PAEWG-01-15

First Meeting of the Protected Areas and Ecosystems Working Group (PAEWG1)

18-19 March 2019

CCAMLR's Vulnerable Marine Ecosystems bioindicator taxa: a relevant tool for benthic ecoregionalisation

Relates to agenda item: 3.2 VME Indicator taxa and Encounter protocols

Working paper ☒ Info paper ☐

Delegation of France (Territoires)

Abstract

This document is a short extraction from the article "Benthic ecoregionalisation and conservation issues in the French Exclusive Economic Zone of Kerguelen" submitted the 16/05/2018 to the journal CCAMLR Science. Original article includes full benthic ecoregionalisation modelling results. In this document, a methodological focus is made to highlight the relevance of using the CCAMLR's Vulnerable Marine Ecosystems bioindicator taxa to build benthic ecoregionalisation models, which constitutes a secondary result of the original study.

Recommendations *(working papers only)*

This document is to be discussed within the PAEWG-01 and during the PAEWG-01 physical meeting.

CCAMLR'S VULNERABLE MARINE ECOSYSTEMS BIOINDICATOR TAXA: A RELEVENT TOOL FOR BENTHIC ECOREGIONALISATION

Alexis Martin* (1), Emmanuelle Trouslard (1), Mélyne Hautecoeur (1), Jonathan Blettery (1), Camille Moreau (2), Thomas Saucède (3), Nadia Améziane (4), Guy Duhamel (1), Marc Eléaume (4)

* Corresponding author: Alexis Martin, alexis.martin@mnhn.fr

(1) Département Adaptations du Vivant, Muséum national d'Histoire naturelle, UMR BOREA 7208, 57 rue Cuvier, 75231 Paris Cedex 05, France

(2) Laboratoire de Biologie Marine, Université Libre de Bruxelles, Avenue FD Roosevelt 50, CP 160/15, 1150 Bruxelles, Belgique

(3) UMR 6282 Biogéosciences, Université Bourgogne Franche-Comté, CNRS, 6 bd Gabriel F-21000 Dijon, France

(4) Département Origine et Évolution, Muséum national d'Histoire naturelle, UMR ISYEB 7205, 57 rue Cuvier, 75231 Paris Cedex 05, France

Abstract

This document is a short extraction from the article "Benthic ecoregionalisation and conservation issues in the French Exclusive Economic Zone of Kerguelen" submitted the 16/05/2018 to the journal *CCAMLR Science* for the proceedings of The Second Symposium on Kerguelen Plateau Marine Ecosystems and Fisheries (13–15 November 2017, Hobart, Tasmania). The document, in revision, has been accepted with minor corrections. Full text, figures, results, discussion and references will be available in the original article. Original article includes full benthic ecoregionalisation modelling results and an assessment of the new marine reserve of the Kerguelen EEZ regarding to the benthic conservation issues. In this document, a methodological focus is made to highlight the relevance of using the CCAMLR's Vulnerable Marine Ecosystems bioindicator taxa to build benthic ecoregionalisation models, which constitutes a secondary result of the original study.

1. Introduction

(…) Understanding deep-sea ecosystems structure represents a major scientific issue for conservation and sustainable exploitation of commercial species in the French EEZ [of Kerguelen]. In addition to global change (Smith et al., 2006), benthic ecosystems in the French EEZ of Kerguelen are impacted by an industrial deep-sea longline fishing activity, focusing today on Patagonian toothfish *Dissostichus elegenoides* (Duhamel et al., 2011).

To prevent potential impacts of longline fishing and following the evolution of international regulation, the "Terres Australes et Antarctiques Françaises" (TAAF) administration transposed in 2014 CCAMLR Conservation Measures 22/06 and 22/07 about Vulnerable Marine Ecosystems (VME) conservation (TAAF, 2008) in the French EEZ of Kerguelen (TAAF, 2014). Furthermore, the Muséum national d'Histoire naturelle (MNHN) started to develop in 2015 a permanent VME monitoring program to assess the real impact of the fishery on deep-sea benthic habitats (Martin et al., 2017).

