

	Environmental dataset	Expertise dataset	Spatial dataset
Variables tested	Ammonium, nitrate, phosphate, silicate and oxygen concentration, water temperature, salinity, Spim, Cloud cover, Humidity, air temperature, pressure, wind speed Global averaged fetch (sd, min max et mean for each variables) Current velocity, tide amplitude, bed shear stress based on the current magnitude, Maximum and average wave height Coastal type (not included in the selection)	Average number of inventories, average expertise, average number of observers, average inventory year, percentage of inventories observed between 2017-2018, 2020-2021, and percentage of inventories observed between 2022-2023	MEM 1 to 46
Non collinear variables	Average : water and air temperature, spim, salinité, oxygen concentration Minimum : wind speed, phosphate, oxygen and ammonium concentration Maximum humidity, water temperature, air temperature, salinity Tidal amplitude, average wave height, fetch, current velocity	Average number of inventories, average expertise, average number of observers, percentage of inventories observed between 2017-2018, 2020-2021, and percentage of inventories observed between 2022-2023	(This step was not realized for the spatial variables)

	Coastal type (not included in the selection)		
Output of the forward selection	Average wave height, fetch, tidal amplitude, average salinity, current velocity, minimum phosphate, ammonium and oxygen concentration, maximum air temperature, average air temperature, minimum wind speed Coastal type (not included in the selection)	Average inventory year, average expertise	MEM 1 to 4

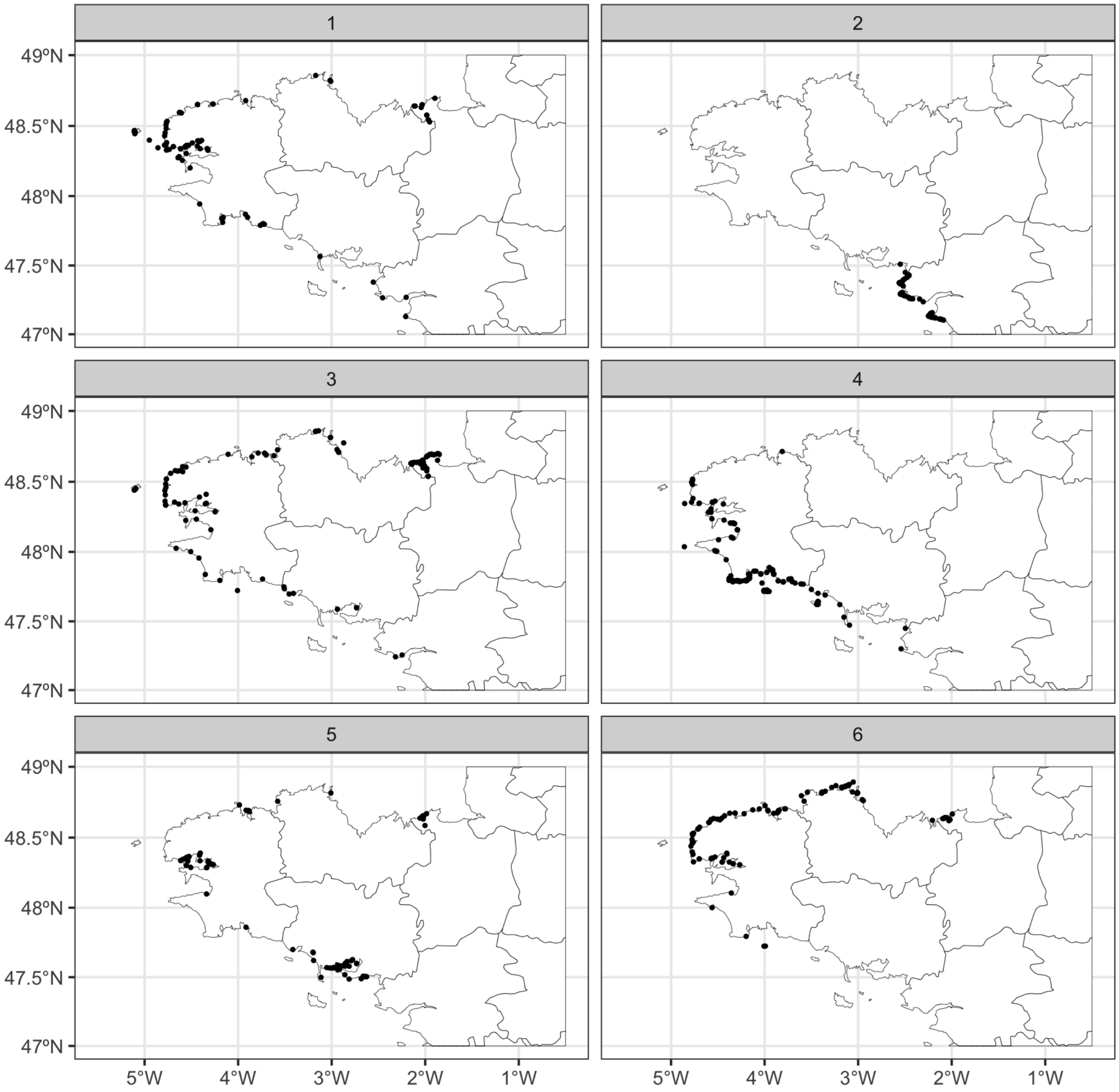


Appendix 8 : Barplot showing the values of the environmental variables selected and their average value.



