



Université de Rennes 1

Master 2 : Modélisation en Ecologie

**CARACTÉRISER LA CONTRIBUTION DES FACTEURS  
ENVIRONNEMENTAUX ET DES INTERACTIONS BIOTIQUES À LA  
VARIABILITÉ DES ASSEMBLAGES BENTHIQUES**

*CLÉMENT VIOLET*

*Laboratoire d'Ecologie Benthique Côtière, Centre IFREMER Brest*

Encadrants :

Dr Martin, *IFREMER, LEBCO, Brest*

Dr Aurélien, *IFREMER, LEBCO, Brest*

Dr Gauthier, *IUEM, LEMAR, Brest*

Dr Jacques, *IUEM, Observatoire Marin, Brest*

**Année 2019 — 2020**



UNIVERSITÉ DE  
**RENNES 1**



# Remerciements

Je souhaiterai remercier en premier lieu mon encadrant principal, Martin Marzloff pour avoir eu la gentillesse de m'accepter en stage à ses côtés. Je souhaite le remercier chaleureusement, lui et Aurélien Boyé pour avoir patiemment répondu à mes trop nombreuses questions tout au long de stage et pour les abondants commentaires qui ont grandement amélioré la qualité de ce manuscrit. Je remercie également Olivier Gauthier pour les heures de discussion passées à décortiquer les modèles dont il est question dans ce manuscrit, ces discussions m'ont permis d'encore approfondir mes connaissances en statistiques. Enfin Jacques Grall pour avoir répondu à mes questions sur la biologie et les interactions que peuvent entretenir ces animaux. Pour finir, je souhaiterai remercier tous les experts des polychètes que j'ai interrogés, sans qui une grande partie de ce travail n'aurait été possible.

# Notations mathématiques

## Indices

Indice & plage de valeurs	Description
$i = 1, \dots, n$	Unité d'échantillonnage
$j = 1, \dots, n_s$	Espèce
$r = 1, \dots, n_r$	Effet aléatoire
$l = 1, \dots, n_l$	Variable latente

## Vecteurs et matrices

Matrice	Description
$Y, y_{ij}$	Abondances observées
$\hat{Y}, \hat{y}_{ij}$	Abondances estimées
$m$	Abondances moyennes
$X, x_{ij}$	Données environnementales
$B, \beta_{ij}$	Coefficients associés à la niche environnementale
$Z, z_{ij}$	Variables latentes (coordonnées de sites)
$\Lambda, \lambda_{ij}$	Facteurs latents (coordonnées d'espèces)
$\Omega, \omega_{ij}$	Corrélations résiduelles interspécifiques

## Fonction et distribution statistique

Fonction / Distribution statistique	Description
$g(.)$	Fonction de lien d'un modèle linéaire généralisé
$\mathcal{L}$	Fonction de log-vraisemblance
$sgn(.)$	Fonction signe
$\mathcal{P}(.)$	Distribution de Poisson log-normale
$\mathcal{N}(.)$	Distribution normale

## Divers

Divers	Description
$y_{.j}$	Vecteur de l'abondance de l'espèce $j$ à travers tous les sites
$r_s$	Corrélation de Spearman

# Table des matières

<b>Remerciements</b>	<b>i</b>
<b>Notations mathématiques</b>	<b>i</b>
<b>Introduction</b>	<b>1</b>
Prédire la distribution d'une espèce . . . . .	2
Prédire la distribution et les dynamiques des assemblages d'espèces . . . . .	2
Comprendre le rôle des interactions biotiques . . . . .	3
Objectifs scientifiques . . . . .	4
<b>Matériel et méthode</b>	<b>5</b>
Cas d'étude . . . . .	5
Jeu de données . . . . .	5
Communauté faunistique . . . . .	6
Emprise spatiale et temporelle de l'étude . . . . .	6
Données environnementales . . . . .	7
Modèle de distribution d'espèces conjointes ( <i>JSDM</i> ) . . . . .	7
<i>Hierachical modelling of Species Communities</i> . . . . .	8
Modèle de Poisson Lognormal . . . . .	9
<i>Generalized linear latent variable models</i> . . . . .	9
Reconstruction des réseaux d'interactions . . . . .	10
Critères de comparaison des modèles implémentés . . . . .	10

Pouvoir explicatif . . . . .	10
Validation croisée . . . . .	11
Comparaison des réseaux inférés . . . . .	11
Validation par experts . . . . .	12
Matériel informatique logiciels utilisés . . . . .	13
<b>Résultats</b>	<b>13</b>
Comparaison des performances des modèles . . . . .	13
Pouvoir explicatif . . . . .	13
Prédiction de l'occurrence et de l'abondance . . . . .	14
Effets de l'environnement . . . . .	18
Réseaux reconstruits . . . . .	19
Analyse des graphes . . . . .	19
Avis des experts du taxon des polychètes . . . . .	22
Coût de calcul . . . . .	23
<b>Discussion</b>	<b>24</b>
Comprendre et prédire la diversité benthique . . . . .	24
Rôle et inférence des interactions biotiques . . . . .	26
Forces et faiblesses des différentes approches . . . . .	28
Perspectives . . . . .	29
<b>Bibliographie</b>	<b>30</b>
<b>Annexes</b>	<b>37</b>
Liste de <i>Species-Distribution Models</i> . . . . .	37
Partitionnement de la variance expliquée . . . . .	38
Espèces de polychètes . . . . .	40
Réseaux reconstruits . . . . .	43
<b>Résumé</b>	<b>47</b>

# Introduction

L'écologie moderne s'intéresse particulièrement à la répartition des espèces sur la surface du globe (Worm & Tittensor 2018) et à la réorganisation de la biodiversité face aux changements environnementaux en cours (Blowes *et al.* 2019; Magurran *et al.* 2019). Une multitude d'approches théoriques et empiriques a ainsi été développée pour comprendre les facteurs qui gouvernent la biogéographie des espèces et prédire les patrons spatiaux et temporels de biodiversités.

La capacité de persistance d'une espèce peut être définie grâce au concept de niche écologique tel que proposé par Hutchinson (Hutchinson 1957). Ainsi, la niche d'une espèce est représentée selon un hypervolume à n-dimensions où chaque dimension correspond à une ressource ou une condition environnementale (Blonder 2017) et permet de décrire les besoins ou la gamme de tolérance physiologique des organismes (Sunday *et al.* 2012). Cette niche de Hutchinson peut être divisée en deux niches : (1) la niche de Grinnell qui s'intéresse plus particulièrement aux contraintes liées aux conditions abiotiques et caractérise la niche fondamentale de l'espèce (Grinnell 1917); (2) La niche d'Elton qui se focalise, principalement sur les interactions biotiques entre les individus (Elton 2001). En considérant le rôle des interactions biotiques en complément des contraintes environnementales, la niche d'Elton permet de décrire la niche réalisée des espèces, qui est en réalité celle observée *in situ*.

Une des manières d'avoir accès à la niche environnementale d'une espèce passe par l'expérimentation afin de délimiter des gammes et seuils de tolérances physiologiques des organismes (Bates & Morley 2020). Néanmoins, pour des contraintes pratiques, ces expériences peuvent difficilement prendre en compte les effets combinés de plusieurs facteurs abiotiques, et se concentrent souvent sur une seule voire deux variables environnementales (par exemple la température et la disponibilité en nourriture, ou encore le pH pour les études marines ciblées sur les changements en cours dans l'océan (Boyd *et al.* 2000; Witman *et al.* 2015)). Par ailleurs, ces expérimentations en mésocosme concernent principalement des fines échelles spatiales, temporelles et écologiques (*i.e* échelle de l'individu plutôt que de la population ou de la communauté (Witman *et al.* 2015)). Ainsi, l'échelle de ces études expérimentales ne permet pas d'appréhender pleinement l'ensemble des phénomènes écologiques structurant à l'échelle de tout un écosystème (Leibold & Chase 2018). Dans un contexte de changements globaux rapides induits par l'Homme, il est nécessaire d'accéder à des échelles spatiales, temporelles et écologiques plus larges (Witman *et al.* 2015; Estes *et al.* 2018), qui soit plus en lien avec les échelles auxquelles la société bénéficie, gère et affecte les écosystèmes (Isbell *et al.* 2017). La modélisation écologique et notamment les modèles de distribution d'espèces, fournissent un outil privilégié pour cela (Edgar *et al.* 2016).

## Prédire la distribution d'une espèce

Parmi les approches permettant d'étudier la répartition spatiale d'espèces, les modèles de distributions d'espèces (*Species Distribution Models; SDM*) constituent sans conteste l'un des outils les plus utilisés sur ces vingt dernières années (Guisan *et al.* 2017; Melo-Merino *et al.* 2020). Les *SDM* reposent sur l'utilisation de variables environnementales pour expliquer et prédire la répartition des espèces, la niche de Grinnel constituant "la colonne vertébrale des modèles de distribution d'espèce" (Gravel *et al.* 2018).

La première implémentation de *SDM* est apparue à la fin des années 80 avec la méthode *BIOCLIM* (Busby 1991). Le développement de ce type d'approche de modélisation a notamment été favorisé par les progrès informatiques qui les rendent faciles à mettre en oeuvre (Araújo *et al.* 2019; Ovaskainen & Abrego 2020). Ainsi, des *SDM* ont utilisé des modèles linéaires généralisés (Hastie & Pregibon 1992), des modèles linéaires additifs (Wood 2011), ou bien encore plus récemment des modèles de machine learning basée sur des arbres (Elith *et al.* 2008)<sup>1</sup>. Ces modèles permettent uniquement de modéliser la distribution d'une seule espèce à la fois, ils sont qualifiés de *Single-species Distribution Model*.

De plus, si ces méthodes permettent de caractériser la niche potentielle des espèces, la niche réalisée et notamment l'aspect eltonien de la niche des espèces reste plus difficile à appréhender (Hortal *et al.* 2015; Dehling & Stouffer 2018). Comprendre le rôle des interactions biotiques dans la distribution des espèces est souvent un facteur limitant nos capacités à prédire les patrons spatiaux de biodiversité.

## Prédire la distribution et les dynamiques des assemblages d'espèces

Si la question de recherche s'intéresse à la communauté dans son entiereté, il existe deux stratégies différentes pour modéliser la distribution de multiples espèces. Il est possible de prédire la distribution des espèces indépendamment et de les assembler en communauté ensuite (*Stack Species Distribution Model* ou *SSDM*), ou bien de prédire la distribution des espèces conjointement (*Joint Species Distribution Model*) (Ferrier & Guisan 2006).

A l'inverse de l'approche climatologique largement appliquée dans le cas des *SDM* ou des *SSDM* qui vise à caractériser la répartition biogéographique des espèces, en ne considérant que très rarement les dynamiques temporelles de manières explicites, les *JSDM* ont pour objectif de caractériser et de prédire la variabilité spatio-temporelle des communautés. Ils permettent par ailleurs d'intégrer des informations spécifiques additionnelles afin de tirer profit des similitudes entre espèces à l'échelle de la communauté, notamment via les traits fonctionnels ou encore la phylogénie (Tableau 1).

<sup>1</sup>Un tableau listant quelques une des méthodes de *SDM* actuellement utilisées sont disponible dans le Tableau 7 en annexe.

De plus, les *SSDM*, en combinant des modèles faits indépendamment pour chaque espèce, ne tiennent pas compte du filtre biotique que les espèces exercent les unes sur les autres et négligent donc l'aspect eltonien de leur niche (Zurell *et al.* 2019). Au contraire, les *JSDM* ont été développés en prenant en compte les patrons de co-occurrence à l'échelle d'une communauté afin de modéliser conjointement la distribution de chaque espèce au sein de toute la communauté en tenant compte des filtres abiotiques (Warton *et al.* 2015). Cette structure de co-occurrence fournit alors des informations potentielles sur les interactions entre espèces (Ovaskainen *et al.* 2017a).

Tableau 1 : Quelques uns des modèles de *JSDM* et leurs fonctionnalités (adapté de Ovaskainen & Abrego (2020)). Les colonnes “traits fonctionnels” et “phylogénie” indiquent si les modèles prennent en compte ces données de façon explicite.

<i>Joint species distribution model</i>	Effets spatiaux	Effets temporels	Traits fonctionnels	Phylogénie	Référence
<i>Bayesian ordination and regression analysis (BORAL)</i>	Implicite	Implicite	Oui	Non	Hui (2016)
<i>Hierachical modelling of Species Communities (HMSC)</i>	Explicite	Explicite	Oui	Oui	Ovaskainen <i>et al.</i> (2017b)
<i>Generalized linear latent variable models (GLLVM)</i>	Implicite	Implicite	Oui	Non	Niku <i>et al.</i> (2019b)
<i>Poisson lognormal model (PLN)</i>	Implicite	Implicite	Non	Non	Chiquet <i>et al.</i> (2019)

## Comprendre le rôle des interactions biotiques

Bien que le rôle des interactions biotiques dans la structure des communautés et la coexistence des espèces est depuis longtemps reconnu (Hardin 1960; Paine 1966), leur influence à larges échelles a pendant longtemps été négligée. Ainsi, l'explication des patrons biogéographique de diversité s'est concentrée sur le rôle des filtres abiotiques (Tittensor *et al.* 2010; Soininen 2014). La contribution des interactions biotiques aux patrons de biodiversité à larges échelles est à ce jour mal appréhendée (Wisz *et al.* 2012). Leur importance est néanmoins variable dans le temps, dans l'espace et selon les systèmes biologiques (Leibold & Chase 2018). Par exemple, la compétition pour l'espace et la prédation sont localement des facteurs extrêmement structurants pour les communautés benthiques des estrans rocheux. La compétition entre les balanes et les moules est un exemple classique en écologie intertidale de compétition pour le substrat (Hawkins *et al.* 2019). De même, le rôle de la prédation des étoiles de mer sur les balanes et sur les moules est à l'origine du concept d'espèce « clé de voute » (Paine 1966). A contrario, les plages de sable sont moins structurées par les interactions que les estrans rocheux et sont généralement plus structurées par les forces hydrodynamiques et autres facteurs abiotiques (Defeo & McLachlan 2005; Woodin *et al.* 2019). Clarifier le rôle relatif

des interactions biotiques par rapport aux filtres environnementaux dans différents systèmes et à différentes échelles paraît alors primordial.

Observer et quantifier les interactions dans les écosystèmes marins est une tâche difficile. Pour les écosystèmes pélagiques, il est possible d'utiliser différentes techniques comme l'observation directe, l'étude des contenus stomacaux ou bien l'analyse isotopique d'acides gras (Majdi *et al.* 2018). Lorsque ces méthodes sont difficiles à mettre en place, comme c'est le cas pour les écosystèmes benthiques aux larges échelles qui nous intéressent ici, Majdi *et al.* (2018) proposent d'utiliser des méthodes d'inférences statistiques. Ainsi, les *JSDM* peuvent être des outils intéressants pour mieux comprendre les interactions biotiques des écosystèmes intertidaux benthiques.

## Objectifs scientifiques

Différentes implémentations de ces modèles conjoints de distribution d'espèces (Tableau 1) ont vu le jour ces dernières années sous l'influence forte de théorie issue de l'écologie végétale concernant le rôle des interactions dans la structuration des communautés de plantes à larges échelles spatiales et temporelles (Warton *et al.* 2015). De récents travaux se sont intéressés à comparer certaines de ces méthodes. Norberg *et al.* (2019) ont comparé trente-trois modèles de distributions d'espèces sur des jeux de données de présence/absence de communautés végétales. Momal *et al.* (2020) quant à eux ont comparé la capacité de reconstruction de réseaux d'interaction à partir de résultats de modèles *JSDM* entraînés sur des données d'abondance d'une communauté de poissons et sur des données de métabarcoding issu d'une communauté fongique. Niku *et al.* (2019b) quant à eux présentaient le *JSDM* qu'ils ont développé, et son application à une communauté de fourmis.

Le but global de ce travail est d'explorer le potentiel de différentes méthodes de *JSDM* pour expliquer et prédire la diversité benthique. Ainsi, ce travail vise à exploiter 8 ans de données de suivi côtier (programme *REBENT* ; [www.rebent.fr](http://www.rebent.fr)) afin de caractériser l'influence des processus biotiques et abiotiques dans la variabilité des communautés de macrofaune benthique. En effet, la structure des *JSDM* nous permet de quantifier l'influence des facteurs abiotiques sur les patrons spatiotemporels de biodiversité benthique et d'appréhender les interactions biotiques à l'échelle de la communauté. Pour ce faire, nous allons nous concentrer sur trois implantations différentes des *JSDM* : l'un des meilleurs cadres de modélisation décrits par Norberg *et al.* (2019) (*HMSC* pour *Hierarchical Modelling of Species Community*), celui utilisé par Momal *et al.* (2020) (*PLN* pour *Poisson Lognormal Model*) et celui développé par Niku *et al.* (2019b) (*GLLVM* pour *Generalised Linear Latent Variable Model*). L'application de ces trois méthodes sur les mêmes données d'abondance faunistiques permettra par la même occasion de comparer la performance de ces modèles, puisqu'à notre connaissance aucune étude n'a comparé ces trois méthodes lorsqu'elles sont appliquées aux écosystèmes benthiques. L'innovation de cette étude est donc double : effectuer la première comparaison simultanée de ces trois modèles, et effectuer une des premières approches de *JSDM* sur de la faune benthique inféodée aux habitats meubles.