Moreover, the Marine Reserve of Kerguelen, managed by TAAF administration and originally limited to some coastal areas, has been recently extended to a large portion of deep-sea environments in the French EEZ in 2016 and 2017 (TAAF, 2017). (…)

Here we (…) identify sub-regions characterized by homogeneous benthic assemblages using a measurable and repeatable methodology (…) [and highlight the relevance of using the CCAMLR's Vulnerable Marine Ecosystems bioindicator taxa to build benthic ecoregionalisation models**].**

2. Materials and methods

Approach and choice of the method

Ecoregionalisation method is used in this work as a "process and output of identifying and mapping broad spatial patterns based on physical and biological attributes through classification methods used for planning and management purposes" (Reygondeau et al., 2014). The choice of this approach aims to meet the first main objective of this study, to characterise ecological sub-regions based on homogeneous benthic assemblages and to contribute to our understanding of deep-sea benthic ecology of the study area. Various ecoregionalisation methods have been developed and are commonly used in ecosystem-based spatial management studies (e.g. Berline et al., 2014 ; Blasi et al., 2015, Duranton, 2006; Hill et al., 2018; Hornsmann et al., 2008; Hunter et al., 2018; O'Hara, 2008). Given its relevance to our objectives the Generalized Dissimilarity Modeling (GDM) technique has been selected for analysis (Ferrier et al., 2007). GDM is a "statistical technique for modeling spatial variation in biodiversity between pairs of geographical locations. It allows to make predictions and map biological patterns by transforming environmental predictor variables" (Ferrier et al., 2007). To perform the GDM and produce interpretable results in terms of benthic assemblages, a complete analysis pipeline has been developed, using R software (R Core Team, 2015) and includes different steps, from dataset assembly to statistical analyses. The choice of this robust and proven modeling technique (e.g. Ives and Helmus, 2011; Koubbi et al., 2010; Leathwick et al., 2010; Leitao et al., 2015), associated to the development of a pipeline, aim to process a complete analysis using a measurable and repeatable methodology allowing inter-comparison between models.

Biological data

POKER (POissons de KERguelen) is a program led by the MNHN, Paris, and has been designed to monitor commercial and by-catch fish species within the French EEZ of Kerguelen (Duhamel et al., 2011). This monitoring program is based on systematic random stratified trawl surveys over the northern part of the Plateau, from 85 to 1,000 m depth. POKER II cruise (October 2010) included a marine invertebrate by-catch sampling program (Duhamel et al., 2011). The 40 mm mesh bottom trawl used during POKER was designed for catching demersal fish, and benthic invertebrates collected as by-catch were limited to macro-epibenthic invertebrate. All organisms collected have been sorted, weighted, identified, and photographed (Eléaume et al., 2011). Furthermore, representative sub-samples of specimens have been collected and fixed in 96% ethanol to be later identified using molecular and anatomical approaches (Eléaume et al., 2011). A set of presence/absence data for 111 taxa and morphotypes identified at various taxonomic levels (fig. 1) has been produced for the 209 stations of the survey (fig. 2).