Appendix 9 : Significant canonical axes of the dbMEM (Distance-Based Moran's Eigenvector Maps)

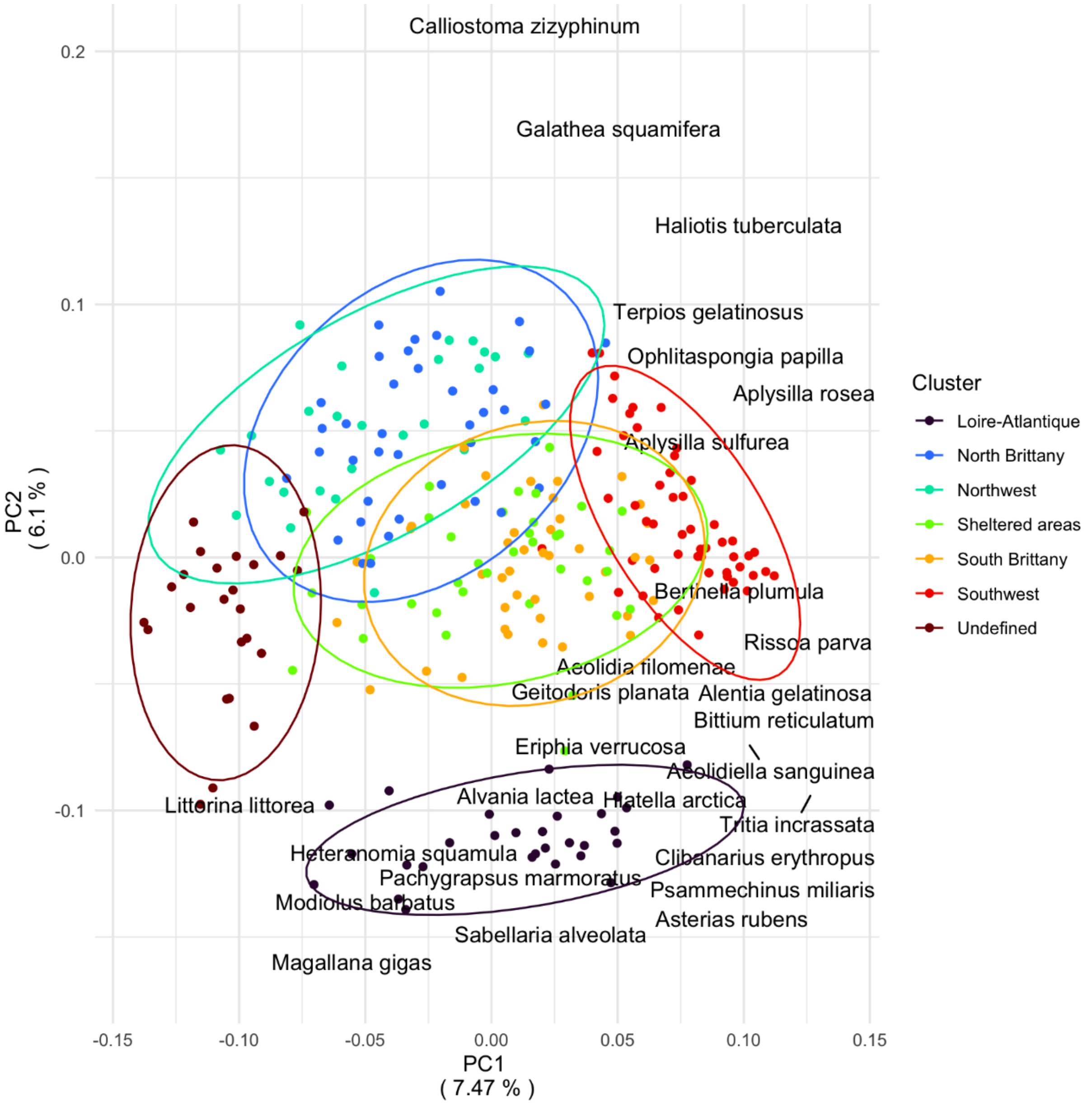
Appendix 10 : Evaluation scores of the seven models with 2 metrics (ROC and TSS), for the 3 algorithms used (GAM, GLM, RF).