# Matériel et méthode

## Cas d'étude

### Jeu de données

Le jeu de donnée utilisé dans ce travail est issu du *REseau de surveillance BENthique*. Ce réseau de surveillance a été créé en 1999 à la suite du naufrage du pétrolier Erika. Le but de ce réseau est d'acquérir des connaissances sur les habitats benthiques côtiers et de détecter des changements de la diversité biologique de ces habitats (Derrien-Courtel *et al.* 2017). Le protocole de suivis se concentre sur la macrofaune (> 1mm) et la méthodologie détaillée est présentée dans Boyé *et al.* (2019). Seulement deux de ces habitats ont été retenus pour comparer les trois méthodes de modélisations : les herbiers de zostères et les sédiments meubles nus.

Les zostères sont des phanérogames marines qui lorsque leur densité est forte, créent des “herbiers”. En Bretagne, il existe deux espèces de zostères : la zostère marine (*Zostera marina*) et la zostère naine (*Zostera noltii*). Ces herbiers forment des habitats pour de nombreuses espèces. Ils ont un rôle fonctionnel important pour beaucoup d’espèces en tant que zone de nurserie, de reproduction et de nourrissage (McDevitt-Irwin *et al.* 2016 ; Lefcheck *et al.* 2019). Cet habitat est connu pour sa plus grande richesse spécifique par rapport à d’autres habitats comme les sédiments meubles nus (Hily & Bouteille 1999 ; Sunday *et al.* 2016 ; Boyé *et al.* 2017). Le suivi du *REBENT* se concentre sur les herbiers de *Z. marina*, car cette espèce fait déjà partie du livre rouge des espèces menacées (Waycott *et al.* 2009 ; Derrien-Courtel *et al.* 2017). Toutefois, les mesures de protections mises en place vis-à-vis de cet habitat semblent porter leurs fruits en Europe (Santos *et al.* 2019).

Les plages de sable abritent des communautés dominées par les crustacés les mollusques et les polychètes (Defeo & McLachlan 2005). Cet habitat est contraint principalement par trois facteurs abiotiques : les marées, la houle et le sédiment. L’interaction entre ces trois facteurs crée un grand éventail de plages différentes : de la plage où la mer ne se retire que très peu à marée basse aux plages dont de vastes étendues de sable sont découvertes lors des grandes marées (Defeo & McLachlan 2005). La richesse spécifique sur ces plages est fortement influée par la hauteur des marées et dans une moindre mesure par la température de l’eau et la taille des sédiments (Defeo & McLachlan 2013 ; Defeo *et al.* 2017). Ces plages de sable fin ont aussi un rôle écologique important de nurserie pour de nombreuses espèces de poissons plats (Quillien *et al.* 2017). Toutefois, les plages de sable sont des biocénoses menacées. Ce sont des habitats à la fois menacés par les changements globaux (hausse du niveau des mers, érosion...) et par les activités anthropiques telles que l’urbanisme ou bien les activités récréatives (Defeo *et al.* 2009).

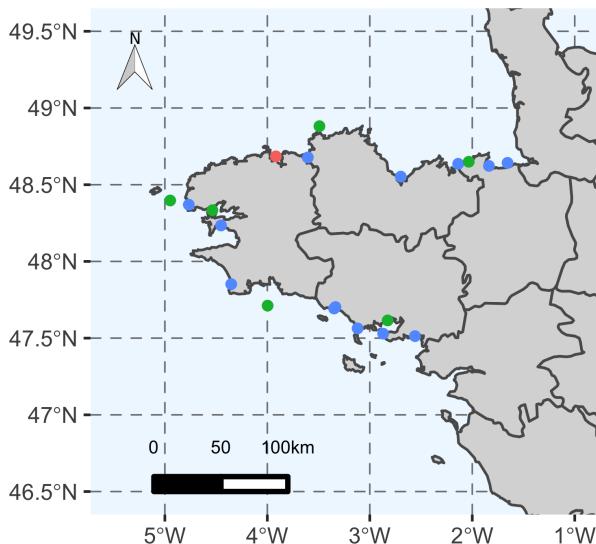
## Communauté faunistique

Ce travail se concentre sur les assemblages de polychètes, du phylum des annélides (Lecointre & Le Guyader 2001). Cette classe a été choisie, car composée de taxons aux modes de vie très divers, dont l'écologie a déjà fait l'objet d'un travail approfondi (Boyé *et al.* 2019). Il est classique de catégoriser les polychètes en plusieurs groupes en fonction de leurs modes de vie : certains sont nageurs, ou tubicoles, ou bien encore benthiques fouisseurs. Leurs régimes alimentaires sont aussi très variés, il existe des polychètes prédateurs, mais également des brouteurs, des suspensivores ou psamivores (Jumars *et al.* 2015). Toutefois, la taxonomie et l'écologie de ces animaux restent assez mal connues par la communauté scientifique (Troudet *et al.* 2017). En particulier, les interactions écologiques (compétition pour l'espace ou la nourriture, voire prédation) au sein de ces communautés de polychètes laissent largement méconnues.

## Emprise spatiale et temporelle de l'étude

Vingt-trois sites le long de la façade bretonne ont été retenus, représentant 215 échantillons collectés sur une période s'étalant entre 2006 et 2014. Les données considérées contiennent 96 taxa différents de polychètes (et le jeu de données d'apprentissage ne contient toutefois que 92 taxa en raison de l'exclusion des 3 sites de validation). Il est important de noter que certains sites présentent deux habitats (*i.e.* herbiers, sédiments nus) associés aux mêmes conditions environnementales (étant donné la faible résolution spatiale des variables environnementales). Afin de tester la performance des modèles, vingt sites ont été retenus pour l'entraînement des différents modèles et trois pour la validation croisée (fig. 1).

1



2

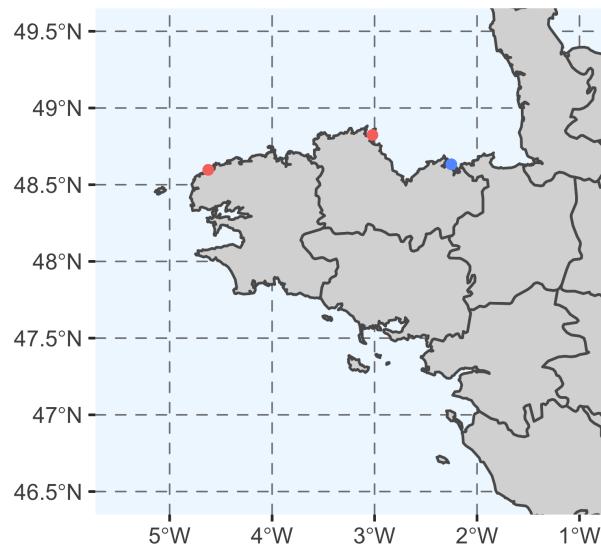


Figure 1 : Carte des sites échantillonnés. Les points bleus et verts représentent respectivement les sites à sédiments meubles et les herbiers. Les points rouges représentent les sites ayant les deux habitats. 1. Sites utilisés pour entraîner les modèles. 2. Sites utilisés pour valider les modèles

## Données environnementales

Six variables environnementales, collectées lors de travaux précédents sur ces communautés (Boyé *et al.* 2019) ont été sélectionnées pour expliquer la variabilité de ces communautés (tableau : 2). La salinité, la température et la vitesse des courants proviennent de la base de données publique *PREVIMER* basée sur les résultats des modèles de *MARS3D*. Le fetch a été calculé à partir des polygones terrestres disponibles dans *OpenStreetMap*. Les variables granulométriques ont été échantillonnées *in situ* (protocole détaillé dans Boyé *et al.* (2017)). Il est important de noter que, dans le but d'utiliser des polynômes pour modéliser des relations non linéaires, l'ensemble des variables explicatives ont été centrées et transformées via la fonction *poly*. Cependant, étant donné le grand nombre d'échantillons, nous avons été contraints d'utiliser des polynômes de degré un, ce qui revient à modéliser une simple relation linéaire, afin de réduire le temps de calcul.

Tableau 2 : Variables environnementales utilisées dans l'ensemble des modèles.

Abréviation	Définition	Unité
Fetch	Fetch moyen	Km
MO	Concentration en Matière organique	%
SAL_sd	Ecart-type de la salinité de l'eau	PSS-78
TEMP_sd	Ecart-type de la température de l'eau	°C
CURR_mean	Force moyenne des courants	m.s <sup>-1</sup>
Mud	Concentration de boue dans les sédiments	%
Trask_So	Indice de Trask - Homogénéité du sédiment	Aucune

## Modèle de distribution d'espèces conjointes (*JSDM*)

Bien que chaque implémentation des *JSDM* soit différente et permettent de prendre en compte différents types de données (Tableau 1), ces modèles sont tous des extensions multivariées à variables latentes des modèles linéaires généralisés classiques (Hui 2016; Ovaskainen *et al.* 2017b; Chiquet *et al.* 2019; Niku *et al.* 2019b). Un modèle à variables latentes (*LVM*) avec une variable explicative peut être écrit comme suit<sup>2</sup> :

$$y_{ij} = g(x_i \times \beta_j + Z_{i.} \times \lambda_j.) \quad (1)$$

Avec  $y_{ij}$  représentant l'abondance prédictive de l'espèce  $j$  au site  $i$ .  $g(.)$  une fonction de lien ;  $x_i$  le vecteur de la variable environnementale au site  $i$ ,  $\beta_j$  le coefficient de l'effet environnemental associé pour l'espèce  $j$ . Enfin, avec  $Z_{i.}$  est la matrice de variable latente associée aux sites et qui joue le rôle des prédicteurs manquants et  $\lambda_j$  la matrice de poids associés à l'espèce  $j$  qui donnent une estimation des

<sup>2</sup>L'ensemble des notations mathématiques sont présentées dans le glossaire.

corrélations résiduelles entre espèces  $\Omega$ , avec (Warton *et al.* 2015) :

$$(Z_i \times \lambda_j) \sim \mathcal{N}(0, \Omega) \quad (2)$$

Outre la prise en compte via les variables latentes d'éventuelles variables explicatives manquantes et des corrélations résiduelles entre espèces. L'avantage majeur des *LVM* est que l'estimation de matrice de corrélation entre espèces est plus simple que par rapport à un modèle linéaire à effets mixtes généralisés. La matrice de coordonnées des espèces ( $\lambda$ ) dans le cas d'un *LVM* a au plus autant de colonnes que de variables latentes (Équation 2). Dans le cas d'un modèle linéaire à effets mixte généralisé celle-ci à autant de colonnes que d'espèces (Warton *et al.* 2015). Ainsi, le nombre de variables latentes utilisées par le modèle est donc un paramètre crucial, puisqu'il permet de faire un compromis entre précisions de la matrice de corrélation résiduelle et la réduction du temps de calcul et des degrés de liberté utilisés (Warton *et al.* 2015). Trois cadres de modélisations ont été sélectionnés : *HMSC*, *GLLVM* et *PLN* et exploré cinq variantes de modèles ces modèles. L'ensemble des caractéristiques de ces variantes de modèles sont présentées dans le Tableau 3 .

### ***Hierarchical modelling of Species Communities***

*HMSC* est un modèle mixte linéaire généralisé, hiérarchique et multivarié, ajusté par inférence bayésienne (Ovaskainen & Abrego 2020). La particularité de ce modèle est qu'il est hiérarchique, ainsi chaque effet aléatoire donne lieu à sa propre matrice de corrélation résiduelle (Ovaskainen *et al.* 2017b) et qu'il peut s'implémenter avec une vaste gamme de distribution (Tikhonov *et al.* 2019a). L'Équation 3 présente la formulation mathématique d'un modèle n'utilisant que des variables environnementales et un nombre  $n_r$  d'effets aléatoires (il serait aussi possible d'ajouter des données de traits et de phylogénie pour mieux caractériser les niches abiotiques des taxa similaires).

$$y_{ij} = g\left(x_i \times \beta_j + \sum_{r=1}^{n_r} Z_{ir} \times \lambda_{rj}\right) \quad (3)$$

Etant un modèle bayésien, la distribution a posteriori est échantillonnée grâce à la méthode MCMC (Ovaskainen *et al.* 2017a). L'utilisation de l'inférence bayésienne permet à l'utilisateur de ne pas spécifier le nombre de variables latentes à utiliser pour chaque effet aléatoire. Le modèle ajuste le nombre de variables latentes pour que celles non significatives soient tronquées (Ovaskainen & Abrego 2020).

Trois modèles *HMSC* ont été créés : un premier sans l'inclusion d'effets aléatoires (*HMSC\_reg*) qui ne prend donc pas en compte les corrélations résiduelles entre espèces et l'aspect eltonien de leur niche; un second avec un seul effet aléatoire lié à l'échantillon (*HMSC\_samp*) et un dernier avec trois facteurs aléatoires liés respectivement à l'année, au site et à l'habitat. Ce dernier modèle permet

ainsi de partitionner les associations résiduelles entre association temporelle, spatiale ou liée à l'habitat (sédiment nu ou herbiers). Les calculs ont été réalisés avec le package *Hmsc* (Tikhonov *et al.* 2019a, b). Chaque modèle dispose de quatre chaînes de Markov et chaque chaîne effectue 1,5 million d'itérations avant de se stopper. L'étape de *burn-in* supprime les 500 000 premières itérations de chaque chaîne. Les chaînes sont échantillonnées toutes les mille itérations. Les priors par défaut ont été utilisés. Avant de regarder les résultats des différents modèles du framework *HMSC*, la validité de ces modèles a été inspectée. La bonne convergence des chaînes a été vérifiée à l'aide de l'outil de diagnostic de Gelman-Rubin (Gelman & Rubin 1992). Enfin, le nombre d'échantillons indépendants pour chaque paramètre a été jugé satisfaisant (supérieur à 1000 échantillons indépendants pour l'ensemble des paramètres).

### **Modèle de Poisson Lognormal**

Le modèle de Poisson lognormal est linéaire mixte et multivarié. Son implémentation ne permet de modéliser qu'une seule sorte de distribution : la distribution (conditionnelle) de Poisson lognormale. Équation 4. Un modèle simple avec uniquement des variables environnementales peut être écrit de cette façon :

$$y_{ij}|Z_{ij} \sim \mathcal{P}(\exp\{x_i \times \beta_j + Z_{ij}\}) \quad (4)$$

Le vecteur latent  $Z_i$  prend en compte les variations d'abondance non expliquées par les variables environnementales incluses dans le modèle (Momal *et al.* 2020). Cette variable latente agit comme un effet aléatoire lié à l'échantillon (Momal *et al.* 2020). Dans ce cadre de modélisation, il y a autant de variables latentes différentes qu'il y a d'espèces  $n_s$  et chaque variable latente suit une loi normale de moyenne 0 et de variance  $\Omega^{-1}$ , où  $\Omega$  est la matrice de corrélations résiduelles entre espèces :

$$Z_i \sim N(0_{n_s}, \Omega^{-1}) \quad (5)$$

Le modèle a été créé avec le package R *PLNmodels* (Chiquet *et al.* 2019). Pour comparer avec les autres implémentations des JSDM testées ici, aucun terme d'offset n'a été ajouté au modèle. Les paramètres par défaut ont également été utilisés pour ajuster le modèle.

### ***Generalized linear latent variable models***

*GLLM* est un modèle mixte linéaire généralisé et multivarié ajusté par la méthode du maximum de vraisemblance qui peut être écrit par l'Équation 1. Ce modèle présente une manière innovante de maximiser la vraisemblance en utilisant une approximation variationnelle gaussienne de la log-vraisemblance pour le cas où la fonction de lien serait des données de comptage surdispersé, binaires ou encore ordinaires (Niku *et al.* 2019b). Cette méthode de maximisation de la log-vraisemblance per-

met d'accélérer les calculs. Par exemple, comparativement à une autre implémentation, *Boral* (Hui 2016), d'un même modèle, *GLLVM* ne requiert que quelques minutes au lieu de quelques heures. En revanche, et contrairement à *HMSC* qui l'estime ou *PLN* qui a une valeur fixe, *GLLVM* demande à l'utilisateur de choisir le nombre de variables latentes qu'il utilisera le modèle (Niku *et al.* 2019a).

Le modèle a été créé avec le package R *gllvm* (Niku *et al.* 2019a, b). Le modèle utilise une distribution négative binomiale et vingt variables latentes.

Tableau 3 : Descriptif des modèles utilisés. \* Lorsque des effets aléatoires sont définis pour les modèles *HMSC*, un nombre infini de variables latentes peut être généré, seules les variables latentes importantes sont conservées (voir Bhattacharya & Dunson (2011) pour plus d'informations).