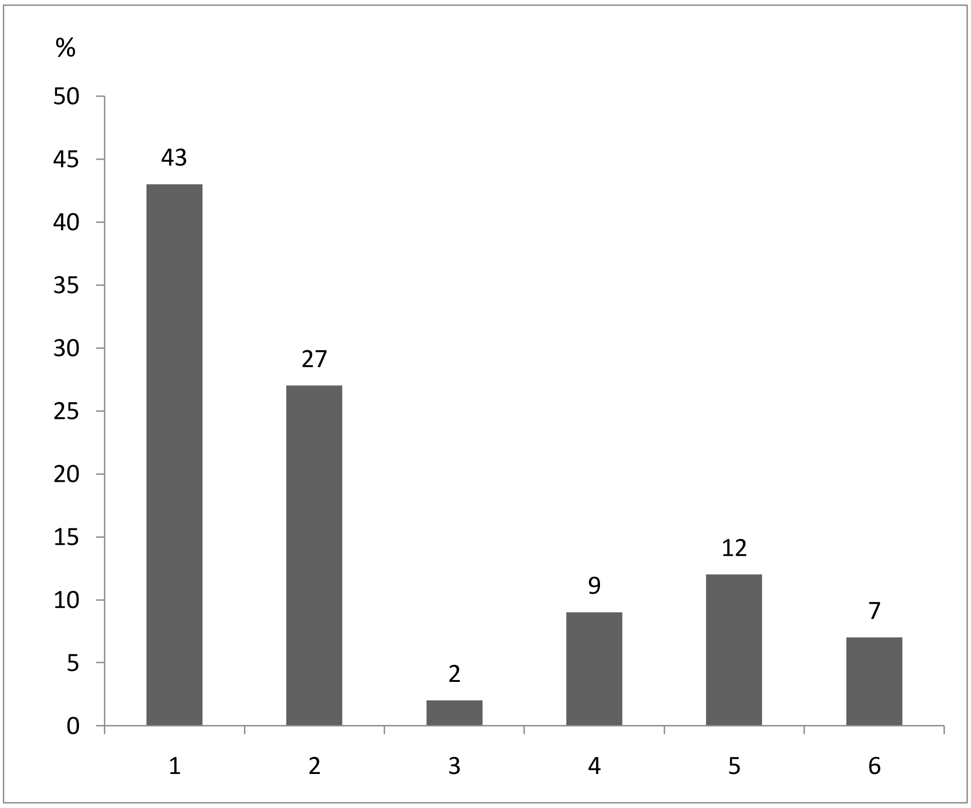

Fig.1: taxonomic identification levels given in percentage for the 111 taxa of the POKER II invertebrates dataset; 1: species or morphotype; 2: genus; 3: family; 4: order; 5: class; 6: phylum

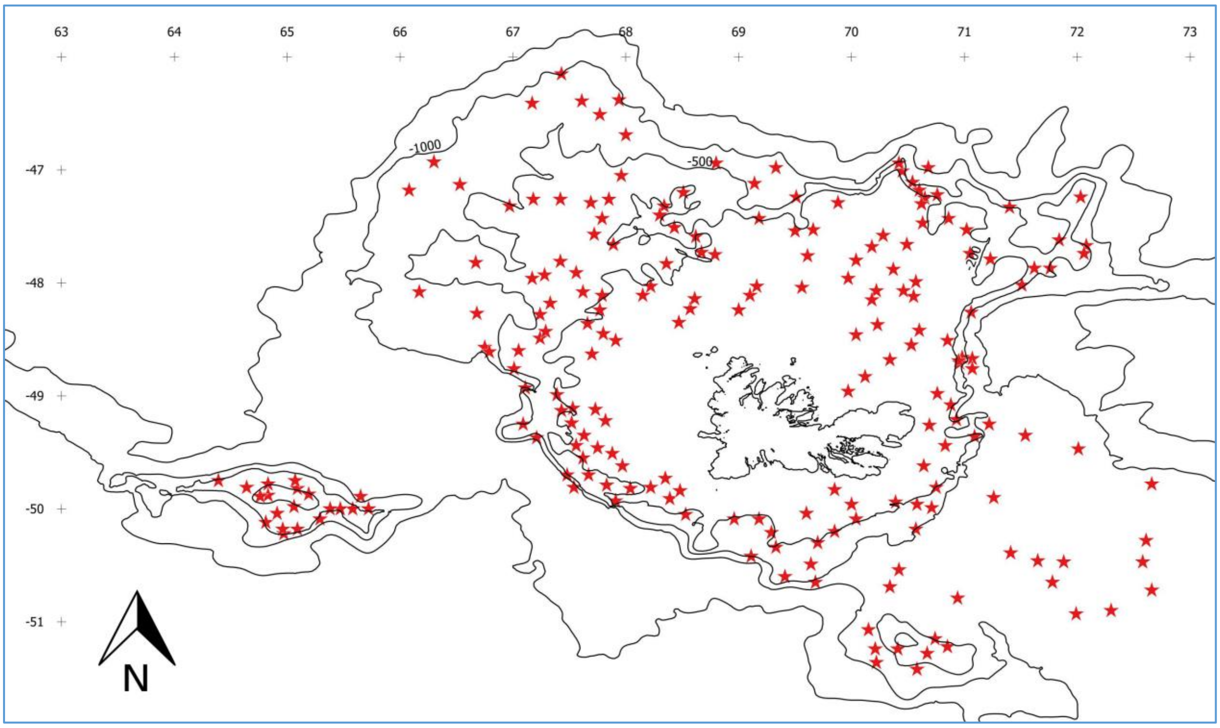

Fig.2: geographical distribution of POKER II cruise sampling stations over the French Exclusive Economic Zone of Kerguelen

Four types of grouping have been used to pool observations together and produce ecoregionalisation models: life history traits (trophic mode and motility), protection status and taxonomic identification. Scientific literature (Gutt, 2007; Shojaei, 2016) and taxa listed by CCAMLR as VME bio-indicators (mostly engineer taxa and/or sessile and/or suspension feeders) (CCAMLR, 2008) have been used to define relevant groups and classify all the observations in specific tables.