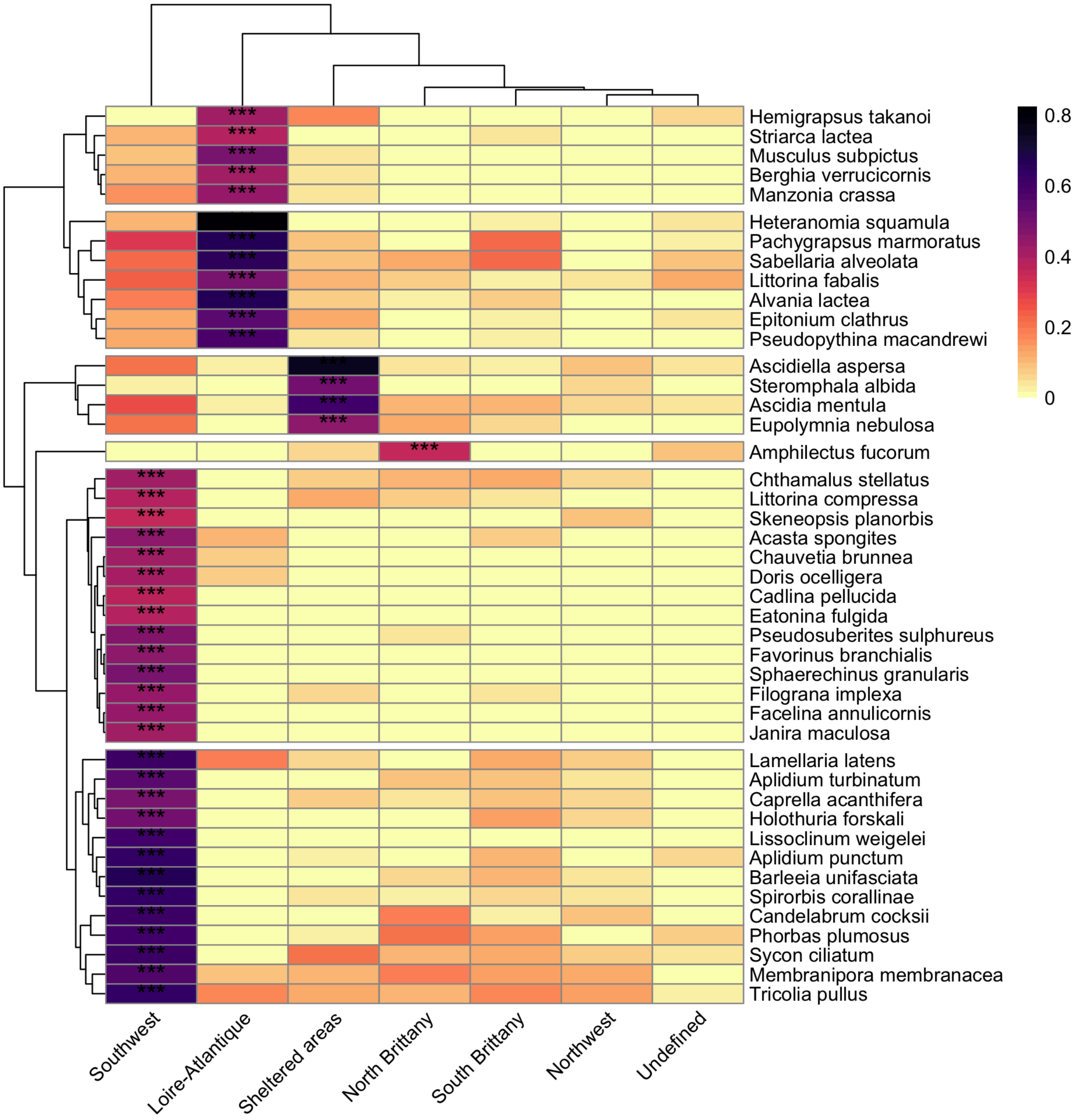
Appendix 11 : Cartography of the clusters repartition, from the first clusterization at the inventory scale. Dots represent inventory locations.

Appendix 12 : Table of the characteristics of the 8 clusters defined at the scale of the pixels

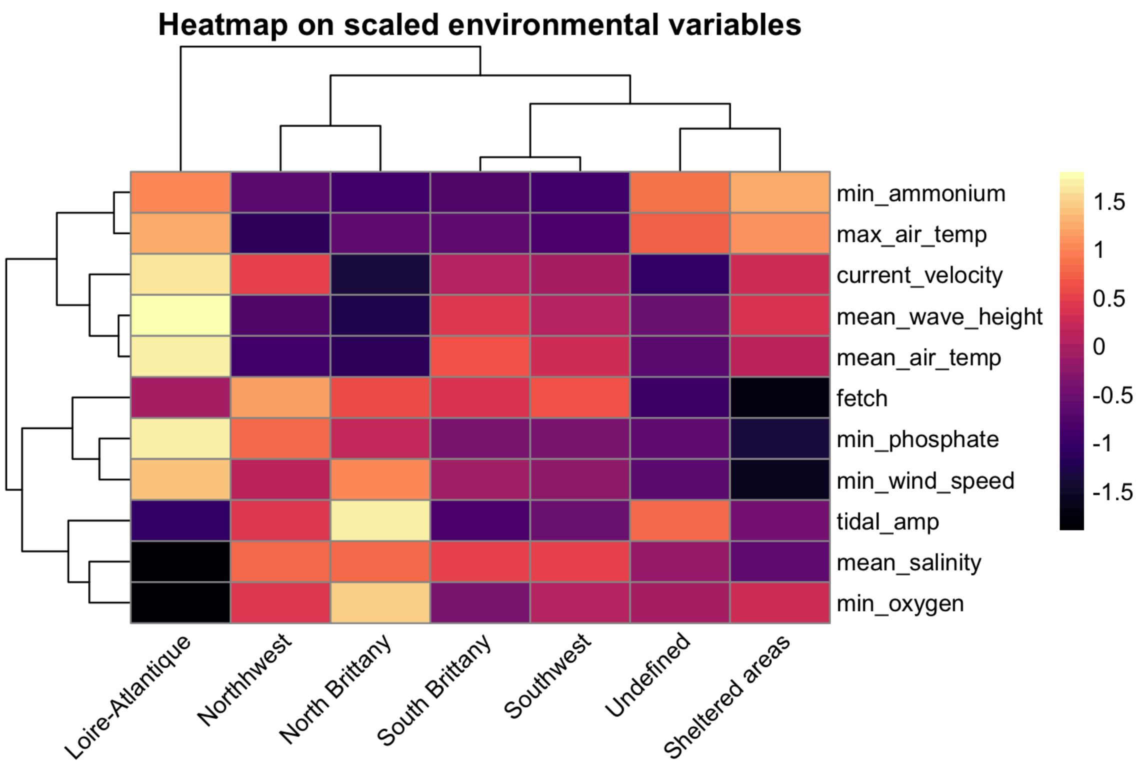
clusters	Average year of the inventories per pixels	Average species richness per pixel	Total number of pixels	Prevalence (%)	Average number of transects per pixel	% of inventories in 2017-2018 per pixel	% of inventories in 2020-2021 per pixel	% of inventories in 2022-2023 per pixel	Average expertise	Average number of observer per pixel
South Brittany	2020	54,3	43	16	1,4	11,67	36,67	15	4,41	1,16
Undefined	2020	20,83	24	9	1,29	16,13	45,16	9,68	3,30	1,09
Northwest	2019	29	26	10	1,96	27,45	15,69	5,88	4,37	1,35
Sheltered areas	2020	65,14	37	14	2,05	11,84	23,68	10,53	3,88	1,35
Loire-Atlantique	2021	57,1	29	11	2,59	12	14,67	20	3,96	1,24
North Brittany	2021	42,31	39	15	2,28	3,37	25,84	28,09	2,08	1,05
Southwest	2021	93,94	52	19	4,75	3,64	17,81	19,43	4,56	1,90
Cluster 8	2018	5,61	18	7	1,11	60	15	0	4,22	1,00
total	2020	46,03	268	100	2,18	18,26	24,32	13,58	3,85	1,27