Nom du modèle	Frame-work	Distribution statistique	Nombre de facteur aléatoire	Nombre de variables latentes
<i>HMSC_reg</i>	<i>HMSC</i>	Poisson lognormal	0	0
<i>HM-</i>	<i>HMSC</i>	Poisson lognormal	1	$n_l \in \mathbb{N}^*$
<i>SC_samp</i>				
<i>HMSC_hier</i>	<i>HMSC</i>	Poisson lognormal	3	$n_l \in \mathbb{N}^*$
<i>PLN</i>	<i>PLN</i>	Poisson lognormal	1	$n_l = 92$
<i>GLLVM</i>	<i>GLLVM</i>	Negative binomial	0	20

## Reconstruction des réseaux d'interactions

Afin d'appréhender les processus écologiques structurant les communautés de polychètes, des réseaux d'interactions potentielles ont été reconstruits grâce au package *EMtree* à partir des matrices de corrélations résiduelles issues de chaque modèle. Le principe de l'algorithme contenu dans ce package est d'inférer des interactions entre espèces en utilisant des arbres couvrants (graphes connectant tous les noeuds sans aucune boucle). La probabilité qu'un lien fasse partie du réseau d'intérêt est égale à la somme des probabilités conditionnelles associées à chaque arbre couvrant (Momal *et al.* 2020).

## Critères de comparaison des modèles implantés

### Pouvoir explicatif

Le pouvoir explicatif de chaque modèle pour chaque taxon est donné par une mesure de pseudo- $R^2$ , dénommée après  $SR^2$ . Ainsi, pour la distribution de Poisson, *HMSC* estime le  $SR^2$  comme la corrélation de Spearman entre les données d'abondance observées et prédictes (Ovaskainen & Abrego 2020). Cette mesure est calculée de la façon suivante pour l'espèce  $j$  :

$$SR_j^2 = sgn(r_s(y_{.j}, \hat{y}_{.j})) \times r_s(y_{.j}, \hat{y}_{.j})^2 \quad (6)$$

Où  $sgn$  est la fonction signe,  $\hat{y}$  est l'abondance estimée et  $r_s$  la corrélation de Spearman. Cette même mesure a été appliquée sur l'ensemble des autres modèles.

## Validation croisée

Parmi les vingt-trois sites étudiés, trois ont été utilisés pour effectuer de la validation croisée sur les modèles. Deux de ces trois sites présentent les deux habitats et le dernier présente uniquement des sédiments meubles. La performance des modèles a été comparée sur deux critères : la prédiction de l'abondance de chaque espèce et la prédiction de leur occurrence.

**Prédiction de l'abondance** Chaque modèle entraîné a été utilisé pour prédire l'abondance des espèces présentes dans les trois sites de validation. La qualité de la prédiction a été évaluée par la racine carrée de l'erreur quadratique moyenne (*RMSE*).

**Prédiction de l'occurrence** Pour prédire l'occurrence de chaque espèce à partir des données d'abondance prédites par les modèles, deux approches ont été suivies. Dans un premier temps, une espèce a été considérée présente si son abondance prédite était supérieure à 0. Cependant, étant donné les incertitudes existantes autour des valeurs d'abondances prédites (cf. partie résultat sur l'abondance), il a été décidé d'établir dans un second temps un seuil limite, modèle et espèce spécifique, pour convertir les données d'abondance prédites en présence/absence, de manière analogue aux seuils définis dans les *SDM* pour convertir les cartes de probabilité de présence en carte de présence/absence (Ashcroft *et al.* 2017; Guisan *et al.* 2017). Afin de définir ce seuil, une première étape a consisté à créer pour chaque espèce une courbe *ROC* (de l'anglais *Receiver Operating Characteristic*) basée sur le nombre de vrais et faux positifs de l'occurrence selon différents seuils d'abondance. Puis, pour chaque courbe *ROC*, le *J* de Youden a été calculé grâce à l'Équation 7 et la valeur maximale de cet indice a donné le seuil que nous avons appliqué pour déterminer la présence prédite de chaque espèce. Le *J* de Youden a été sélectionné, car cette statistique maximise le nombre de vrais positifs et de vrais négatifs, tout en donnant le même poids aux faux positifs et faux négatifs. Sa valeur maximale est de 1, lorsqu'il n'y a pas de faux positifs ou négatifs et de 0 lorsque qu'il y a la même proportion de faux positifs et négatifs que de vrais positifs et négatifs.

$$J = \frac{\text{vrais positifs}}{\text{vrais positifs} + \text{faux négatifs}} + \frac{\text{vrais négatifs}}{\text{vrais négatifs} + \text{faux positifs}} - 1 \quad (7)$$

## Comparaison des réseaux inférés

Les réseaux reconstruits sont comparés à l'aide de métriques issues de la théorie des graphes appliquée aux réseaux probabilistes (Poisot *et al.* 2015). Ces outils sont issus du package EcologicalNetworks.jl (Poisot *et al.* 2019). Quatre métriques ont été sélectionnées : le nombre de liens, la connectance, la centralité de Katz et l'emboitement. Le nombre de liens et la connectance sont des pro-

priétés fondamentales de la structure des réseaux trophiques et des réseaux d'interactions (Martinez 1992). Dans un réseau probabiliste, le nombre de liens est calculé grâce à la somme des probabilités sur l'ensemble du réseau ( $l = \sum A_{ij}$ ), il ainsi possible de calculer la variance du nombre de liens ( $var(l) = \sum A_{ij}(1 - A_{ij})$ ). Par exemple, la résistance d'un réseau écologique aux perturbations est par proportionnelle à sa connectance. La connectance est le nombre de liens  $l$  divisé par le nombre de liens possibles ( $S \times (S - 1)/2$ ). Cette mesure comprise entre 0 et 1 peut être associée à la résistance globale d'un réseau écologique aux perturbations. La centralité de Katz permet de connaître l'importance de chaque espèce dans le réseau. L'emboîtement des interactions est un indice qui traduit à quel point les interactions réalisées par les espèces spécialistes (celles impliquées dans peu d'interactions) est un sous-ensemble de celles réalisées par les espèces généralistes (impliquées dans beaucoup d'interactions). Cette mesure est comprise entre 0 et 1. Cette mesure de l'imbrication est importante pour comprendre la dynamique d'un réseau d'interaction, car l'hypothèse a été faite que des réseaux fortement imbriqués promeuvent une plus grande diversité en minimisant la compétition en particulier dans les réseaux mutualistes (Bastolla *et al.* 2009).<sup>3</sup>

Pour tous les modèles incluant au moins une variable latente (tous sauf *HMSC\_reg*), un réseau a été reconstruit grâce à la matrice de corrélation résiduelle  $\Omega$ . Pour la méthode *HMSC\_hier*, un réseau a été reconstruit pour chaque effet aléatoire, chacun ayant sa propre matrice. Un méta-réseau a également été créé en moyennant les probabilités sur l'ensemble des méthodes. L'ensemble des différents réseaux sont représentés en annexe (Figure 13 à 17).

## **Validation par experts**

Douze paires d'interactions potentielles ont été soumises au jugement d'un panel de cinq experts des polychètes. Les experts ont évalué la vraisemblance de chaque interaction sur une échelle de 0 à 4 :

- 0. Ces deux taxa n'interagissent pas;
- 1. Ces deux taxa ne peuvent probablement pas interagir;
- 2. Pas d'avis sur l'interaction de ces deux taxa;
- 3. Ces deux taxa peuvent probablement interagir;
- 4. Ces deux taxa interagissent.

Les experts ont effectué cette évaluation sans avoir connaissance des probabilités d'interactions inférées par les modèles. Du fait du petit nombre d'experts, les réponses ont été agrégées en trois catégories : interaction possible, interaction impossible et ne se prononce pas. La concordance de l'avis des experts a été mesurée à l'aide du  $W$  de Kendall (Legendre 2005). Enfin, les experts ont aussi été interrogés de manière plus qualitative sur la vraisemblance des réseaux reconstruits par la méthode *EMtree*.

---

<sup>3</sup>Pour le calcul des différentes métriques, se référer à Poisot *et al.* (2015) et Poisot *et al.* (2019).

## **Matériel informatique logiciels utilisés**

Les calculs pour tous les modèles ont été réalisés par le supercalculateur de l'IFREMER DATARMOR. Le processeur de chaque noeud de calcul du cluster HPC est un CPU Intel E5-2680 v de 14 coeurs cadencés à 2,40 GHz. Chaque modèle n'utilisait pas plus d'un seul cœur.

Les modèles ont été mis en place à l'aide du langage de programmation R 3.6.1 (R Core Team 2019). Les résultats ont été analysés conjointement grâce aux langages R (R Core Team 2019) et Julia 1.2.0 (Bezanson *et al.* 2017).

## **Résultats**

### **Comparaison des performances des modèles**

#### **Pouvoir explicatif**

Les capacités explicatives des différents modèles permettent de les séparer en deux groupes distincts (Figure 2). Les modèles *PLN*, *GLLVM* et *HMSC\_reg* expliquent en moyenne autour de 10% de la variabilité des espèces contre 25% pour les modèles *HMSC\_samp* et *HMSC\_hier*. De plus, *HMSC\_samp* et *HMSC\_hier* expliquent jusqu'à 60 à 80% de la variation des espèces les mieux expliquées alors que les modèles *PLN*, *GLLVM* et *HMSC\_reg* ne dépassent pas les 40%. Enfin, ces deux groupes de modèles diffèrent aussi dans l'identité des espèces les mieux expliquées, alors qu'entre les modèles d'un même groupe, les mêmes espèces ressortent de façon extrêmement consistante. Néanmoins, les performances de l'ensemble de ces modèles en termes de pouvoir explicatif sont à mettre en perspective du grand nombre d'espèces ayant un  $SR^2$  inférieur à 0,10.

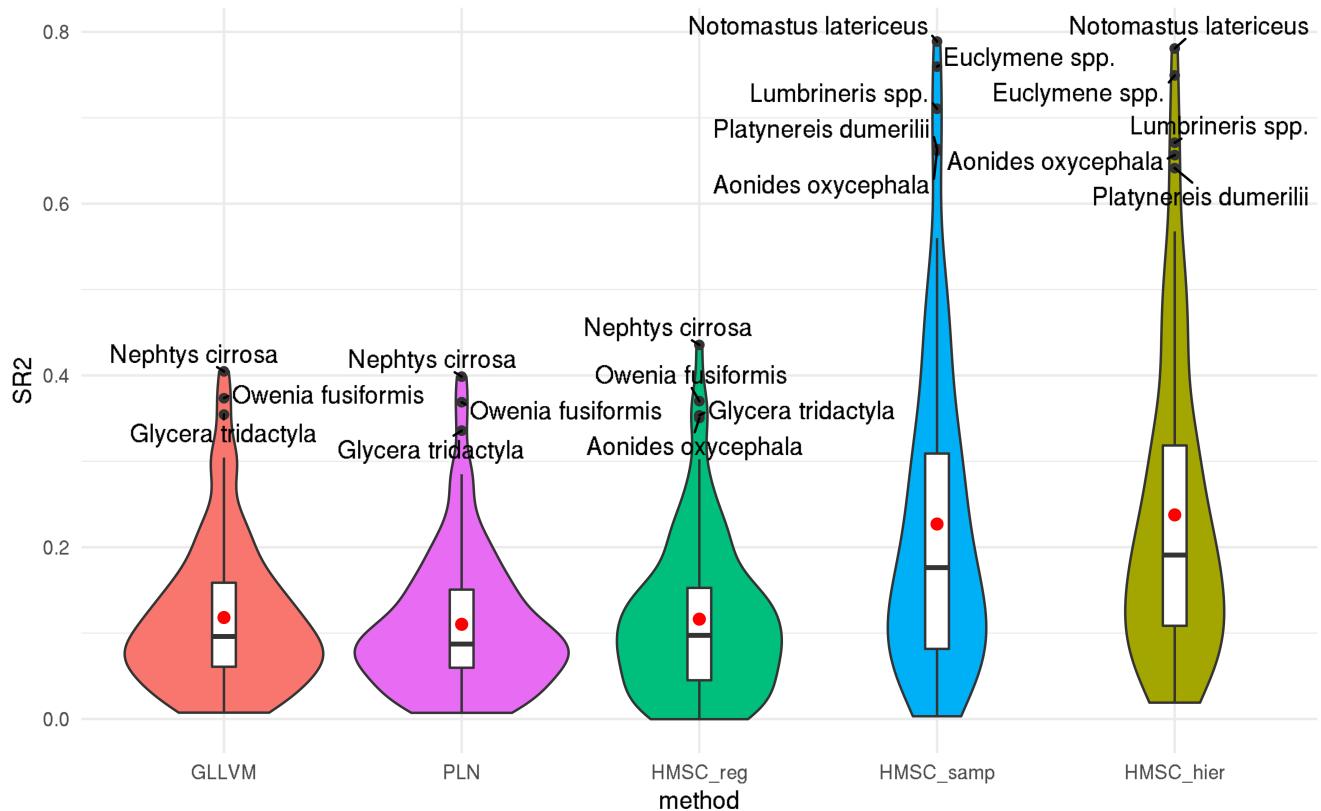


Figure 2 : Distribution du SR<sup>2</sup> de l'abondance des 92 taxa de polychètes pour chaque modèle. Les points rouges représentent la valeur moyenne du SR<sup>2</sup> pour chaque modèle. Les points noirs représentent les espèces ayant une valeur extrême de SR<sup>2</sup>.

### Prédiction de l'occurrence et de l'abondance

La validation croisée sur trois sites tests (exclus de l'apprentissage) permet la comparaison du pouvoir prédictif des modèles en termes de richesse spécifique et d'abondance, et ce sur les 8 années de suivis de ces trois sites.

**Abondance** Aucun modèle ne parvient à prédire de manière satisfaisante l'abondance des espèces à chaque site. Le modèle le moins mauvais est le modèle *HMSC\_hier* dont la plus faible valeur de RMSE pour une espèce est de 24 individus et la plus forte valeur de RMSE est de l'ordre de 10<sup>4</sup>. En moyenne ce modèle présente un RMSE de l'ordre de 3,63 × 10<sup>3</sup>. Bien que cette erreur moyenne soit importante, elle est plus faible que celle des autres modèles (entre 6,53 × 10<sup>3</sup> et 1,09 × 10<sup>3</sup>). Le modèle GLLVM apparaît comme le moins fiable pour estimer l'abondance, avec une erreur minimale de prédiction de l'ordre de 10<sup>3</sup>1.

**Occurrence** Si aucun modèle n'arrive pleinement à reproduire les richesses observées aux sites de validation et leurs dynamiques temporelles ((Figure 3, Figure 4 et Figure 5), les prédictions de l'occurrence des espèces semblent bien plus réalistes que celles concernant leur abondance.

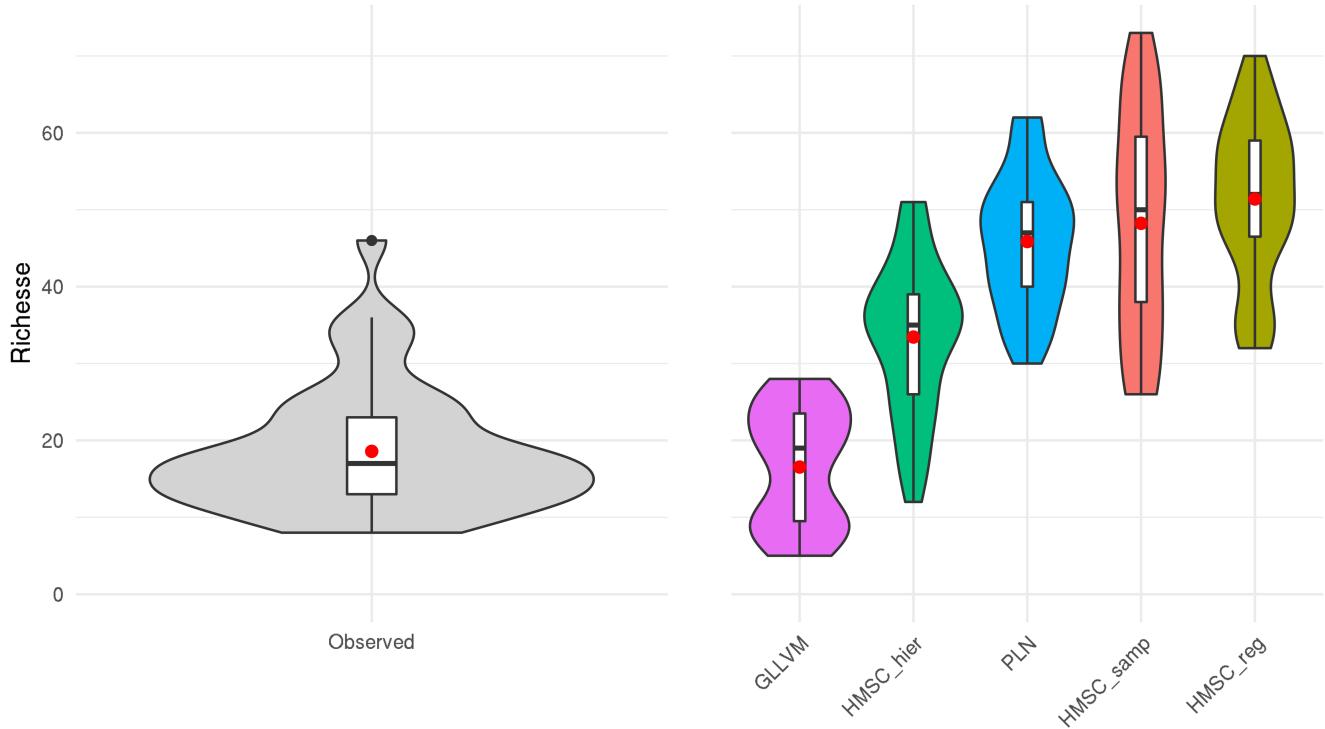


Figure 3 : Variabilité spatiale et temporelle de la richesse spécifique observée et prédictée par les différentes méthodes de *JSDM* pour les 3 sites de validation. Les points rouges correspondent à la richesse moyenne des sites de validation.