Environmental data

Fifteen environmental variables representing seafloor and sea surface conditions have been extracted or derived (Guillaumot et al., 2016) from available international databases (Boyer et al., 2013; Douglass et al., 2013; Mccoy, 1991; Smith and Sandwell, 1997), and processed so as to fit species postulated life span, species observed bathymetric distribution and POKER II spatial coverage (Duhamel et al., 2011) (tab. 1). Environmental statistics were averaged over the 2005-2012 period and encompass depths ranging from 85 to 1000 m with spatial extension from 46° to 52° South and 63° to 73° East. Spatial resolution was set to a common 0.1°. Poker II transects mid-points were used to assign environmental statistics to biological occurrences.

Prior to modeling, explanatory factors have been selected using the Spearman correlation test with a threshold of $rs <= 0,66$ (Mukaka, 2012). Whenever two variables displayed rs > 0.66, a rule of thumb was applied and variables considered to have the highest explanatory power were retained.

Table 1: environmental factors used for the ecoregionalisation modelling with references

Environmental predictors	References
Depth	Guillaumot *et al.* (2016), derived from Smith and Sandwell (1997)
Average of surface temperature	World Ocean Circulation Experiment (2013)
Variance of surface temperature	World Ocean Circulation Experiment (2013)
Average of temperature near seafloor	Guillaumot *et al.* (2016), derived from World Ocean Circulation Experiment (2013) sea surface temperature layers
Variance of temperature near seafloor	Guillaumot *et al.* (2016), derived from World Ocean Circulation Experiment (2013) sea surface temperature layers
Average of surface salinity	World Ocean Circulation Experiment (2013)
Variance of surface salinity	World Ocean Circulation Experiment (2013)
Average of salinity near seafloor	Guillaumot *et al.* (2016), derived from World Ocean Circulation Experiment 2013 sea surface salinity layers
Variance of salinity near seafloor	Guillaumot *et al.* (2016), derived from World Ocean Circulation Experiment 2013 sea surface salinity layers
Average of chlorophyll concentration in summer	MODIS AQUA (NASA) (2010)
Geomorphology	ATLAS ETOPO2 2014 (Douglass *et al.* 2014)
Sediments	McCoy (1991), updated by Griffiths 2014 (unpublished).
Seafloor slope	Smith and Sandwell (1997)
Average of oxygen concentration near seafloor	Guillaumot *et al.* (2016), derived from World Ocean Circulation Experiment 2013 sea surface oxygen concentration layers
Ruggedness	Guillaumot *et al.*(2016), derived from bathymetric layer

Statistical analyses

GDM technique is based on analysis of biological dissimilarity between sampling stations according to the variation of related environmental factors (Ferrier et al., 2007).

GDM technique takes into account biological and environmental datasets and produces a dissimilarity matrix that best reflects biological/environment interactions. Each presence/absence biological table is transformed into a dissimilarity matrix using Jaccard Index (Goslee et al., 2007). This biological dissimilarity matrix is compared to environmental layers and a new table is produced that includes for all possible pairs of sampling stations biological distance and environmental statistics. A Generalized Linear Model (GLM) (McCullagh and Nelder, 1989) is fitted to this table, with the biological dissimilarity as the explained value and the environmental statistics as the explanatory factors. An iterative modeling process is run to select the best GLM and the best set of environmental factors, using the deviance explained as testing criteria (McCullagh and Nelder, 1989).

To predict biological dissimilarity over the whole study area, models are secondarily applied to the main environmental grid. Each ecoregionalisation model results in a new distance matrix where the biological distance index between all the pairs of pixels of the main grid is calculated from the GLM. Predictions are limited to the cells of the main grid that present environmental conditions similar to those observed at the sampling stations. In addition, data gaps are not compensated by interpolation techniques, to allow a critical evaluation of the quality of the modelling outputs.

Modeled predictions are then clustered using Partitioning Around Medoids (PAM), an algorithm that accommodates very large matrices (Kaufman and Rousseeuw, 1987). Each clustering is iterated ten times, starting with three a priori clusters, and finishing the process at twelve clusters. Number of clusters that maximizes “*Sih*” Silouhette Index (Rousseeuw, 1987) is selected.