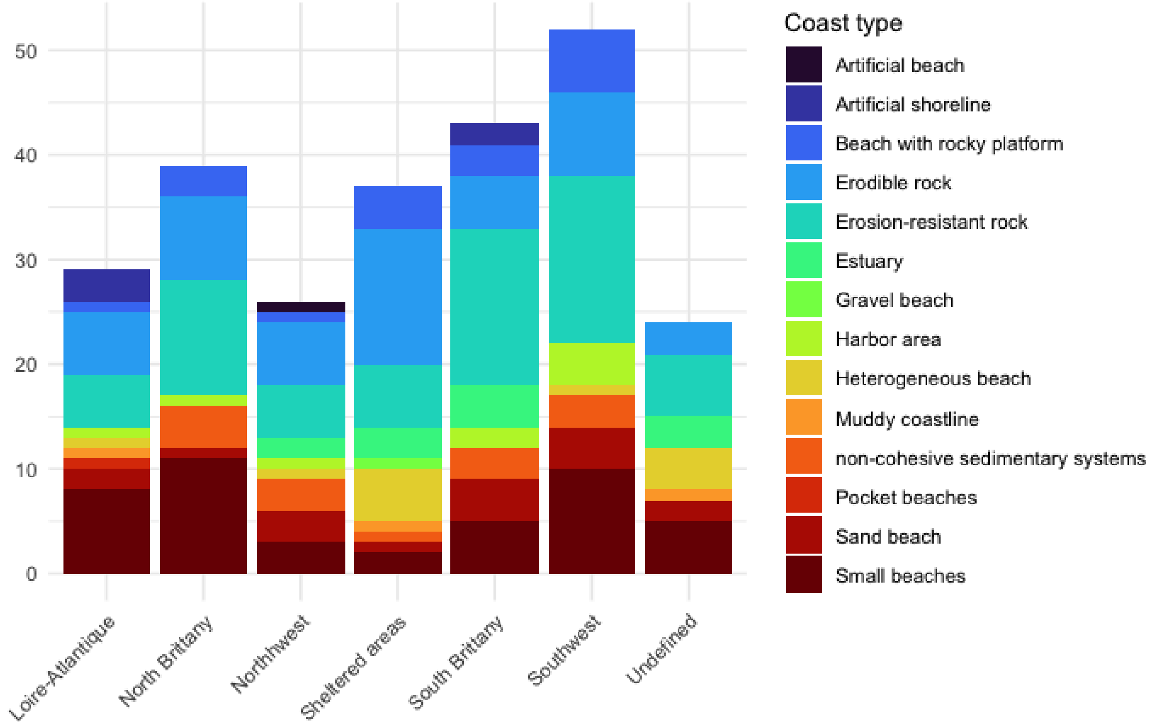
Appendix 13 : PCA biplot of species presence / absence data, Hellinger transformed. The points represent each inventory, the clusters are represented by color, with the ellipse corresponding to their 95% confidence interval (based on multi-normal distribution). For the sake of clarity, the species represented are those with a goodness of fit of at least 0.25 (.i.e. species whose at least 25% variance is represented) in the ordination plane formed by axes 1 and 2.