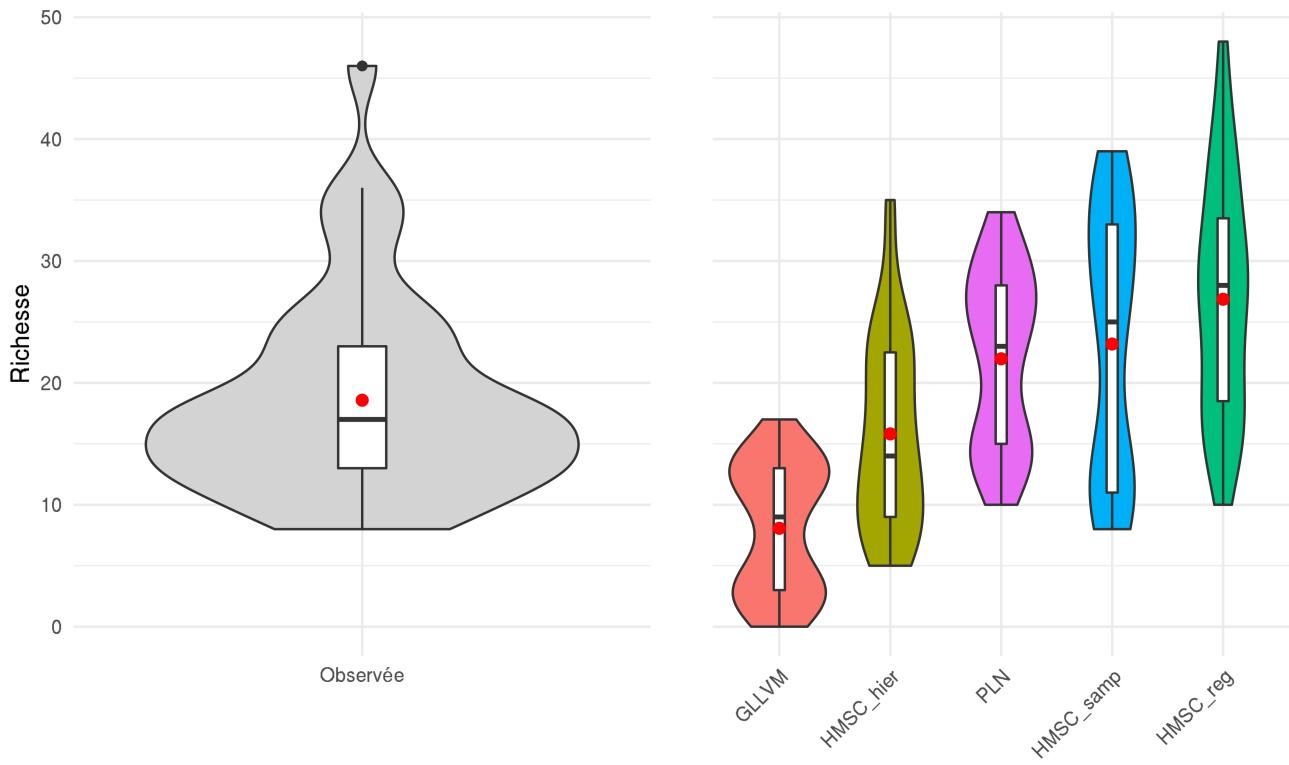


Figure 4 : Variabilité spatiale et temporelle de la richesse spécifique observée, corrigée par les seuils d’abondance et prédite par les différentes méthodes de *JSDM* pour les 3 sites de validation. Les points rouges correspondent à la richesse moyenne des sites de validation.

La richesse moyenne observée aux sites de validation est d’environ 20 espèces. Si l’on considère qu’une espèce est présente à partir du moment où son abundance prédictive est supérieure à 0, *GLLVM* est le modèle dont la richesse prédictive est la plus proche de celle observée, en ayant toutefois tendance à sous-estimer la richesse moyenne (Figure 3, Figure 4 et Figure 5). Cette sous-estimation est encore plus prégnante lorsqu’on utilise un seuil d’abondance adapté pour estimer l’occurrence des espèces (Figure 5). Au contraire, *PLN*, *GLLVM* et *HMSC\_samp* surestiment la richesse des communautés avec une richesse moyenne deux fois supérieure à celle observée si l’on convertit directement l’abondance prédictive des espèces en présence/absence, mais sont dans une gamme de richesse plus cohérente lorsque l’approche par seuil est utilisée. Dans les deux approches, *HMSC\_hier* apparaît comme la meilleure méthode pour estimer l’occurrence des espèces et prédire la richesse des assemblages de polychètes. Sans définir de seuil spécifique de présence effective, le *RMSE* indique que l’erreur sur la prédiction de la richesse spécifique est finalement relativement faible pour l’ensemble des modèles (Tableau 4). Le *RMSE* médian est de 7 pour le modèle *GLLVM* et de 16 pour le modèle *HMSC\_hier*. Mais ce *RMSE médian* est près de deux fois plus élevées pour les modèles *PLN*, *HMSC\_samp* et *HMSC\_reg*.

Tableau 4 : Statistiques descriptives du *RMSE* de la richesse spécifique par modèle. Les modèles sont classés par ordre croissant de *RMSE* maximal. Q1 et Q3 représentent respectivement le premier et troisième quartile du *RMSE* de chaque modèle.

Method	Min	Q1	Median	Moy	Q3	Max
<i>GLLVM</i>	0	3,0	7	7,63	10,0	29
<i>HMSC_hier</i>	1	11,5	16	15,37	20,0	33
<i>PLN</i>	4	21,0	27	27,28	33,0	51
<i>HMSC_reg</i>	6	25,0	33	32,84	40,5	56
<i>HMSC_samp</i>	4	18,5	30	29,65	38,5	62

Les modèles prédisent correctement la variabilité de la richesse inter-site en trouvant tous l'ordre de grandeur de la richesse spécifique pour chaque site (Figure 5). Les modèles prédisent mieux la richesse spécifique dans les herbiers que dans les plages de sable. *HMSC\_hier* est le modèle qui prédit le mieux la variabilité de la richesse spécifique. Ce modèle arrive à également prédire les fluctuations temporelles, notamment la diminution de la richesse spécifique observée pour les herbiers en 2009.

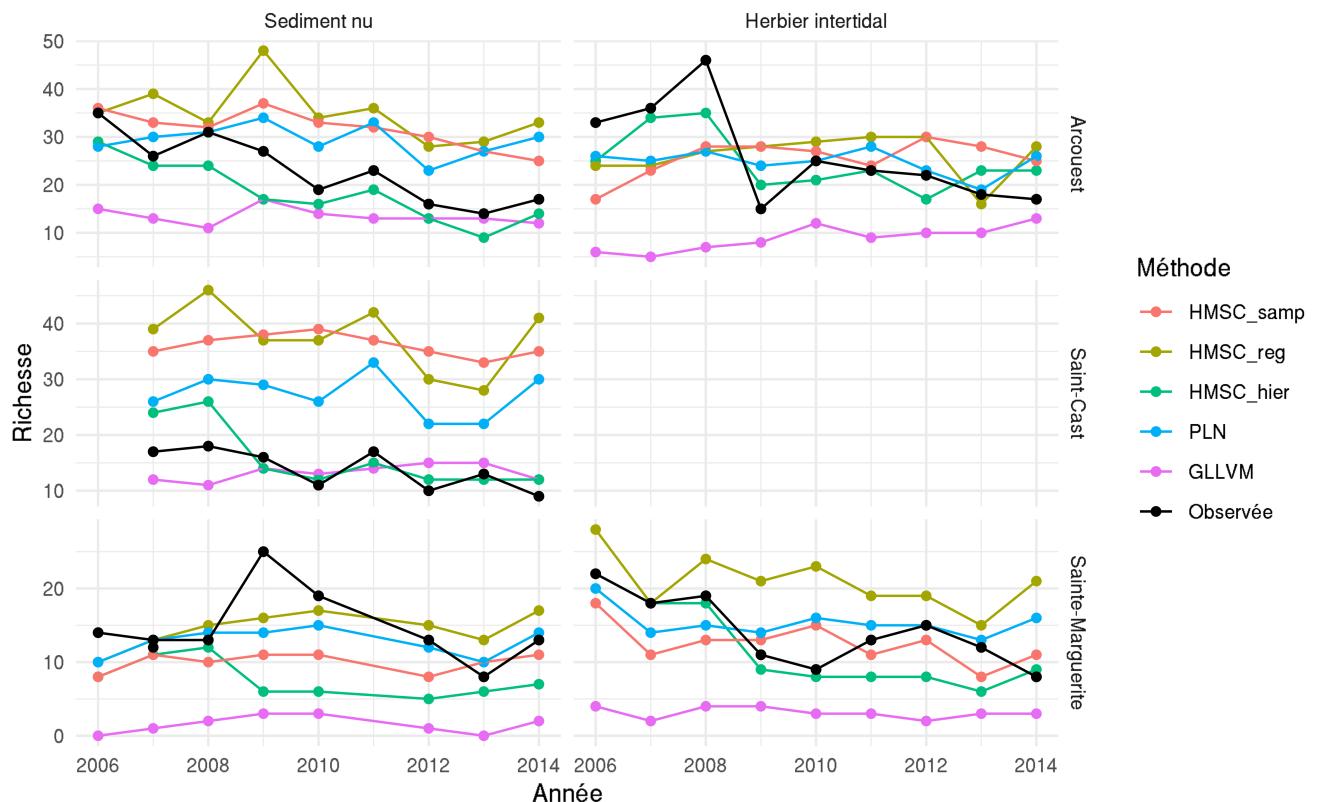


Figure 5 : Variabilité spatio-temporelle de la prédition de la diversité spécifique des communautés de polychètes selon les différentes méthodes de *JSDM*. L'occurrence prédictive est basée sur des seuils de probabilité spécifiques à chaque taxon.

## **Effets de l'environnement**

Les effets des variables environnementales sur l'abondance prédictive des différents taxa sont globalement similaires à travers les différents modèles (Figure 6). Dans la Figure 6 les variables environnementales sont ordonnées par proportion décroissante de variances expliquée. Ainsi, des variables mesurées à l'échelle du site comme l'exposition (Fetch) ou la vitesse moyenne du courant expliquent une plus grande part de variance dans l'abondance des espèces que les variables sédimentologiques comme l'indice de Trask ou la concentration en vase mesurée à une échelle plus locale (Figure 10 à 12 en annexe). L'analyse de cette figure par les experts benthologues a aussi mis en lumière que ces espèces étaient toutes majoritairement inféodées à des environnements de sables fins, ce qui tend à corroborer les effets positifs du Trask observés dans la majorité des modèles, sauf *HMSC\_hier*. Dans l'ensemble, les facteurs aléatoires sont ceux qui expliquent le plus de la variation des espèces, notamment pour celles les mieux expliquées (Figure 1 à 3 en annexe). La façon d'implémenter ces effets aléatoires apparaît donc critique, d'autant que dans *HMSC* ils influencent grandement les coefficients estimés (Figure 5).

Pour tous les modèles, l'effet des variables environnementales est assez net pour les espèces les mieux expliquées (15 premières en partant du haut) et devient moins marqué quand il s'agit d'expliquer l'abondance des espèces les moins bien expliquées (espèces qui se trouvent dans le tiers inférieur de la Figure 6). De plus, tous les modèles montrent un effet positif pour la variable Trask (plus le sédiment est bien trié plus les espèces sont abondantes), à l'exception du modèle *HMSC\_hier*. Avec l'inclusion des effets aléatoire sites, habitats et années, le modèle *HMSC\_hier* montre à l'inverse un effet faiblement négatif, voire négatif, sur l'abondance des espèces. Le modèle *HMSC\_hier* est ainsi celui qui se différencie le plus en termes de coefficients estimés, ce qui semble en lien avec la part de variance importante expliquée par l'ajout des facteurs aléatoires (Figure 10 à 12).

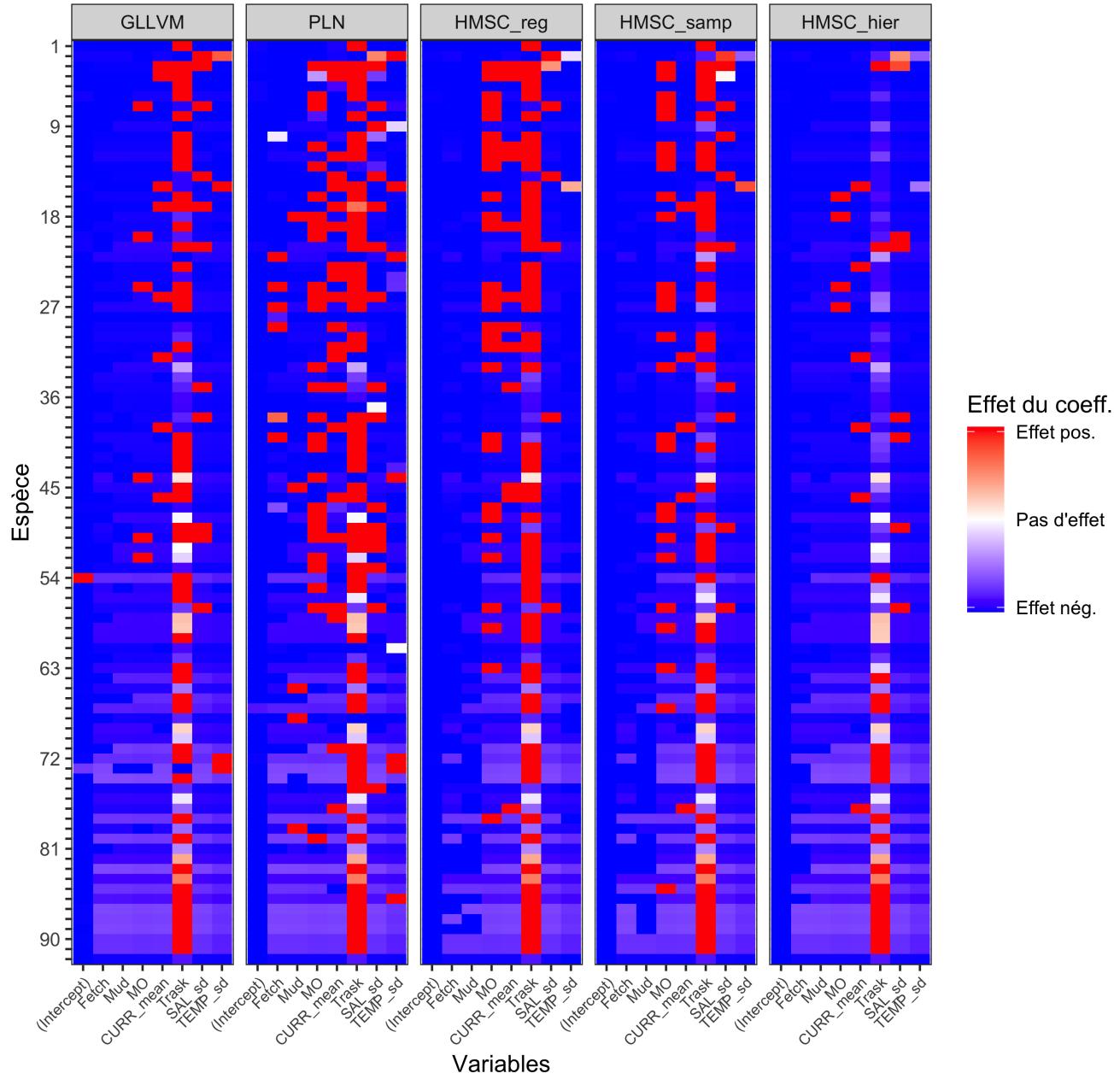


Figure 6 : Effets des variables environnementales sur l'abondance des différentes espèces de poly-chètes. Les variables et les abondances sont centrées et réduites. Les espèces sont ordonnées par  $SR^2$  moyen décroissant. Les variables environnementales sont ordonnées par proportion de variances expliquées décroissante (voir Tableau 2 pour leur description).