For each ecoregionalisation model, ecological sub-regions or ecotypes are characterised. Ecotypes are here defined as spatially delimited areas with shared similar environmental factors and benthic assemblages. For this study, the spatially delimited areas are the areas obtained from the clustering of model predictions. Environmental envelope of each ecotype is summarized using environmental statistics of pixels attributed to each cluster. Benthic assemblages of ecotypes are the taxa detected in sampling stations located within the spatially delimited areas.

To characterize species assemblages in a given spatial area, many techniques are commonly used (e.g. Johnson et al., 1993; Kremen, 1992; Pardo and Armitage, 1997; Ward et al. 1999; Wright and Reeves, 1992). For this study, *IndVal* index is used (Dufrêne and Legendre, 1997). *Indval* attributes a value to each taxon or group of taxa in a given area. This value is the product of a specificity component defined as "*the probability that the surveyed site belongs to the target site group given the fact that the species has been found*", and a sensitivity component defined as "*the probability of finding the species in sites belonging to the site group*" (italics are quoted from Dufrêne and Legendre, 1997). The higher the value of *IndVal* index attributed to a taxon or a combination of taxa, the higher its value as a bioindicator. *IndVal* statistical significance is also tested and a *p-value* is provided (Dufrêne and Legendre, 1997). For this study, we performed *IndVal* analysis by testing bioindicator value for all taxa that were detected in the various ecotypes of each ecoregionalisation model. We also performed the analysis by testing bioindicator value for all combinations of two and three of those taxa.

Ecotypes are projected onto a map using QGIS (QUANTUM GIS Development Team, 2012) and surface overlap with Marine Reserve Strictly Protected Areas (MRSPA) is computed. Percentage of each ecotype located within MRSPAs is here interpreted as a measurement of the protection level of each ecotype. This level of protection is contextualized according to the degree of rarity of each ecotype calculated as the percentage of each ecotype in the whole study area. These results are used in a preliminary evaluation of the relevance of the protection zones for the conservation of deep-sea benthic assemblages.

3. Results

Six out of fifteen uncorrelated environmental factors are retained for modeling: depth, surface temperature mean, sea surface salinity amplitude, chlorophyll a summer mean, slope and sediments.

Twenty-one distinct benthic ecoregionalisation models are produced and projected onto a map also showing MRSPAs localisation. Pixel size indicates the geographical scale of predictions and blank areas indicate missing, uninterpolated data.

Thirteen models have been computed using taxa at various identification levels: Actiniaria (fig. 3a), Annelida (fig. 3b), Arthropoda (fig. 3c), Ascidiacea (fig. 3d), Asteroidea (fig. 3e), Cnidaria (fig. 3f), Corals (fig. 3g), Echinodermata (fig. h3), Echinoidea (fig. 4a), Holothuroidea (fig. 4b), Mollusca (fig. 4c), Ophiuroidea (fig. 4d) and Porifera (fig. 4e). A complete benthic ecoregionalisation model has also been computed by pooling together all 111 taxa occurrence data (fig. 4f). Six models have been computed using life history traits: motility includes Vagile (fig. 4g) and Mobile (fig. 4h); organisms feeding mode includes Detritivore (fig.5a), Necrophagous (fig.5b), Predator (fig.5c) and Suspension feeders (fig.5d). A single model was obtained using CCAMLR Vulnerable Marine Ecosystems (VME) taxa (fig. 5e).

(…)