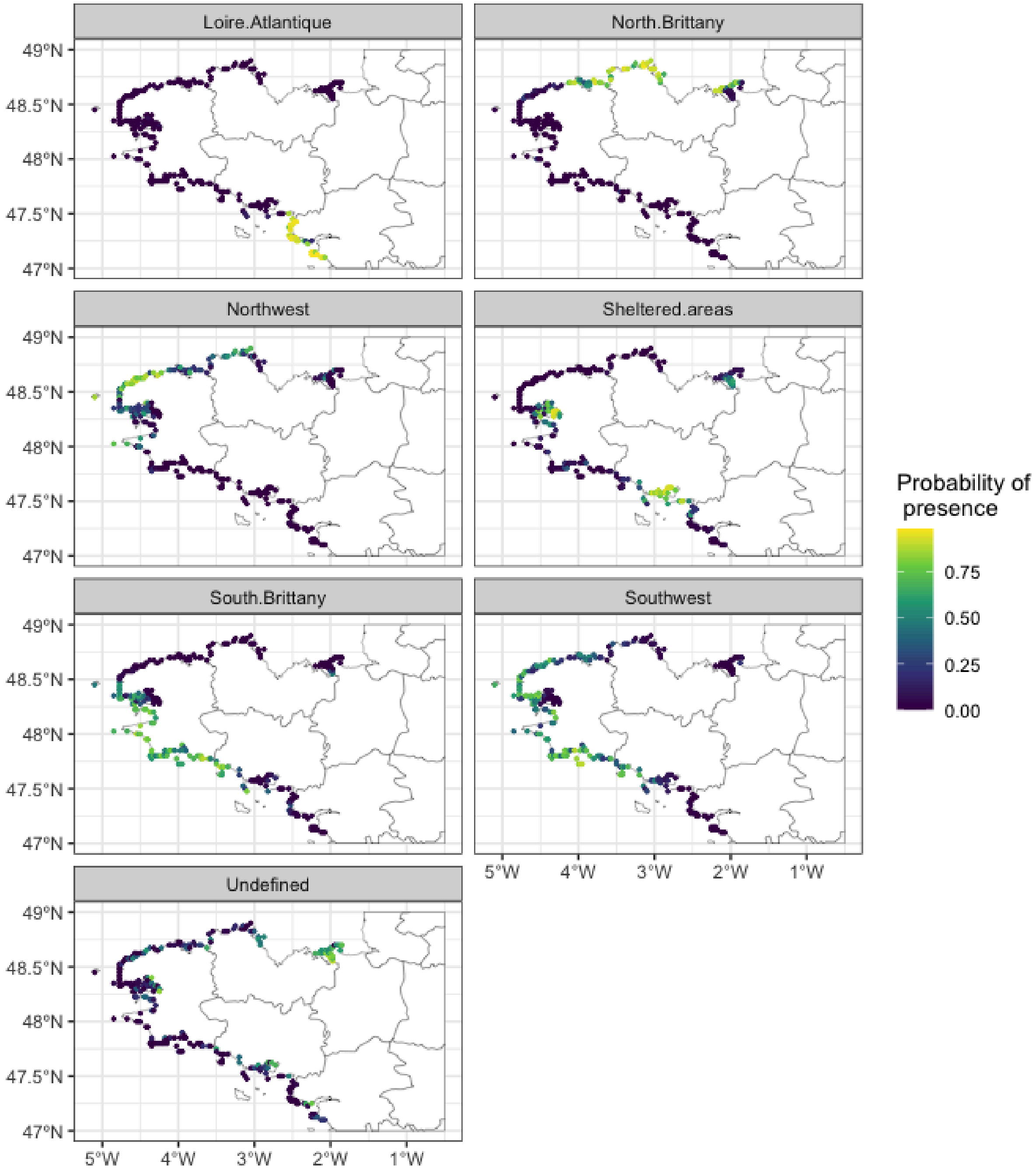
Appendix 14 : Heatmap showing the 7 clusters and their indicator species, represented by the highest Indval index species (> 0.35). The color gradient represents the value of the Indval index, the "***" symbol is present when the value of the Indval index is significant (p < 0.005). The upper dendrogram represents the similarity tree of the clusters based on the Indval value of the species and the one on the left side represent the similarities of the species based on their Indval value in each cluster



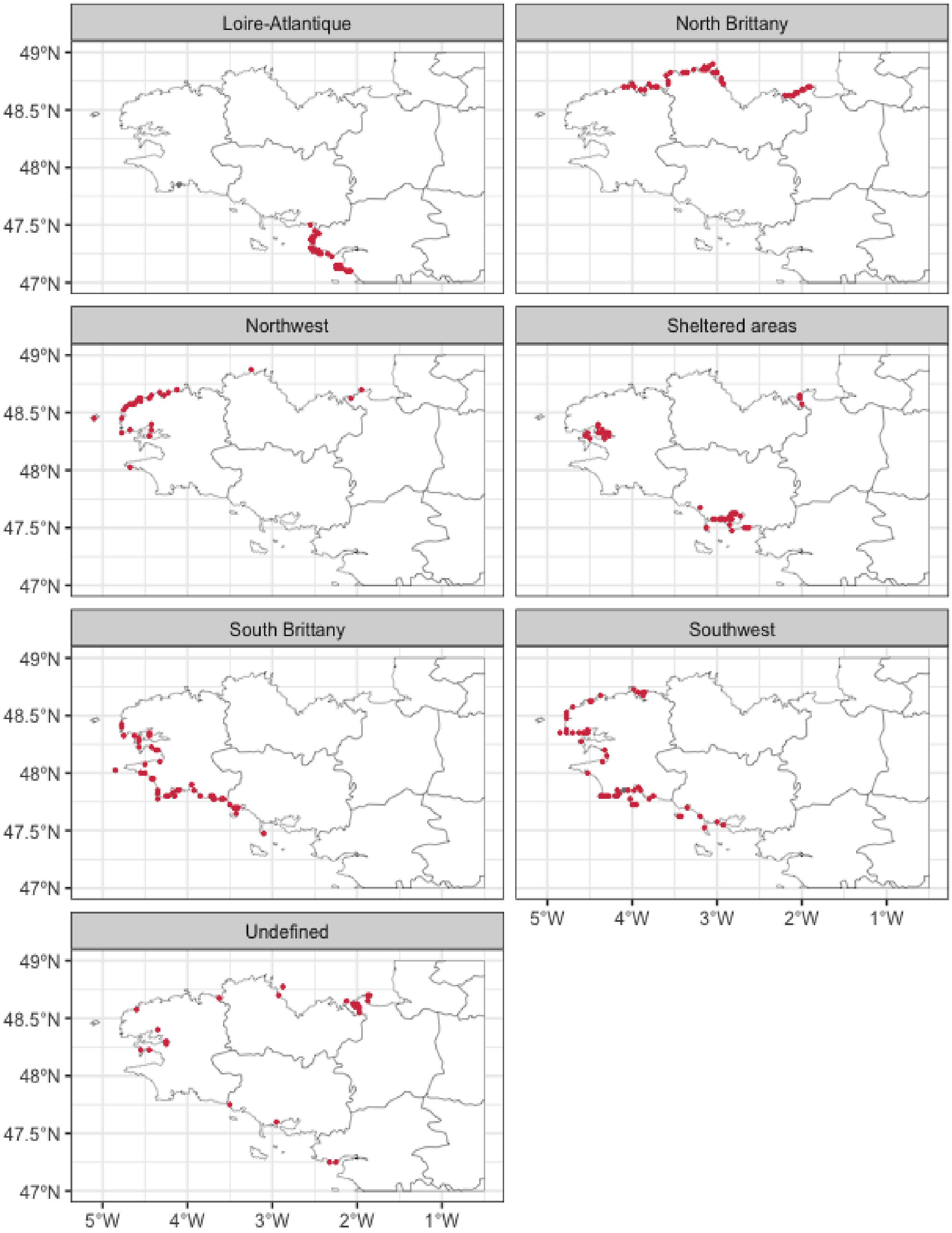
Appendix 15 : Heatmap on the chosen scaled environmental variables by clusters. The color gradient represents the scaled average value of the variables. The upper dendrogram represents the similarity of the clusters based on their environmental variables and the one on the side represents the similarity between the environmental values based on their values in the clusters.