## Réseaux reconstruits

### Analyse des graphes

Tous les réseaux probabilistes reconstruits présentent le même nombre de liens 182 et globalement, il y a peu de différence notable dans la structure des réseaux inférés par les différents modèles (Tableau 5). *L'écart-type du nombre de liens est légèrement plus important pour les effets aléatoires année et*

*habitat* du modèle HMSC\_hier\*, mais reste dans des ordres de grandeur similaires à ceux des autres modèles. La connectance , qui peut en théorie varier entre 0 et 1, est quasiment identique pour tous les modèles et apparaît extrêmement faible. L’imbrication des réseaux est aussi quasi nulle et similaire sur l’ensemble des modèles.

Tableau 5 : Métrique des réseaux d’interactions reconstruits.  $\sigma_l$  représente l’écart-type du nombre de liens.  $C$ , Connectance.  $\eta$ , imbrication des réseaux trophiques.

Méthode	$\sigma_l$	$C$	$\eta$
<i>HMSC_samp</i>	12, 70	0, 022	0, 03
<i>HMSC_hier_annee</i>	13, 10	0, 021	0, 03
<i>HMSC_hier_site</i>	12, 46	0, 021	0, 03
<i>HMSC_hier_habitat</i>	13, 11	0, 021	0, 03
<i>GLLVM</i>	12, 51	0, 021	0, 04
<i>PLN</i>	12, 60	0, 022	0, 04

Le réseau moyen reconstruit sur la base des corrélations résiduelles entre taxons met en évidence un nombre restreint d’interactions fortement probables (Figure 7) en corroborant les faibles connectances estimées sur les réseaux individuels (Tableau 5). Ainsi, beaucoup d’espèces ne semblent pas avoir de liens dans le réseau, avec une majorité des interactions inférées ayant une probabilité inférieure ou égale à 0, 2.

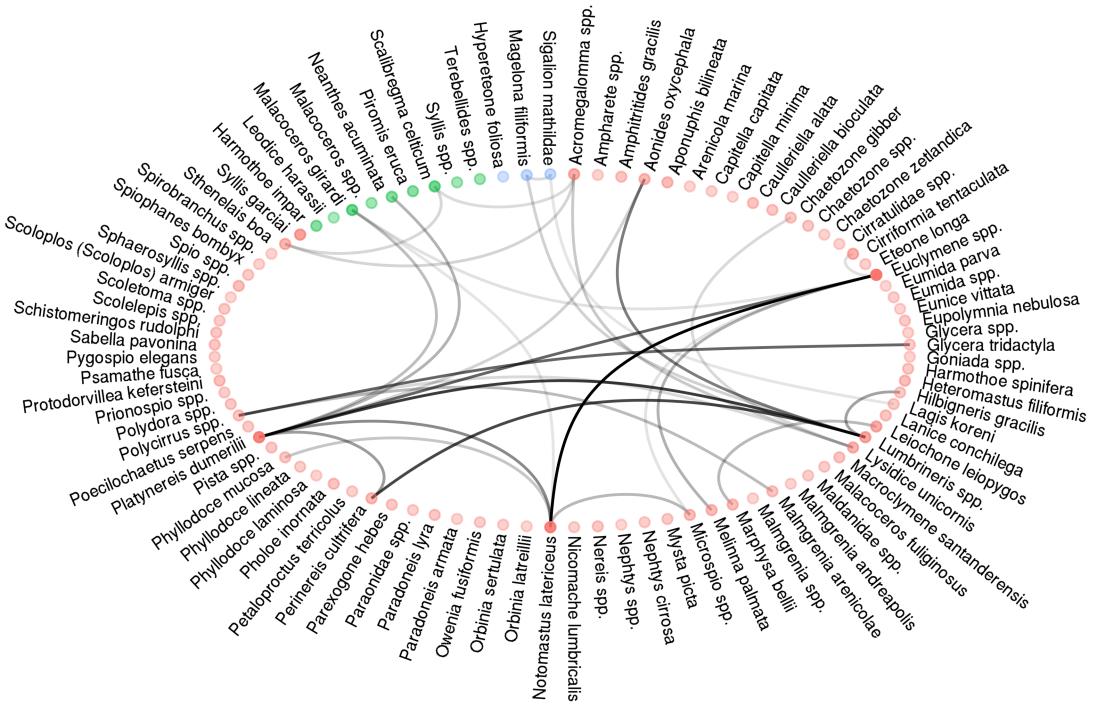


Figure 7 : Réseau inféré moyen qui présente des probabilités d'interaction moyennées sur l'ensemble des réseaux reconstruits décrits dans le Tableau 2 . Les points rouges représentent les taxa de polychètes retrouvés dans les deux habitats, les verts ceux retrouvés uniquement dans les herbiers et les bleus ceux présents seulement dans les sédiments meubles. Seules les arrêtes ayant une probabilité supérieure à 0,2 sont affichées. L'opacité des arrêtes est proportionnelle à leur importance dans le réseau (mesurée par la centralité de Katz)

Plusieurs réseaux d'interactions peuvent être dérivés de *HMSC*, une association résiduelle pouvant être extraite pour chacun des facteurs aléatoires inclus dans le modèle. La Figure 6 présente par exemple le réseau reconstruit sur la base des corrélations résiduelles associées au facteur aléatoire temporel (années). Il est intéressant de noter que seules des corrélations positives entre espèces sont présentes dans ce réseau alors que les réseaux liés aux facteurs sites ou habitats, ou que les réseaux issus des autres modèles incluent aussi des corrélations négatives (Figure 16 et 17 en annexe).

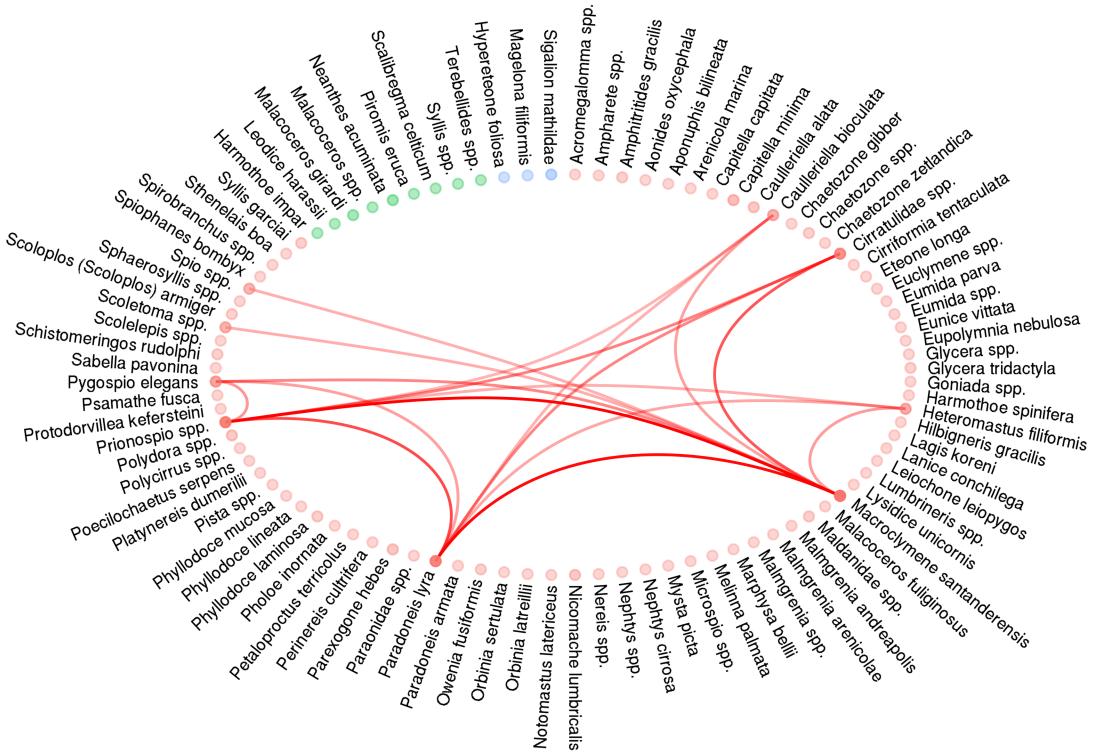


Figure 8 : Réseau reconstruit avec EMTree sur la base des corrélations résiduelles associées à l'effet aléatoire temporelle du modèle *HMSC\_hier*. Les points rouges représentent les taxa de polychètes retrouvés dans les deux habitats, les verts retrouvés uniquement dans les herbiers et les bleus dans les sédiments meubles. Seules les arrêtes ayant une probabilité supérieure à 0, 2 sont affichées. L'opacité des arrêtes est proportionnelle à leur probabilité. L'opacité des points est proportionnelle à leur importance dans le réseau.

### Avis des experts du taxon des polychètes

L'analyse de concordance de Kendall montre qu'il n'y a pas de consensus significatif entre les différents experts quant à la probabilité d'interaction entre les espèces qui leur ont été soumises ( $W = 0, 26$ ,  $p > 0, 2$ ). Résumer ces notes attribuées par les experts en deux catégories seulement (interaction possible ou impossible), ne modifie pas cet absence de consensus entre les experts ( $W = 0, 2$ ,  $p > 0, 44$ ). Cette absence de concordance entre experts rend difficile la validation des interactions inférées par les modèles, mais il semble que les probabilités d'interactions prédites ne correspondent pas à la vision globale des experts (Figure 9). En effet, les interactions prédites sur l'ensemble des modèles comme étant les plus probables sont plus souvent qualifiées d'impossible par les experts que de potentiellement possible. Pour certaines interactions, il faut cependant noter que l'avis des experts était unanime, et il a ainsi été possible d'interpréter qualitativement le sens de certaines interactions prédictes via des discussions directes avec certains experts.

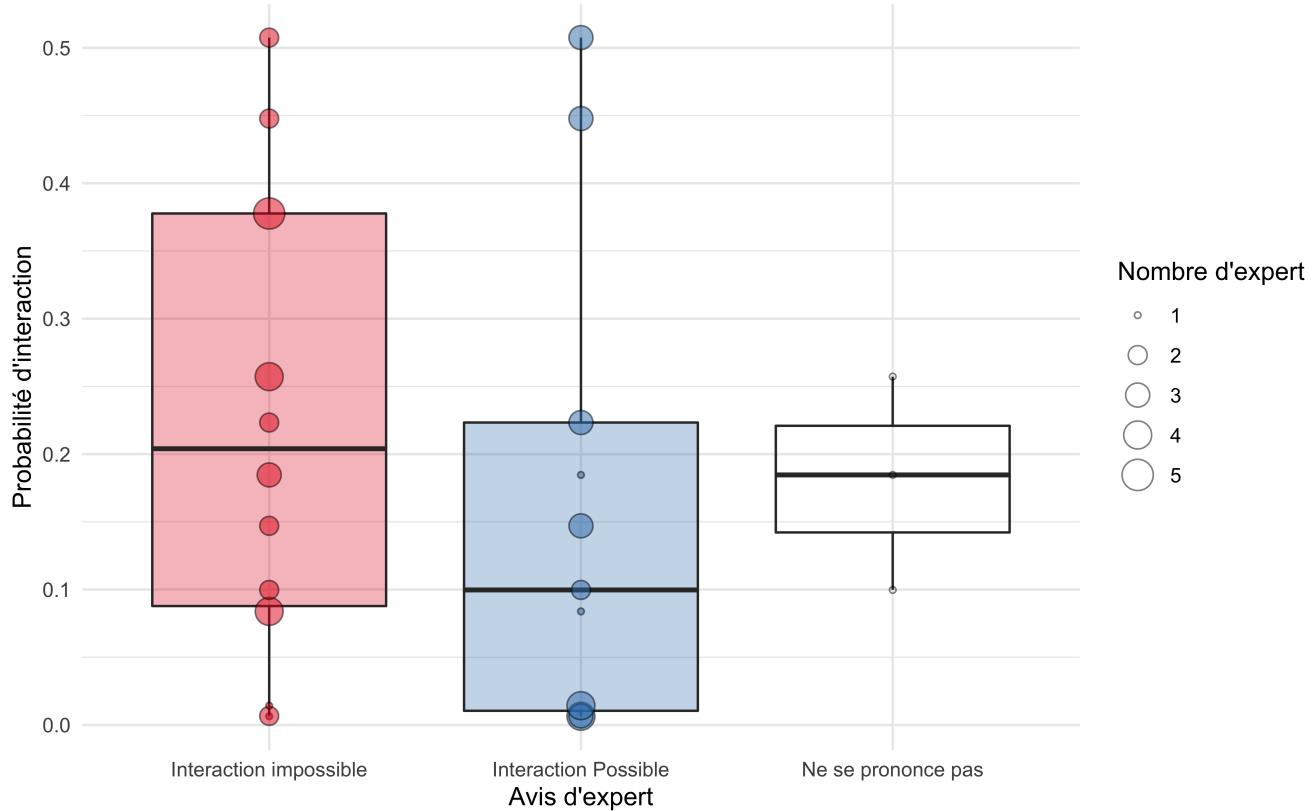


Figure 9 : Avis des experts quant à la possibilité ou d'une interaction en fonction des probabilités d'interaction moyenne prédictes par les modèles. La taille des points est proportionnelle au nombre d'experts ayant donné la même réponse.

## Coût de calcul

Chaque modèle présente des caractéristiques propres en termes de coup de calcul (Tableau 6). Les modèles implémentés sous *HMSC* présentent les temps de calcul les plus longs, mais leur coût en mémoire vive reste cependant faible. Le temps de calcul augmente considérablement avec le nombre d'effets aléatoires passant de plus de 25h pour le modèle le plus simple (sans effet aléatoire) à près de 458h pour le modèle le plus complexe (trois effets aléatoires indépendants). Comme attendu, *GLVM* se révèle être une implémentation particulièrement rapide, étant même plus rapide à ajuster que le modèle sans effet aléatoire *HMSC\_reg*. Toutefois, sa vitesse est largement contrebalancée par le coût important en mémoire vive. Ce modèle a nécessité plus de 68 Go de RAM pour pouvoir être ajusté. *PLN* quant à lui est le modèle le plus économique à ajuster, il n'a besoin que de trois minutes et moins de 500 Mo de RAM pour être ajusté.

Tableau 6 : Coût de calculs des différents modèles.

Modèle	Temps de calcul (h : mn)	RAM (Go)
<i>HMSC_reg</i>	25 : 27	0,49
<i>HMSC_samp</i>	170 : 56	0,69
<i>HMSC_hier</i>	457 : 50	0,73
<i>PLN</i>	00 : 03	0,37
<i>GLLVM</i>	13 : 43	68,1

## Discussion

### Comprendre et prédire la diversité benthique

Les différents modèles présentés ici n'expliquent que partiellement la variabilité des assemblages de polychètes. Ainsi, sur les 92 espèces du jeu données d'apprentissage, seules quelques espèces comme *Notomastus latericeus* (n°1), *Euclymene spp.*, *Lumbrineris spp.* (n°3), *Nephtys cirrosa* (n°2) ou bien *Owenia fusiformis* (n°7) sont convenablement expliqués par les modèles.<sup>4</sup>

Les modèles tels que *HMSC\_hier* et *HMSC\_samp* semblent tirer parti de leur structure hiérarchique pour expliquer convenablement un plus grand nombre d'espèces par rapport aux modèles ne disposant pas d'une telle structure. En effet, ce sont principalement les effets aléatoires qui expliquent le plus de la variation des espèces. Ainsi, le manque de pouvoir explicatif des variables environnementales semble montrer que la niche écologique des espèces de cette communauté est mal définie dans nos modèles. Les variables environnementales sélectionnées ne sont probablement pas mesurées à la bonne échelle, ou leur effet ne peut être distingué correctement des effets spatiaux. Ce manque de pouvoir explicatif des variables environnementales limite la transférabilité de ces modèles à d'autres lieux et environnements.

Nos modèles prédisent relativement bien l'occurrence des espèces et donc la richesse des assemblages et leurs variations dans l'espace et le temps. Ces résultats sont consistants avec les précédentes comparaisons qui ont été faites de ces modèles (Norberg *et al.* 2019; Momal *et al.* 2020). Notamment, la richesse des assemblages est apparue particulièrement bien prédite lorsque des seuils espèce-spécifique ont été utilisés pour convertir l'abondance en présence/absence (Ashcroft *et al.* 2017). L'utilisation de cette approche par seuil, largement inspirée des approches de cartographie à partir de *SDM* (Lawson *et al.* 2013), est inédite à notre connaissance pour l'analyse des prédictions issues de *JSDM*. Une comparaison des performances de cette approche (modélisation de l'abondance puis seuils pour convertir en occurrence) avec de modèles entraînés directement sur des données d'occurrence (ce qui était fait jusqu'à présent, e.g; Norberg *et al.* (2019)) pourrait permettre d'évaluer plus

<sup>4</sup>Les numéros correspondent aux entrées dans le Tableau 8 .

clairement la validité de l'approche. Cependant, il nous semble potentiellement plus prometteur de prendre en compte l'information contenue dans l'abondance des espèces pour entraîner les modèles puis d'utiliser des seuils appropriés pour modéliser la richesse des communautés, que d'utiliser une transformation brute des abondances en présence/absence, ou encore que de négliger complètement l'information contenue dans l'abondance des espèces. Cependant, si l'une des promesses des *JSDM* est bien de pouvoir prédire également l'abondance des espèces, la performance des modèles testés ici est apparue décevante. Le *RMSE* maximal pour le meilleur modèle, *HMSC\_reg*, est ainsi de l'ordre de  $10^4$ , ce qui justifie d'autant plus l'utilisation des seuils espèce-spécifique pour prédire la richesse des communautés à partir de *JSDM* entraîné sur des données d'abondances.