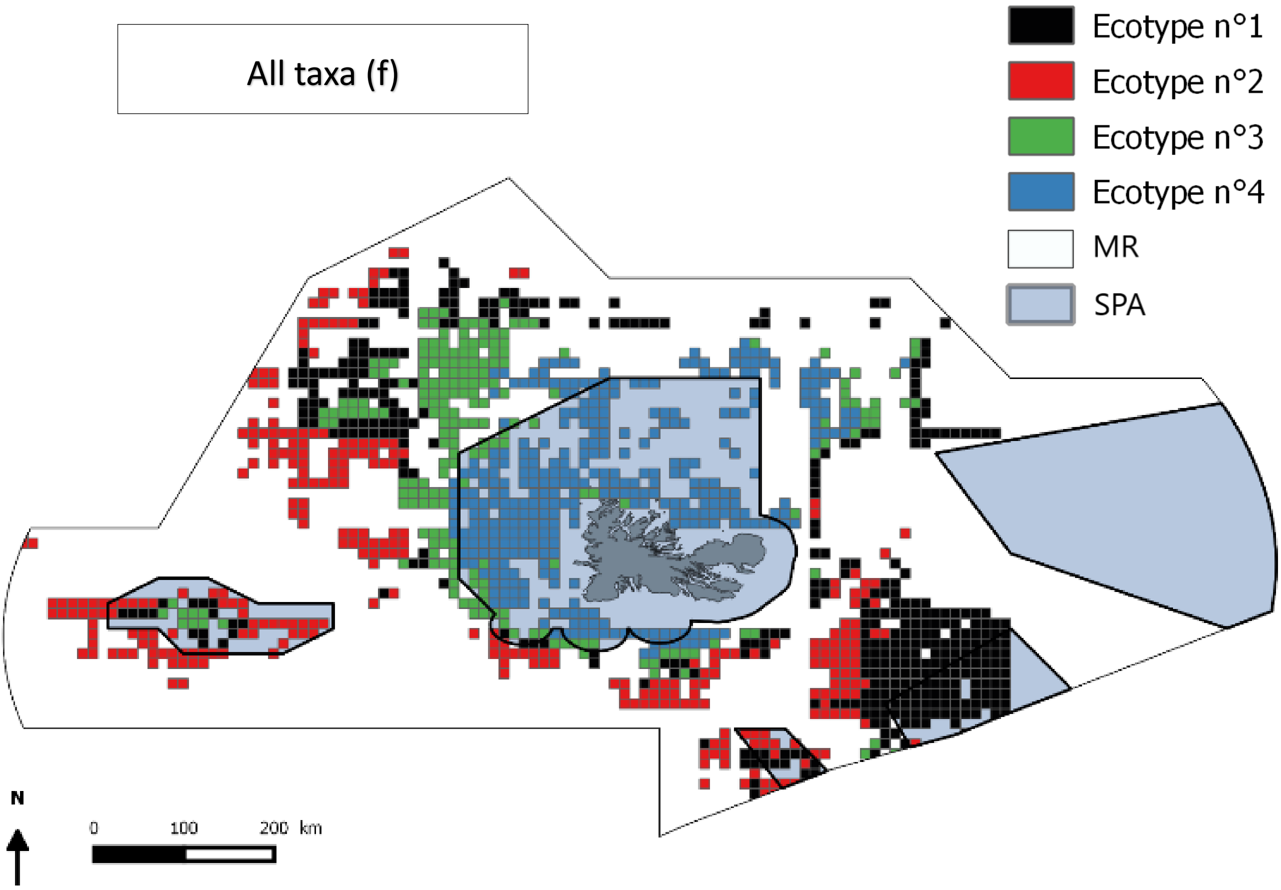

Fig. 4: benthic ecoregionalisation projected into maps showing ecotypes geographical distribution for Echinoidea (a), Holothuroidea (b), Mollusca (c), Ophiuroidea (d), Porifera (e), all taxa (f), vagile organisms (g) and mobile organisms (h), Marine Reserve (MR) and Strictly Protected Areas (SPA)