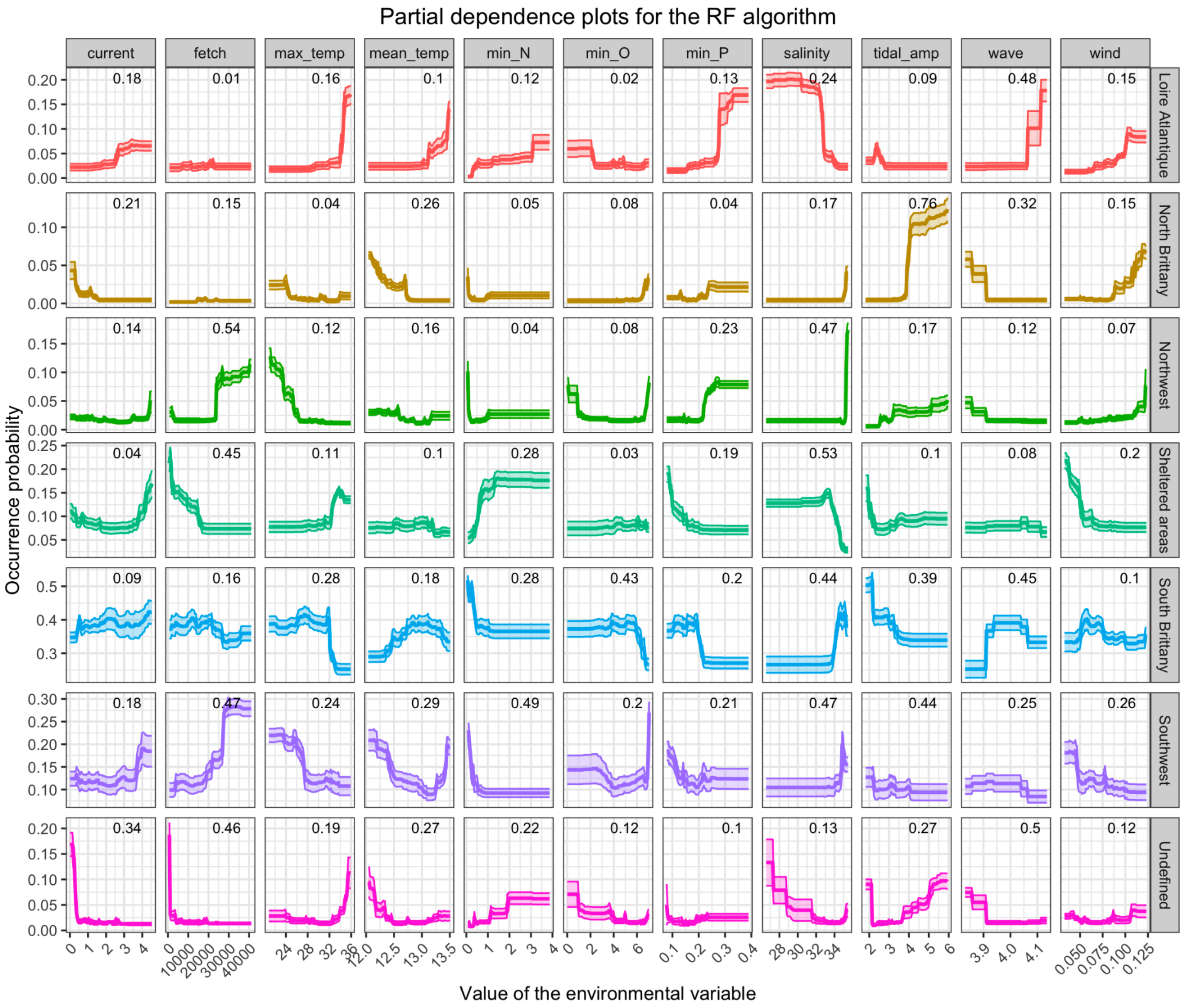
Appendix 16 : Barplot showing coast type by cluster.