Sur ce critère d'abondance, *GLLVM* est le moins performant des modèles testés. Cela peut s'expliquer par plusieurs facteurs. Notamment, la distribution négative binomiale a été utilisée pour le modèle *GLLVM* contrairement aux autres modèles qui utilisent la distribution de Poisson lognormal. Dans le cadre plus classique de *SDM*, cette distribution négative binomiale montre en général de plus faibles capacités à modéliser l'abondance des espèces que d'autres distributions (Potts & Elith 2006). Ce manque de performance prédictif a également été montré face à la distribution de Poisson lognormal (Trinh *et al.* 2013). Il semblerait que la structure de la variance de la distribution négative binomiale permette trop de surdispersion par rapport à d'autres distributions, et donc donne des prédictions moins bonnes. De plus, l'utilisation d'effets aléatoires au niveau de l'observation comme le permet la distribution de Poisson lognormale permet de mieux prendre en compte la surdispersion des données d'abondances en permettant plus de souplesse au modèle (Harrison 2014). Enfin, il semblerait que l'utilisation dans *GLLVM* d'approximation variationnelle pour améliorer les vitesses de calcul, puisse affecter la précision des résultats par rapport aux autres approches (Warton *et al.* 2015). Dans l'ensemble, la faible qualité de la prédiction de l'abondance peut en partie être expliquée par le faible pouvoir explicatif des variables environnementales utilisées pour entraîner nos modèles. En ayant une meilleure connaissance ou une meilleure prise en compte de la niche écologique de ces espèces de polychètes, il serait possible d'améliorer significativement la qualité prédictive de ces modèles (Leach *et al.* 2016 ; Barbaro *et al.* 2019). Dans le cadre de cette étude, le nombre de variables environnementales prises en compte dans les modèles a avant tout été limité pour des questions de coûts de calculs et de nombre de paramètres à estimer. En prenant en compte toutes les données existantes du suivi REBENT, il serait possible d'utiliser plus de variables environnementales ainsi que des polynômes de degrés plus élevés pour modéliser des relations non linéaires des espèces avec leur environnement. Si cela était initialement envisagé, les temps de calcul des modèles se sont révélés prohibitifs pour ce genre d'approches, et le choix a été fait de se limiter aux variables potentiellement les plus intéressantes sur la base de travaux précédents (Boyé 2019).

## Rôle et inférence des interactions biotiques

Prendre en compte les interactions dans nos modèles améliore globalement leurs capacités de prédictions si nous les comparons aux résultats d'un modèle sans effet aléatoire tel que *HMSC\_reg*, qui se retrouve toujours parmi les moins bons modèles. Ce résultat est en accord avec la littérature, puisque la prise en compte des interactions biotiques améliore les capacités prédictives des modèles linéaires (Wiss *et al.* 2012; Gavish *et al.* 2017; Staniczenko *et al.* 2017). Bien que la prise en compte des interactions par les différents modèles se fait de la même manière à l'aide de variables latentes, l'implémentation de ces variables latentes change entre les modèles. Cette différence dans la façon d'inclure ces interactions se répercute sur le rôle estimé des facteurs environnementaux dans les modèles. L'indice de Trask a une influence positive pour beaucoup d'espèces selon la plupart des modèles, mais explique une faible variance. Cependant, en prenant en compte un plus grand nombre de facteurs aléatoires, le modèle *HMSC\_hier* a démontré que son effet était largement inversé (et majoritairement négatif). Cela semble suggérer que l'indice de Trask est une variable fortement liée au site.

Malgré le fait que l'avis de nos experts ne concorde pas quant à la probabilité des liens, certains liens du réseau moyen ont été identifiés par les experts comme pouvant correspondre à des interactions proies-prédateurs. Par exemple, celles entre *Perinereis cultrifera* (n°12) et *Lumbrineris spp.* (n°3), *Magelona filiformis* (n°9) et *Sigalion mathildae* (n°22) ou bien encore *Scalibregma celticum* (n°27) et *Sthenelais boa* (n°40). Par ailleurs, la probabilité forte d'interaction entre *Platynereis dumerilii* (n°10) et *Euclymene spp.* (n°5) et encore *Notomastus latericeus* (n°1) pourrait selon les experts traduire de la compétition trophique et spatiale. En effet, *Platynereis dumerilii* (n°10) est un broutteur de microphytobenthos et les deux autres espèces sont des dépositoires (Jumars *et al.* 2015). Ces trois taxa vont donc acquérir leur nourriture au même endroit, sur ou près de la surface du sédiment. De plus, ces trois taxa sont également ceux avec la plus grande centralité dans le réseau moyen. Cette importance dans le réseau est peut-être liée à la forte dominance des espèces dépositoires ou broutteuses dans les habitats suivis (Boyé *et al.* 2019).

La structure de ces réseaux reconstruits assez particulière : la valeur de la connectance est dix fois plus petite que celle observée pour d'autres réseaux benthiques (Martinez 1992; Dunne *et al.* 2004). Les réseaux reconstruits présentent également une imbrication très faible par rapport à ce que montre la littérature des réseaux trophiques benthique (Nordström *et al.* 2015). Ces résultats, ainsi que le faible nombre d'interactions ayant une probabilité supérieure à 0,2 est probablement lié à la nature des communautés étudiées, et en particulier au focus de ce travail sur les polychètes. En effet, des analyses de traits biologiques de ces assemblages de polychètes ont suggéré précédemment une prédominance de filtres biotiques et de processus neutres, et un faible rôle des interactions biotiques, dans la structure de ces assemblages de polychètes (Boyé *et al.* 2019). Les résultats de cette étude, et notamment les faibles connectances observées sur l'ensemble des modèles, semblent confirmer ces

hypothèses. Cela peut aussi s'expliquer par le fait que les polychètes ne forment pas un réseau à part entière et que de nombreuses interactions avec d'autres phylums ne sont ainsi pas représentées. En effet, il nous manque beaucoup d'acteurs pour obtenir un réseau d'interaction d'une communauté benthique complète, certains poissons et d'autres invertébrés par exemple, sont des prédateurs de polychètes (Jankowska *et al.* 2019), et ne sont pas inclus dans les réseaux que nous avons reconstruits, alors que certains bivalves peuvent avoir un rôle facilitateur important (Gagnon *et al.* 2020), qui ne sont pas non plus pris en compte.

Le nombre de liens identiques à travers tous les réseaux reconstruits peut sembler étrange au premier regard. C'est une conséquence de la manière dont sont générés ces graphes à l'aide de l'algorithme *EMtree* et de la façon dont sont calculés les liens dans le cadre des réseaux écologiques probabilistes (Poisot *et al.* 2015). Pour reconstruire ces réseaux *EMtree* parcourt l'ensemble des arbres couvrants possibles. Dans notre cas, notre communauté pour le jeu de donnée d'apprentissage compte  $p = 92$  espèces, et il y a  $p^{p-2}$  arbres couvrant possible différent, soit un total de  $5,51 \times 10^{176}$ . Pour rappel, dans un cadre de réseaux probabilistiques, le nombre de liens est calculé comme étant la somme des probabilités contenue dans la matrice . *EMtree* parcourant toujours le même espace d'arbre couvrant possible. C'est pourquoi la somme des probabilités de la matrice d'interaction (nombre de liens) est toujours identique, quelle que soit la méthode, mais qu'il est possible d'avoir de la variabilité dans le nombre de liens et la connectance, car chaque lien n'a pas la même probabilité en fonction du réseau.

Le manque de concordance entre nos différents experts quant à la probabilité des interactions reconstruites s'explique de différente façon. Tout d'abord, les polychètes sont des animaux qui restent assez mal connus actuellement (Troudet *et al.* 2017). Il reste de nombreuses interrogations quant à leurs modes de vie et les possibles interactions qu'il existe au sein de ce clade (Jumars *et al.* 2015; Boyé *et al.* 2019). De plus, certains patrons de co-occurrence ont semblé étranges à nos experts. Le très fort nombre de corrélations positives obtenues par l'effet aléatoire temporelle du modèle *HMS\_hier* n'était pas attendu (Figure 8). En effet, l'hypothèse initiale était que les associations résiduelles dans le temps seraient plus à même de représenter des interactions biotiques que des effets résiduels liés à la variabilité spatiale ou à l'habitat associés à des variables abiotiques non considérées dans les modèles. Dans cette hypothèse, des interactions de type proie préation ou compétition se serait alors traduite par des associations négatives dans le temps. Le fait que seules des interactions positives soient retrouvées pourrait s'expliquer par différents facteurs non exclusifs tels que (1) une prépondérance d'effets facilitateur dans ces communautés benthiques intertidales (Bertness & Leonard 1997; Gribben *et al.* 2019) (2) l'échelle spatio-temporelle du suivi REBENT : à larges échelles, des prédateurs auront tendance à être là où leurs proies se trouvent, résultant en des associations positives, là où des dynamiques temporelles à plus fines échelles montreraient des associations négatives (e.g. modèle de Lokta-Volterra). Enfin, Thurman *et al.* (2019) en utilisant des données empiriques ont montré que plus une espèce interagissait avec d'autres, plus la corrélation résiduelle entre ces espèces dans un *JSDM* était faible et tendait à être positif. Etant donnée la grande diversité de notre jeu de

données, cela pourrait expliquer à la fois les faibles probabilités d'interactions et le grand nombre d'interactions positives observées.

Enfin, malgré les limitations soulevées précédemment, il est important de noter que notre méthodologie de reconstructions des réseaux d'interactions nous permet d'éviter plusieurs des écueils résumés par (Blanchet *et al.* 2020) quant à l'utilisation faite classiquement des *JSDM* pour inférer des interactions entre espèces. Tout d'abord, notre travail utilise des données d'abondance qui sont plus riche en informations que les données de présence/absence. De plus, nous utilisons une méthode basée sur les graphes pour reconstruire les interactions, ce qui évite que les interactions indirectes soient confondues avec des interactions directes(Chiquet *et al.* 2019 ; Momal *et al.* 2020), renforçant ainsi la vraisemblance écologique des interactions inférées avec des probabilités fortes.

## Forces et faiblesses des différentes approches

Dans leur ensemble, les modèles testés dans ce travail sont assez prometteurs. Leur capacité à prédire convenablement l'occurrence des espèces et leurs dynamiques spatiales et temporelles pourrait servir à évaluer l'impact de certains changements environnementaux sur la composition et la richesse des communautés. Cependant, l'importance des effets aléatoires dans la capacité prédictive de ces modèles pourrait limiter leur transférabilité (Yates *et al.* 2018) à d'autres zones géographiques ou à d'autres habitats. En revanche, la capacité de modèles comme *HMSC*, grâce à leur structure hiérarchique, à découper les corrélations résiduelles entre espèces selon différents facteurs aléatoires offre la possibilité d'obtenir une vision plus fine sur les possibles interactions et associations interspécifiques dans l'espace et le temps.

Bien que certains modèles semblent sensiblement plus performants que d'autres, aucun n'est parfait : "Tous les modèles sont faux, mais certains sont utiles." (Box 1979). Ainsi, chaque modèle représente un trade-off entre une plus grande précision des paramètres estimés et un coût de calcul plus élevé. *HMSC\_hier* est le modèle le plus précis et le plus performant en termes prédictifs dans notre étude, mais cela au prix d'un temps de calcul extrêmement élevé (i.e. des centaines d'heures de calculs sur le supercalculateur DATARMOR). Ainsi, si la question de recherche se trouve être la prédiction de l'occurrence, il est préférable de se tourner vers *GLVM*, notamment si les délais de la recherche sont réduits et si la mémoire vive nécessaire est disponible, car sa capacité à prédire la richesse des assemblages reste satisfaisante par rapport à *HMSC*. Si l'on ne dispose pas assez de mémoire vive, une alternative représentant un bon compromis pourrait être *PLN*, bien que la simplicité de ce modèle (une seule distribution possible) puisse lui faire manquer de polyvalence (par exemple sur des données de présence/absence ou de biomasses qui nécessiterait d'autres distributions, Niku *et al.* (2019b)).

Pour ce qui est d'inférer les interactions entre espèces à partir des données d'observations grâce à ces modèles joints, Blanchet *et al.* (2020) mettent en avant plusieurs limites, tant d'un point de vue

théorique que pratiques, qu'il semble important de rappeler ici. Un signal de co-occurrence peut par exemple, provenir de variables environnementales mal prises en compte par le modèle. Bien que les variables latentes puissent capturer ce signal, il n'est pas possible d'en différencier le signal biotique du signal d'une ou plusieurs variables environnementales manquantes (Blanchet *et al.* 2020). Lorsque le signal de corrélation entre taxa est faible, le nombre d'échantillons nécessaire pour détecter une interaction est extrêmement élevé. L'ordre de grandeur du nombre d'échantillons nécessaire peut dépasser de l'ordre du millier, ce qui n'est pas atteint dans ce travail. Enfin, les modèles de JSDM sont incapables de prédire les interactions négatives très fortes (e.g. cas de l'exclusion compétitive), car ces dernières sont incompatibles avec la coexistence (Blanchet *et al.* 2020). Ainsi, ces méthodes ne pourront sûrement pas dans un avenir proche, et en l'état, permettre d'inférer de manière réaliste des réseaux écologiques. Elles peuvent en revanche orienter les recherches vers des interactions probables et apporter des informations utiles pour mettre en perspectives des études réalisées à plus fines échelles, comme le montrent les retours des experts dans cette étude. Pour cela, il est important d'éviter les principaux écueils soulevés par Blanchet *et al.* (2020), ce que visait la méthodologie employée dans cette étude, notamment par l'utilisation d'*EMtree* et de données d'abondances.

## Perspectives

Certains des modèles utilisés dans ce travail présentent la propriété d'être assez flexible. *HMSC* et *GLLVM* peuvent en effet inclure des traits fonctionnels ou la phylogénie pour améliorer un peu plus les prédictions, notamment pour les espèces rares. En effet, pour modéliser des taxa les plus rares *HMSC* "emprunte" de l'information aux taxa les plus communs grâce aux traits et à la phylogénie. Ce type de modèle fait l'hypothèse que deux espèces aux traits similaires ou proche sur le plan phylogénétique réagissent de façon analogue aux gradients environnementaux (Ovaskainen & Abrego 2020). En plus d'améliorer les résultats, l'inclusion des traits ou de la phylogénie pourrait fournir une aide à l'interprétation des résultats grâce à une perspective un peu plus mécaniste. Une autre voie d'amélioration possible pour *HMSC* serait d'implémenter de façon explicite des effets aléatoires spatiaux et temporels en prenant en compte des notions de distances (Ovaskainen *et al.* 2017b).

Ces modèles pourraient également bénéficier d'un autre axe de validation se basant cette fois-ci sur la diversité. Une suite logique aux analyses spatiotemporelles mises en place dans ce travail de recherche serait les analyses de trajectoires des communautés telles que décrites par de Cáceres *et al.* (2019). L'utilisation des métriques proposées par de Cáceres *et al.* (2019) permettraient de comparer les trajectoires des communautés prédictes avec les trajectoires observées. Il serait ainsi possible de voir si certains des *JDSM* modélisent adéquatement la dynamique temporelle des communautés en termes de composition et non plus seulement en termes de richesse comme dans cette étude.

Pour ce qui est des perspectives sur l'inférence des réseaux d'interactions, il convient de noter que les réseaux d'interactions en écologie présentent des caractéristiques particulières (Pascual & Jennifer 2006). C'est pourquoi il serait intéressant de tester les capacités de reconstruction des réseaux d'inter-

actions de ces *JDSM* via des simulations, sur des réseaux aux propriétés connus basés sur des données d'abondances générées *in silico*. Cela permettrait de mieux comprendre si certaines propriétés particulières comme la faible imbrication des réseaux d'interaction observée est un artefact numérique ou bien une réelle propriété du réseau lié aux assemblages de polychètes étudiés ici. Les interactions inférées quant à elles pourraient profiter d'autres moyens de validation en plus des dires d'experts comme des méthodes issues de l'écologie trophique et notamment expérimentales (isotopes, contenus stomachaux; Majdi *et al.* (2018)). Toutefois, le manque de connaissance expérimentale sur ces taxa peut représenter le frein le plus important à la mise en place de cette forme de validation.

Enfin, si ce travail de par son focus réduit aux communautés de polychètes illustre bien la problématique commune en écologie marine du manque de données fiables sur tous les taxons à l'échelle de tout un écosystème, les méthodes testées ici peuvent possiblement trouver des applications à des jeux de données plus exhaustifs grâce aux dernières méthodes de métabarcoding (Djurhuus *et al.* 2020). Allier les *JDSM* à ces méthodes omiques permettrait d'obtenir une meilleure vision d'ensemble de la biodiversité et donc des réseaux d'interaction écologiques. Pour cela, l'un des principaux défis auxquels feront face les *JDSM* concernera l'optimisation de leur coût de calcul.