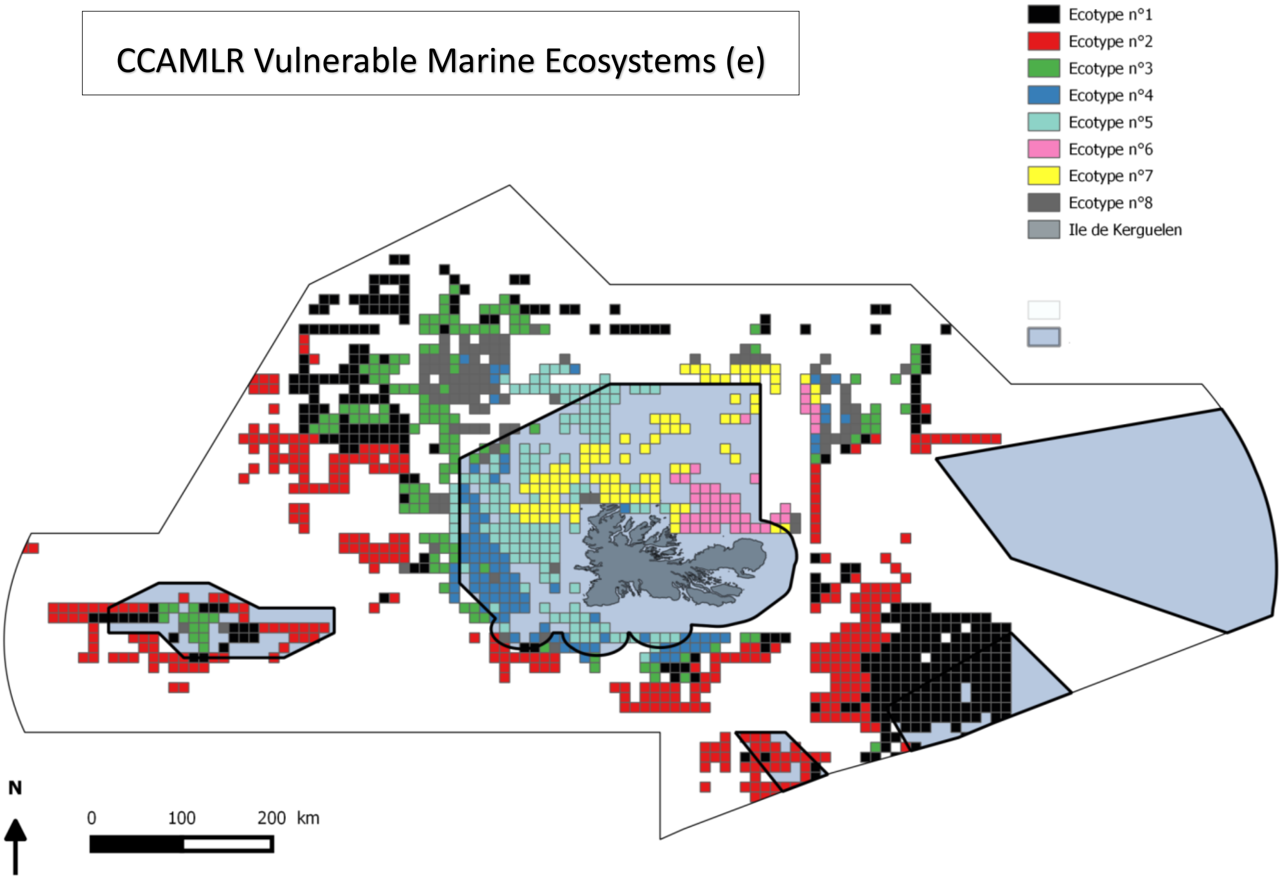

Fig. 5: benthic ecoregionalisation projected into maps showing ecotypes geographical distribution for detritivore organisms (a), necrophagous organisms (b), predator organisms (c), suspension feeders (d) and CCAMLR Vulnerable Marine Ecosystems indicators (e), Marine Reserve (MR) and Strictly Protected Areas (SPA)

Table 2: quality indices for each ecoregionalisation model: number of ecotypes, number of taxa supporting the modeling process, *Sih* index of the clustering, number of *IndVal* bioindicators and number of ecotypes including *IndVal* bioindicators for each output.

Group	**ecoregionalisation model**	**number of ecotypes**	**number of taxa**	***Sih* index**	**number of *IndVal* bioindicators**	**number of clusters including bioindicators**
Taxa	Actiniaria	4	7	0.7	0	0
	Annelida	3	2	0.84	2	2
	Arthropoda	3	10	0.73	4	3
	Ascidiacea	3	3	0.93	2	1
	Asteroidea	4	30	0.75	88	3
	Cnidaria	3	12	0.73	9	2
	Corals	5	5	0.87	0	0
	Echinodermata	4	35	0.69	143	4
	Echinoidea	3	4	0.69	2	2
	Holothuroidea	3	10	0.78	14	2
	Mollusca	3	10	0.88	0	0
	Ophiuroidea	4	8	0.68	0	0
	Porifera	4	3	0.87	2	1
Feeding mode	Detritivore	3	18	0.85	24	2
	Necrophagous	5	10	0.67	10	3
	Predator	6	37	0.73	275	6
	Suspension feeder	3	21	0.93	4	1
Motility	Sessile	5	26	0.7	45	3
	Vagile	8	52	0.7	41	3
Prot. status	CCAMLR VME	8	23	0.77	64	4
Complete	All Taxa	4	111	0.69	1179	4

Across all ecoregionalisation models number of ecotypes varies from 3 to 8 with a mode at 4 (tab. 2). For the taxa-based ecoregionalisation models, the number of ecotypes varies from 3 to 5, with a mode at 3 and a mean of 3.53. For the life-traits-based models, the number of ecotypes varies from 3 to 8, with a mode at 3 and a mean of 4.6. Maximum number of ecotypes is found when models are run using Vagile or CCAMLR VME groups.