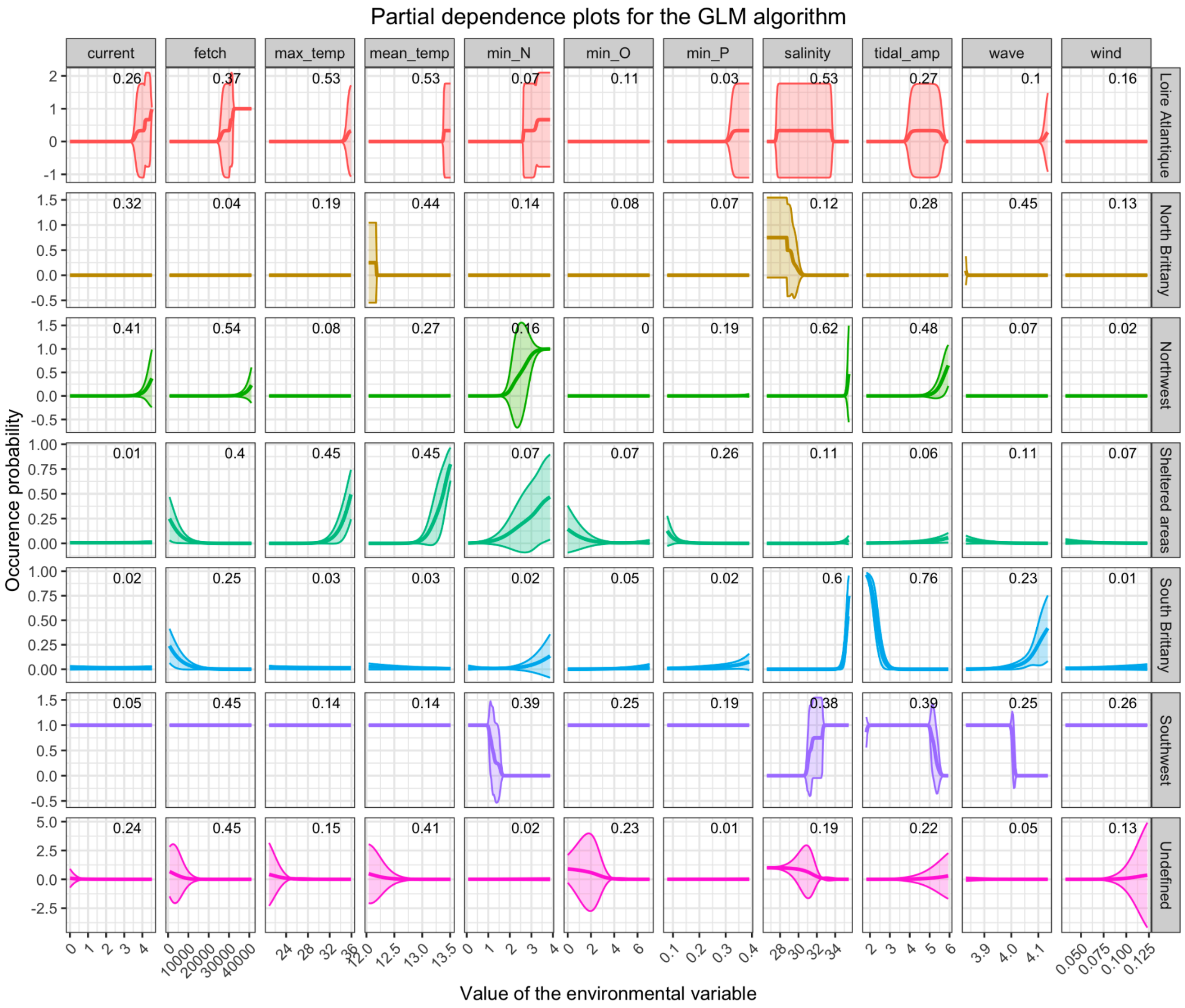
Appendix 17 : Probability of presence of the seven clusters predicted by the models. For the sake of clarity, the pixels were represented as points.