## Bibliographie

- Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F. & Early, R. *et al.* (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5, eaat4858.
- Ashcroft, M.B., King, D.H., Raymond, B., Turnbull, J.D., Wasley, J. & Robinson, S.A. (2017). Moving beyond presence and absence when examining changes in species distributions. *Global Change Biology*, 23, 2929-2940.
- Barbaro, L., Allan, E., Ampoorter, E., Castagneyrol, B., Charbonnier, Y. & Wandeler, H.D. *et al.* (2019). Biotic predictors complement models of bat and bird responses to climate and tree diversity in European forests. *Proceedings of the Royal Society B : Biological Sciences*, 286, 20182193.
- Bastolla, U., Fortuna, M.A., Pascual-García, A., Ferrera, A., Luque, B. & Bascompte, J. (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, 458, 1018-1020.
- Bates, A.E. & Morley, S.A. (2020). Interpreting empirical estimates of experimentally derived physiological and biological thermal limits in ectotherms. *Canadian Journal of Zoology*, 98, 237-244.
- Bertness, M.D. & Leonard, G.H. (1997). THE ROLE OF POSITIVE INTERACTIONS IN COMMUNITIES : LESSONS FROM INTERTIDAL HABITATS. *Ecology*, 78, 1976-1989.
- Bezanson, J., Edelman, A., Karpinski, S. & Shah, V.B. (2017). Julia : A fresh approach to numerical computing. *SIAM review*, 59, 65-98.

- Bhattacharya, A. & Dunson, D.B. (2011). Sparse Bayesian infinite factor models. *Biometrika*, 98, 291-306.
- Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*.
- Blonder, B. (2017). Hypervolume concepts in niche- and trait-based ecology. *Ecography*, 41, 1441-1455.
- Blowes, S.A., Supp, S.R., Antão, L.H., Bates, A., Bruelheide, H. & Chase, J.M. *et al.* (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science*, 366, 339-345.
- Box, G.E.P. (1979). Robustness in the Strategy of Scientific Model Building. In : *Robustness in Statistics*. Elsevier, p. 201-236.
- Boyd, P.W., Watson, A.J., Law, C.S., Abraham, E.R., Trull, T. & Murdoch, R. *et al.* (2000). A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature*, 407, 695-702.
- Boyé, A. (2019). Diversité taxinomique et fonctionnelle des habitats benthiques dans l'espace et dans le temps : une perspective régionale et décennale. PhD thesis. Université de Bretagne Occidentale, Université de Montréal.
- Boyé, A., Legendre, P., Grall, J. & Gauthier, O. (2017). Constancy despite variability : Local and regional macrofaunal diversity in intertidal seagrass beds. *Journal of Sea Research*, 130, 107-122.
- Boyé, A., Thiébaut, Grall, J., Legendre, P., Broudin, C. & Houbin, C. *et al.* (2019). Trait-based approach to monitoring marine benthic data along 500 km of coastline. *Diversity and Distributions*, 25, 1879-1896.
- Busby, J. (1991). BIOCLIM : a bioclimate analysis and prediction system. *Plant protection quarterly*, 6, 8—9.
- Cáceres, M.D., Coll, L., Legendre, P., Allen, R.B., Wiser, S.K. & Fortin, M.-J. *et al.* (2019). Trajectory analysis in community ecology. *Ecological Monographs*, 89, e01350.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y. & Cho, H. *et al.* (2019). *xgboost : Extreme Gradient Boosting*.
- Chiquet, J., Robin, S. & Mariadassou, M. (2019). Variational Inference for sparse network reconstruction from count data. In : *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research (éd. Chaudhuri, K. & Salakhutdinov, R.). PMLR, Long Beach, California, USA, p. 1162-1171.
- Crookston, N.L. & Finley, A.O. (2007). yaImpute : An R Package for kNN Imputation. *Journal of Statistical Software*, 23.
- Defeo, O., Barboza, C.A.M., Barboza, F.R., Aeberhard, W.H., Cabrini, T.M.B. & Cardoso, R.S. *et al.*

- (2017). Aggregate patterns of macrofaunal diversity : An interocean comparison. *Global Ecology and Biogeography*, 26, 823-834.
- Defeo, O. & McLachlan, A. (2005). Patterns, processes and regulatory mechanisms in sandy beach macrofauna : a multi-scale analysis. *Marine Ecology Progress Series*, 295, 1-20.
- Defeo, O. & McLachlan, A. (2013). Global patterns in sandy beach macrofauna : Species richness, abundance, biomass and body size. *Geomorphology*, 199, 106-114.
- Defeo, O., McLachlan, A., Schoeman, D.S., Schlacher, T.A., Dugan, J. & Jones, A. et al. (2009). Threats to sandy beach ecosystems : A review. *Estuarine, Coastal and Shelf Science*, 81, 1-12.
- Dehling, D.M. & Stouffer, D.B. (2018). Bringing the Eltonian niche into functional diversity. *Oikos*, 127, 1711-1723.
- Derrien-Courtel, S., Androuin, T., Ar Gall, E., Aublet, E., Bouriat, A. & Boyé, A. et al. (2017). *Le REBENT-II Bretagne - Surveillance du Benthos du littoral breton.* ( No. Rapport final-Vf2-11/03/2019). REBENT.
- Djurhuus, A., Closek, C.J., Kelly, R.P., Pitz, K.J., Michisaki, R.P. & Starks, H.A. et al. (2020). Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nature Communications*, 11.
- Dunne, J., Williams, R. & Martinez, N. (2004). Network structure and robustness of marine food webs. *Marine Ecology Progress Series*, 273, 291-302.
- Edgar, G.J., Bates, A.E., Bird, T.J., Jones, A.H., Kininmonth, S. & Stuart-Smith, R.D. et al. (2016). New Approaches to Marine Conservation Through the Scaling Up of Ecological Data. *Annual Review of Marine Science*, 8, 435-461.
- Elith, J., Leathwick, J.R. & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802-813.
- Elton, C. (2001). *Animal ecology*. University of Chicago Press, Chicago.
- Estes, L., Elsen, P.R., Treuer, T., Ahmed, L., Caylor, K. & Chang, J. et al. (2018). The spatial and temporal domains of modern ecology. *Nature Ecology & Evolution*, 2, 819-826.
- Ferrier, S. & Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, 43, 393-404.
- Gagnon, K., Rinde, E., Bengil, E.G.T., Carugati, L., Christianen, M.J.A. & Danovaro, R. et al. (2020). Facilitating foundation species : The potential for plantbivalve interactions to improve habitat restoration success. *Journal of Applied Ecology*, 57, 1161-1179.
- Gavish, Y., Marsh, C.J., Kuemmerlen, M., Stoll, S., Haase, P. & Kunin, W.E. (2017). Accounting for biotic interactions through alpha-diversity constraints in stacked species distribution models. *Methods in Ecology and Evolution*, 8, 1092-1102.

- Gelman, A. & Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7, 457-472.
- Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D. & Nyman, T. *et al.* (2018). Bringing Elton and Grinnell together : a quantitative framework to represent the biogeography of ecological interaction networks. *Ecography*, 42, 401-415.
- Gribben, P.E., Angelini, C., Altieri, A.H., Bishop, M.J., Thomsen, M.S. & Bulleri, F. (2019). Facilitation Cascades in Marine Ecosystems : A Synthesis and Future Directions. In : *Oceanography and Marine Biology*. CRC Press, p. 127-168.
- Grinnell, J. (1917). The Niche-Relationships of the California Thrasher. *The Auk*, 34, 427-433.
- Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017). *Habitat Suitability and Distribution Models*. Cambridge University Press.
- Hardin, G. (1960). The Competitive Exclusion Principle. *Science*, 131, 1292-1297.
- Harrison, X.A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, 2, e616.
- Hastie, T.J. & Pregibon, D. (1992). Generalized linear models. In : *Statistical models in S* (éd. Chambers, J.M. & Hastie, T.J.). Chapman & Hall/CRC, p. 195-247.
- Hawkins, S.J., Pack, K.E., Firth, L.B., Mieszkowska, N., Evans, A.J. & Martins, G.M. *et al.* (2019). The Intertidal Zone of the North-East Atlantic Region : Pattern and Process. In : *Interactions in the Marine Benthos : Global Patterns and Processes*, Systematics Association Special vol. Series (éd. Hawkins, S.J., Bohn, K., Firth, L.B. & Williams, G.A.). Cambridge University Press, p. 7-46.
- Hily, C. & Bouteille, M. (1999). Modifications of the specific diversity and feeding guilds in an intertidal sediment colonized by an eelgrass meadow (*Zostera marina*) (Brittany, France). *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie*, 322, 1121-1131.
- Hortal, J., Bello, F. de, Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46, 523-549.
- Hui, F.K.C. (2016). boral – Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r. *Methods in Ecology and Evolution*, 7, 744-750.
- Hutchinson, G.E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415-427.
- Isbell, F., Gonzalez, A., Loreau, M., Cowles, J., Diaz, S. & Hector, A. *et al.* (2017). Linking the influence and dependence of people on biodiversity across scales. *Nature*, 546, 65-72.
- Jankowska, E., Michel, L.N., Lepoint, G. & Włodarska-Kowalczuk, M. (2019). Stabilizing effects of

- seagrass meadows on coastal water benthic food webs. *Journal of Experimental Marine Biology and Ecology*, 510, 54-63.
- Jumars, P.A., Dorgan, K.M. & Lindsay, S.M. (2015). Diet of Worms Emended : An Update of Poly-chaete Feeding Guilds. *Annual Review of Marine Science*, 7, 497-520.
- Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2013). Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, 5, 54-64.
- Leach, K., Montgomery, W.I. & Reid, N. (2016). Modelling the influence of biotic factors on species distribution patterns. *Ecological Modelling*, 337, 96-106.
- Lecointre, G. & Le Guyader, H. (2001). *Classification phylogénétique du vivant*. 3<sup>e</sup> edn. Belin.
- Lefcheck, J.S., Hughes, B.B., Johnson, A.J., Pfirrmann, B.W., Rasher, D.B. & Smyth, A.R. *et al.* (2019). Are coastal habitats important nurseries? A meta-analysis. *Conservation Letters*, e12645.
- Legendre, P. (2005). Species associations : the Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics*, 10, 226-245.
- Leibold, M.A. & Chase, J.M. (2018). *Metacommunity Ecology, Volume 59*. Princeton University Press.
- Liaw, A. & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2, 18-22.
- Magurran, A.E., Dornelas, M., Moyes, F. & Henderson, P.A. (2019). Temporal \$\backslash\$upbeta\$ diversityA macroecological perspective. *Global Ecology and Biogeography*, 28, 1949-1960.
- Majdi, N., Hette-Tronquart, N., Auclair, E., Bec, A., Chouvelon, T. & Cognie, B. *et al.* (2018). There's no harm in having too much : A comprehensive toolbox of methods in trophic ecology. *Food Webs*, 17, e00100.
- Martinez, N.D. (1992). Constant Connectance in Community Food Webs. *The American Naturalist*, 139, 1208-1218.
- McDevitt-Irwin, J., Iacarella, J. & Baum, J. (2016). Reassessing the nursery role of seagrass habitats from temperate to tropical regions : a meta-analysis. *Marine Ecology Progress Series*, 557, 133-143.
- 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.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. (2019). *e1071 : Misc Functions of the Department of Statistics, Probability Theory Group (Formerly : E1071)*, TU Wien.
- Milborrow, S., Hastie, T., Tibshirani, R., Miller, A. & Lumley, T. (2019). *earth : Multivariate Adaptive Regression Splines*.
- Momal, R., Robin, S. & Ambroise, C. (2020). Tree-based inference of species interaction networks from abundance data. *Methods in Ecology and Evolution*, 11, 621-632.

- Niku, J., Brooks, W., Herliansyah, R., Hui, F.K.C., Taskinen, S. & Warton, D.I. (2019a). *gllvm : Generalized Linear Latent Variable Models*.
- Niku, J., Hui, F.K.C., Taskinen, S. & Warton, D.I. (2019b). *gllvm* : Fast analysis of multivariate abundance data with generalized linear latent variable models in r. *Methods in Ecology and Evolution*.
- Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J. & Anttila, J. et al. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, e01370.
- Nordström, M.C., Aarnio, K., Törnroos, A. & Bonsdorff, E. (2015). Nestedness of trophic links and biological traits in a marine food web. *Ecosphere*, 6, art161.
- Ovaskainen, O. & Abrego, N. (2020). *Joint Species Distribution Modelling : With Applications in R. Ecology, Biodiversity and Conservation*. Cambridge University Press.
- Ovaskainen, O., Tikhonov, G., Dunson, D., Grøtan, V., Engen, S. & Sæther, B.-E. et al. (2017a). How are species interactions structured in species-rich communities? A new method for analysing time-series data. *Proceedings of the Royal Society B : Biological Sciences*, 284, 20170768.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L. & Dunson, D. et al. (2017b). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20, 561-576.
- Paine, R.T. (1966). Food Web Complexity and Species Diversity. *The American Naturalist*, 100, 65-75.
- Pascual, M. & Jennifer, D. (2006). *Ecological networks : linking structure to dynamics in food webs*. Oxford University Press, Oxford New York.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.
- Poisot, T., Bélisle, Z., Hoebelke, L., Stock, M. & Szefer, P. (2019). EcologicalNetworks.jl : analysing ecological networks of species interactions. *Ecography*, 42, 1850-1861.
- Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2015). The structure of probabilistic networks. *Methods in Ecology and Evolution*, 7, 303-312.
- Potts, J.M. & Elith, J. (2006). Comparing species abundance models. *Ecological Modelling*, 199, 153-163.
- Quillien, N., Nordström, M.C., Bris, H.L., Bonsdorff, E. & Grall, J. (2017). Green tides on inter- and subtidal sandy shores : differential impacts on infauna and flatfish. *Journal of the Marine Biological Association of the United Kingdom*, 98, 699-712.
- R Core Team. (2019). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Santos, C.B. de los, Krause-Jensen, D., Alcoverro, T., Marbà, N., Duarte, C.M. & Katwijk, M.M. van *et al.* (2019). Recent trend reversal for declining European seagrass meadows. *Nature Communications*, 10.
- Soininen, J. (2014). A quantitative analysis of species sorting across organisms and ecosystems. *Ecology*, 95, 3284-3292.
- Staniczenko, P.P.A., Sivasubramaniam, P., Suttle, K.B. & Pearson, R.G. (2017). Linking macroecology and community ecology : refining predictions of species distributions using biotic interaction networks. *Ecology Letters*, 20, 693-707.
- Sunday, J.M., Bates, A.E. & Dulvy, N.K. (2012). Thermal tolerance and the global redistribution of animals. *Nature Climate Change*, 2, 686-690.
- Sunday, J.M., Fabricius, K.E., Kroeker, K.J., Anderson, K.M., Brown, N.E. & Barry, J.P. *et al.* (2016). Ocean acidification can mediate biodiversity shifts by changing biogenic habitat. *Nature Climate Change*, 7, 81-85.
- Thurman, L.L., Barner, A.K., Garcia, T.S. & Chestnut, T. (2019). Testing the link between species interactions and species co-occurrence in a trophic network. *Ecography*.
- Tikhonov, G., Opedal, O., Abrego, N., Lehikoinen, A. & Ovaskainen, O. (2019a). Joint species distribution modelling with HMSC-R. *bioRxiv*.
- Tikhonov, G., Ovaskainen, O., Oksanen, J., de Jonge, M., Opedal, O. & Dallas, T. (2019b). *Hmsc : Hierarchical Model of Species Communities*.
- Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D. & Berghe, E.V. *et al.* (2010). Global patterns and predictors of marine biodiversity across taxa. *Nature*, 466, 1098-1101.
- Trinh, G., Rungie, C., Wright, M., Driesener, C. & Dawes, J. (2013). Predicting future purchases with the Poisson log-normal model. *Marketing Letters*, 25, 219-234.
- Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, 7.
- Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S. & Walker, S.C. *et al.* (2015). So many variables : joint modeling in community ecology. *Trends in Ecology & Evolution*, 30, 766-779.
- Waycott, M., Duarte, C.M., Carruthers, T.J.B., Orth, R.J., Dennison, W.C. & Olyarnik, S. *et al.* (2009). Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proceedings of the National Academy of Sciences*, 106, 12377-12381.
- Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J. & Damgaard, C.F. *et al.* (2012). The role of biotic interactions in shaping distributions and realised assemblages of species : implications for species distribution modelling. *Biological Reviews*, 88, 15-30.