Number of ecotypes across models is not related to the number of taxa used to fit the models. For example CCAMLR VME ecoregionalisation results in 8 ecotypes and is supported by 23 taxa (tab. 2).

In contrast, Asteroidea ecoregionalisation results in 4 ecotypes, supported by 30 taxa, and the complete ecoregionalisation, supported by 111 taxa, results in 4 ecotypes.

Number of taxa used in models varies from 2 for Annelida to 111 for Complete (tab. 2). Taxa-based models include from 2 (Annelida) to 35 (Echinodermata) taxa, and life history traits-based models include 10 (Necrophagous) to 52 (Vagile) taxa. The mean is 20.57 taxa across all models.

Sih index associated to each clustering varies from 0.67 (Necrophagous) to 0.93 (Ascidiacea and Suspension feeders), with a mean at 0.77 across all groupings (tab. 2).

For taxa-based models, the number of *IndVal* bioindicators varies from 0 to 143 combinations of taxa (tab. 2). The highest value is obtained for Echinodermata model. For life history traits-based models, the number of *IndVal* bioindicators varies from 24 combinations of taxa for the Detritivore ecoregionalisation to 275 for Predator. Model derived from Complete dataset includes an extreme value of 1179 combinations of taxa as *IndVal* bioindicators.

Number of *IndVal* bioindicators for each model and the number of ecotypes including *IndVal* bioindicators are not related to 1- number of taxa supporting the modelling process or 2- number of resulting ecotypes. For example, only 4 *IndVal* bioindicators could be computed in a unique ecotype for the suspension feeder ecoregionalisation, with 21 taxa used for the GDM and 3 ecotypes which could be characterized (tab. 2). This can be compared to the necrophagous ecoregionalisation, built with only 10 taxa, but including 10 combinations of taxa as *IndVal* bioindicators for 3 ecotypes of the model.

The number of clusters including *IndVal* bioindicators within each ecoregionalisation model varies from 0 to all the clusters of the output (tab. 2). For example, Predator model allowed to compute *IndVal* bioindicators for all the 6 ecotypes.

Finally, the most parcimonious and informative benthic ecoregionalisation model is obtained using CCAMLR VME grouping. This model uses 23 taxa and displays 8 ecotypes, with a *Sih* index of 0.77 equivalent to the mean obtained for the whole set of outputs (tab. 2). This model is retained for further analyses of the effect of environmental factors on biological dissimilarity, and to evaluate the relevance of the protection zones of the new Marine Reserve.

Projection onto a map of VME ecoregionalisation (fig. 5e) shows an homogeneous geographical distribution for each of the 8 benthic assemblages: no strong fragmentation can be observed, most of the pixels appearing to be located in continuous areas of pixels with same computed ecotype. This strong spatial structuration can be observed both on the plateau and the south-western “Skiff Bank” seamount.
(…)

4. Discussion

Quality of the models

Assessing the minimum number of biological records required to compute ecological models constitutes a major methodological issue, especially for data-poor areas such as freshwater and marine habitats (e.g. Bevilacqua et al., 2009; Fraschetti et al. 2011; Guillaumot et al. 2017; Ramos-Merchante and Prenda, 2017; Terlizzi et al., 2009; van Proosdij et al., 2016). Our study constitutes a contribution to solve this issue using an ecoregionalisation approach.

The quality indices for each ecoregionalisation model show that 1- number of ecotypes is independent from the number of taxa included in a given model and that 2- number of *IndVal* bioindicators and number of ecotypes including *IndVal* bioindicators are independent from both the number of sampled taxa and the number of ecotypes resulting from the modeling process.

In the light of this analysis, the biological indicator value of the benthic assemblages appears to be more efficient for ecoregionalisation and ecotypes characterization than the number and the diversity of sampled taxa. Indeed, increasing the quantity of biological observations by pooling together specialist and ubiquitous species, can result in a decrease of model statistical power. This is exemplified here by the highly informative VME model.

(…)

5. Conclusion

Our results show that increasing the number of taxa in models may have adverse consequences and may decrease the quality of analyses. Analyzing datasets restricted to relevant groupings results in more informative ecoregionalisation models. Here we show that VME-based ecoregionalisation unexpectedly appears more informative than the model based on the complete dataset. This finding is of particular interest because VME organisms are regularly collected by longliners and can constitute a source of relevant data for modeling benthic ecosystems at a larger scale including CCAMLR areas.

(…)