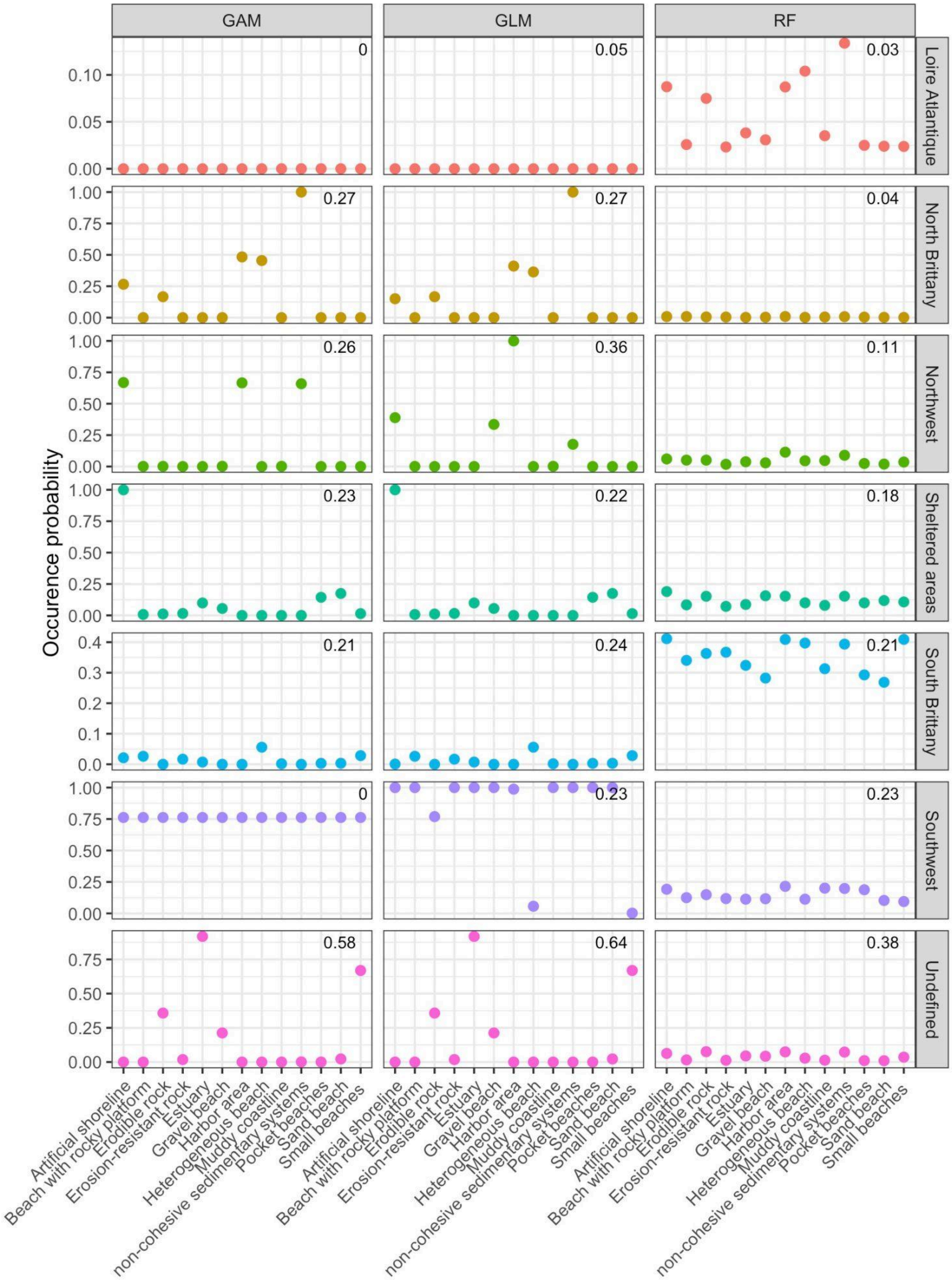
Appendix 18 : Predicted occurrence of the seven communities, calculated by converting the probabilities to occurrence (based on the ROC metric) to find the threshold for each model. For the sake of clarity, the pixels were represented as red dots.



Appendix 19 : Partial dependence plots for the RF algorithm showing the influence of each variables (current velocity, fetch, maximum and average ar temperature, minimum dissolved concentration of ammonium, oxygen, and phosphate, Salinity, tidal amplitude, mean wave height and wind speed), in columns, on the predicted probability of presence of each cluster, in rows and by color. The line represents the mean values of the 10 runs and the ribbon represents the standard deviation of the values. The importance of the variables are printed in the top right corner of each facet.



Appendix 20 : Partial dependence plots for the GLM algorithm showing the influence of each variables (current velocity, fetch, maximum and average ar temperature, minimum dissolved concentration of ammonium, oxygen, and phosphate, Salinity, tidal amplitude, mean wave height and wind speed), in columns, on the predicted probability of presence of each cluster, in rows and by color. The line represents the mean values of the 10 runs and the ribbon represents the standard deviation of the values. The importance of the variables are printed in the top right corner of each facet.



Appendix 21 : Partial dependence plots showing the influence of coast type on the predicted probability of presence of each cluster (in columns) for the 3 algorithms used (RF, GAM, GLM, in rows). The importance of the variables are printed in the top right corner of each fac

Abstract

In this study, we used the OBCE citizen-science program to describe and model the biogeography of rocky shore macrofaunal communities along Brittany's coastlines, while identifying the drivers of the identified biogeographic patterns. For this purpose, we used multivariate analyses to identify homogeneous clusters of communities and highlight the environmental conditions characterizing these clusters. We further used spatial distribution models to identify the drivers of spatial variations in the probability of presence of each cluster. We found 8 clusters that were distributed following longitudinal and latitudinal gradients but which were also characteristics of smaller scale patterns such as bays and estuaries. While the identity and distribution of clusters was affected by expertise bias to some extent, its influence was minor relative to the effect of space and environmental conditions. An ecotone cluster was identified in west Brittany as well as a cluster representing the geographical barrier : the Ushant front. The models showed good predictive power and permitted us to precise environmental characteristics of some clusters. These models could further be used to draw continuous maps and future predictions of the clusters.

Keywords : Biogeography, Rocky shore, Macrofauna, Citizen science, SDM

Résumé

Cette étude se base sur les données de l'OBCE, un programme de science participative, pour décrire et modéliser la biogéographie de la macrofaune des estrans rocheux le long des côtes bretonnes et d'identifier ses drivers. Pour ce faire, des analyses multivariées ont été utilisées pour identifier des clusters de communautés homogènes et mettre en avant les conditions environnementales caractéristiques des clusters. Des modèles de distribution spatiale ont par la suite été utilisés pour identifier les drivers des variations spatiales dans la probabilité de présence des clusters. 8 clusters ont été identifiés suivant un gradient longitudinal et latitudinal mais aussi montrant des caractéristiques de patron spatiaux à plus fine échelle comme les baies et les estuaires. L'identité et la distribution des clusters ont été dans une certaine mesure affectées par un biais d'expertise, son influence était cependant mineure au regard de l'effet de l'espace et de l'environnement. Un cluster écotone a été identifié dans l'ouest de la Bretagne, ainsi qu'un cluster représentant la barrière géographique caractérisée par le front d'Ouessant. Les modèles ont montré un bon pouvoir prédictif et ont permis de préciser les caractéristiques environnementales de certains clusters. Ces modèles pourraient ultérieurement être utilisés pour obtenir des cartes continues de la répartition des clusters et des prédictions futures.

Mots clés : Biogéographie, Estrans rocheux, Macrofaune, Science participative, SDM