- Witman, J.D., Lamb, R.W. & Byrnes, J.E.K. (2015). Towards an integration of scale and complexity in marine ecology. *Ecological Monographs*, 85, 475-504.
- Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 73, 3-36.
- Woodin, S.A., Bell, S.S., Grant, J., Snelgrove, P.V.R. & Wethey, D.S. (2019). Shallow Water Muddy Sands of the North-West Atlantic Ocean : Latitudinal Patterns in Interactions and Processes. In : *Interactions in the Marine Benthos : Global Patterns and Processes*, Systematics Association Special vol. Series (éd. Hawkins, S.J., Bohn, K., Firth, L.B. & Williams, G.A.). Cambridge University Press, p. 128-163.
- Worm, B. & Tittensor, D.P. (2018). *A Theory of Global Biodiversity (MPB-60)*. Princeton University Press.
- Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F. & Parnell, S. et al. (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution*, 33, 790-802.
- Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T. & Wüest, R.O. (2019). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, 47, 101-113.

## Annexes

### Liste de *Species-Distribution Models*

Tableau 7 : Quelques-uns des modèles de SDM les plus populaires actuellement (adapté de Ovaskainen & Abrego (2020)).

<i>Single-species distribution model</i>	Référence
<i>Boosted regression trees (BRT)</i>	Elith et al. (2008)
<i>Generalised additive model (GAM)</i>	Wood (2011)
<i>Generalised linear model (GLM)</i>	R Core Team (2019)
<i>Gradient nearest neighbour (GNN)</i>	Crookston & Finley (2007)
<i>Maximum-entropy approach (MaxEnt)</i>	Phillips et al. (2006)
<i>Multivariate adaptive regression spline (MARS-COMM)</i>	Milborrow et al. (2019)
<i>Random forest (RF)</i>	Liaw & Wiener (2002)
<i>Support vector machine (SVM)</i>	Meyer et al. (2019)
<i>Gradient extreme boosting (XGB)</i>	Chen et al. (2019)

## Partitionnement de la variance expliquée

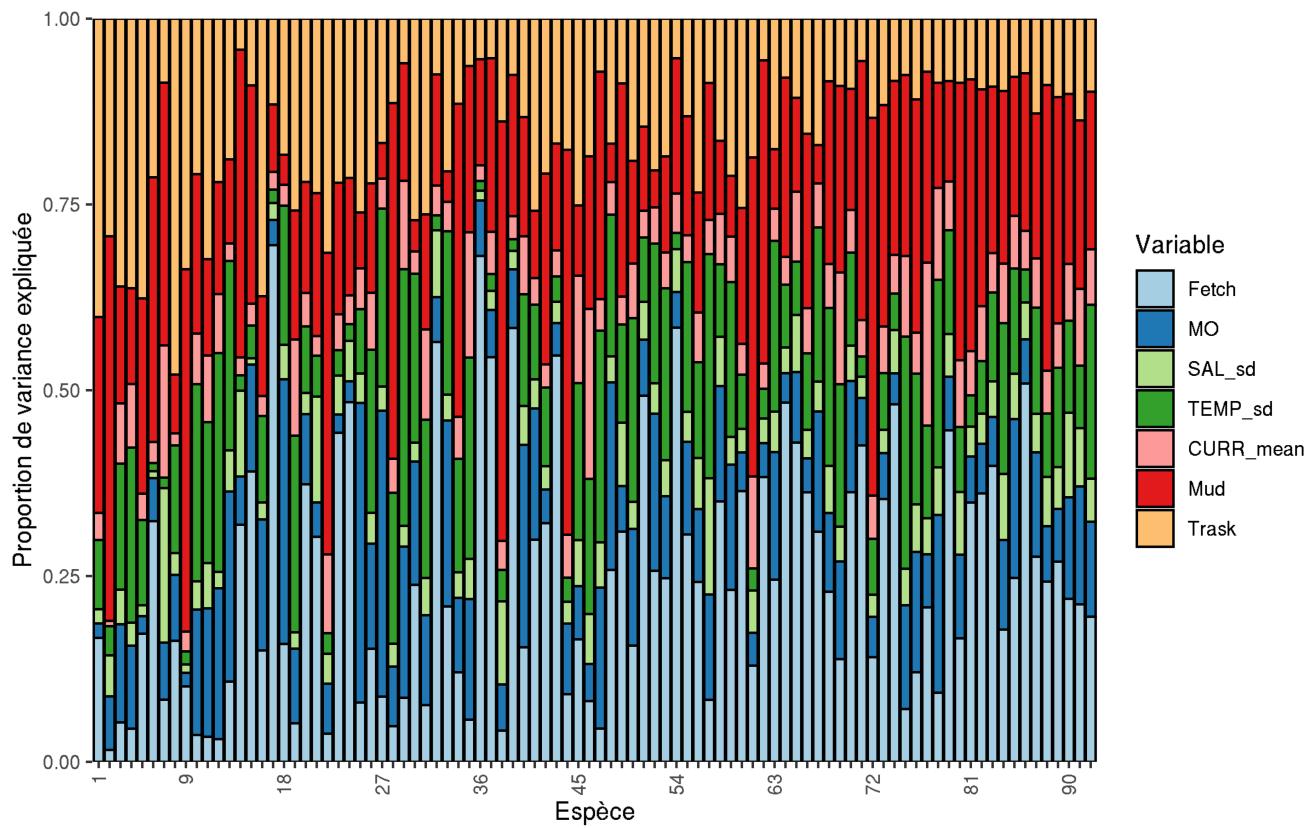


Figure 10 : Partitionnement de la variance de l'abondance des 92 taxons de polychètes expliquée par les différentes variables environnementales du modèle *HMSC\_reg*. Les espèces sont classées par ordre décroissant de  $SR^2$  (voir Tableau 8)

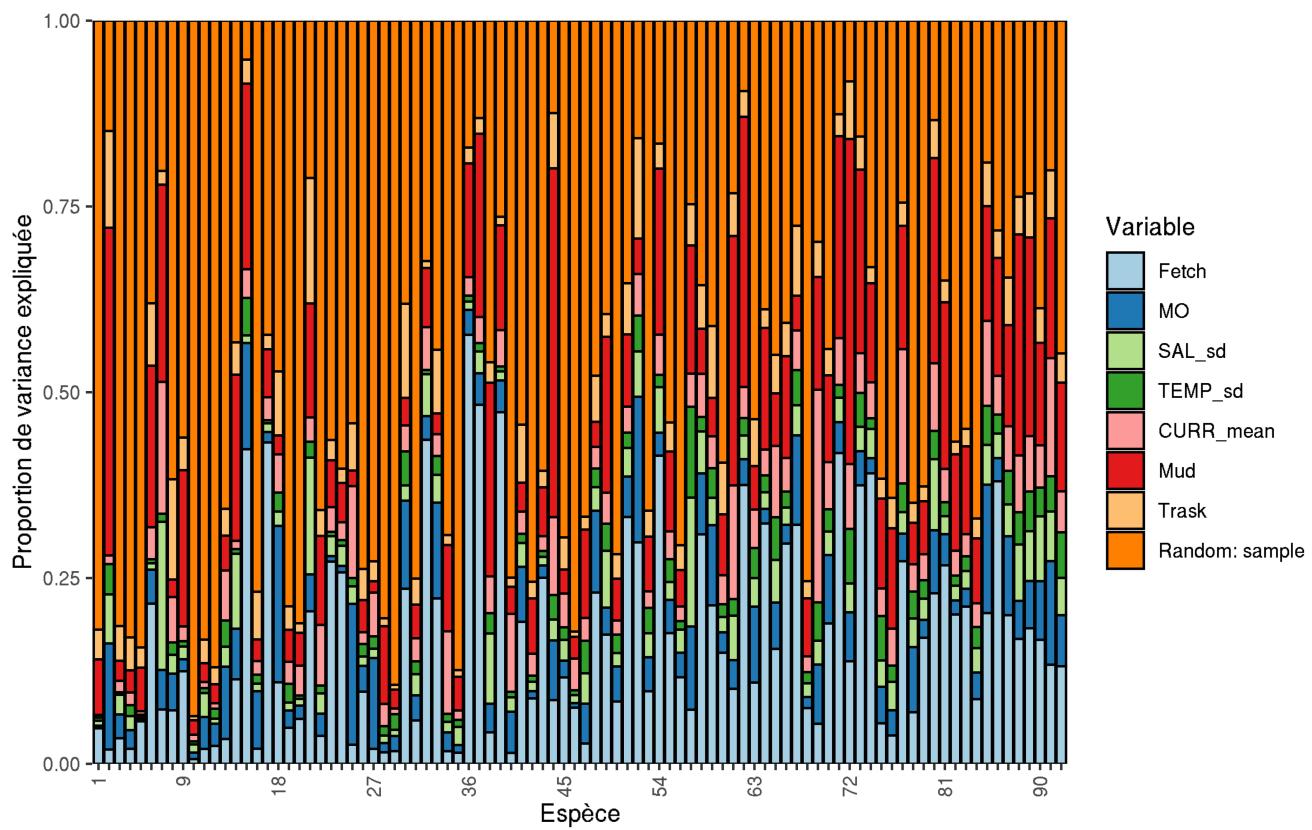


Figure 11 : Partitionnement de la variance de l'abondance des 92 taxons de polychètes expliquée par les différentes variables environnementales du modèle *HMSC\_samp*. Les espèces sont classées par ordre décroissant de  $SR^2$  (voir Tableau 8)

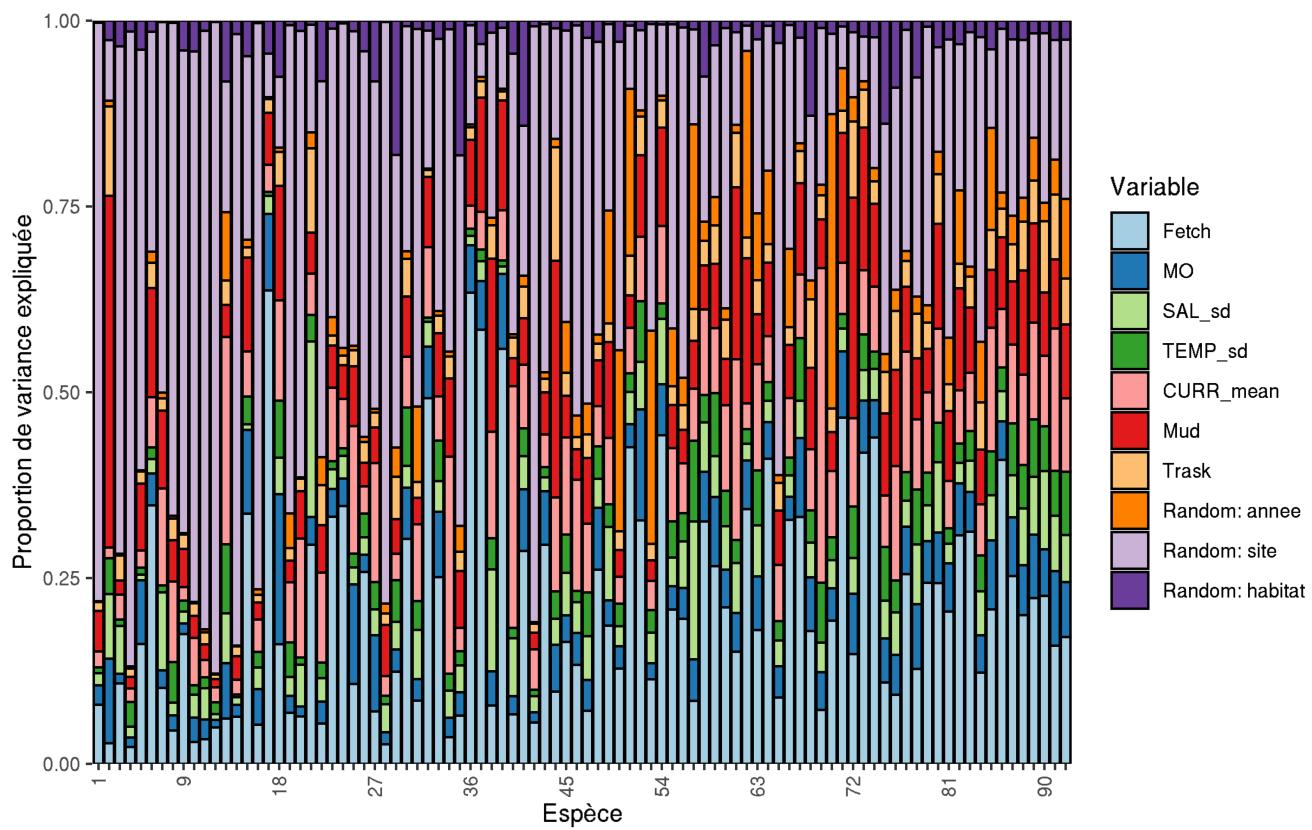


Figure 12 : Partitionnement de la variance de l'abondance des 92 taxons de polychètes expliquée par les différentes variables environnementales du modèle *HMSC\_hier*. Les espèces sont classées par ordre décroissant de  $SR^2$  (voir Tableau 8)

## Espèces de polychètes

Tableau 8 : Tableau de résumant les propriétés de la communauté des Polychètes. Les numéros correspondent à ceux des figures des effets environnementaux et du partitionnement de variance, les espèces sont ordonnées par  $SR^2$  moyen décroissant. Les status trophiques ont été confirmés par l'un de nos experts.

N°	specie	Status trophique	habitat
1	<i>Notomastus latericeus</i>	Dépositivore	Sédiment + Herbier
2	<i>Nephtys cirrosa</i>	Prédateur/Charognard	Sédiment + Herbier
3	<i>Lumbrineris</i> spp.	Prédateur/Charognard	Sédiment + Herbier
4	<i>Aonides oxycephala</i>	Dépositivore	Sédiment + Herbier
5	<i>Euclymene</i> spp.	Dépositivore	Sédiment + Herbier
6	<i>Glycera tridactyla</i>	Prédateur/Charognard	Sédiment + Herbier
7	<i>Owenia fusiformis</i>	Dépositivore	Sédiment + Herbier
8	<i>Lysidice unicornis</i>	Prédateur/Charognard	Sédiment + Herbier
9	<i>Magelona filiformis</i>	Dépositivore	Sédiment

N°	specie	Status trophique	habitat
10	<i>Platynereis dumerilii</i>	Brouteur	Sédiment + Herbier
11	<i>Cirriformia tentaculata</i>	Dépositoire	Sédiment + Herbier
12	<i>Perinereis cultrifera</i>	Brouteur	Sédiment + Herbier
13	<i>Heteromastus filiformis</i>	Dépositoire	Sédiment + Herbier
14	<i>Leiochone leiopygos</i>	Dépositoire	Sédiment + Herbier
15	<i>Scoloplos (Scoloplos) armiger</i>	Dépositoire	Sédiment + Herbier
16	<i>Polycirrus</i> spp.	Dépositoire	Sédiment + Herbier
17	<i>Melinna palmata</i>	Dépositoire	Sédiment + Herbier
18	<i>Pholoe inornata</i>	Prédateur/Charognard	Sédiment + Herbier
19	<i>Cauilleriella alata</i>	Dépositoire	Sédiment + Herbier
20	<i>Acromegalomma</i> spp.	Suspensivore	Sédiment + Herbier
21	<i>Eteone longa</i>	Prédateur/Charognard	Sédiment + Herbier
22	<i>Sigalion mathildae</i>	Prédateur/Charognard	Sédiment
23	<i>Chaetozone</i> spp.	Dépositoire	Sédiment + Herbier
24	<i>Chaetozone gibber</i>	Dépositoire	Sédiment + Herbier
25	<i>Malmgrenia arenicolae</i>	Prédateur/Charognard	Sédiment + Herbier
26	<i>Hilbigneris gracilis</i>	Dépositoire/Prédateur/Charognard	Sédiment + Herbier
27	<i>Scalibregma celticum</i>	Dépositoire	Herbier
28	<i>Spio</i> spp.	Dépositoire	Sédiment + Herbier
29	<i>Malacoceros girardi</i>	Dépositoire	Herbier
30	<i>Harmothoe spinifera</i>	Prédateur/Charognard	Sédiment + Herbier
31	<i>Cauilleriella bioculata</i>	Dépositoire	Sédiment + Herbier
32	<i>Aponuphis bilineata</i>	Prédateur/Charognard	Sédiment + Herbier
33	<i>Piromis eruca</i>	Dépositoire	Herbier
34	<i>Phyllodoce mucosa</i>	Prédateur/Charognard	Sédiment + Herbier
35	<i>Neanthes acuminata</i>	Omnivore	Herbier
36	<i>Ampharete</i> spp.	Dépositoire	Sédiment + Herbier
37	<i>Paradoneis armata</i>	Dépositoire	Sédiment + Herbier
38	<i>Scolelepis</i> spp.	Dépositoire	Sédiment + Herbier
39	<i>Poecilochaetus serpens</i>	Dépositoire	Sédiment + Herbier
40	<i>Sthenelais boa</i>	Prédateur/Charognard	Sédiment + Herbier
41	<i>Harmothoe impar</i>	Prédateur/Charognard	Herbier
42	<i>Marphysa bellii</i>	Prédateur/Charognard	Sédiment + Herbier
43	<i>Lanice conchilega</i>	Dépositoire	Sédiment + Herbier
44	<i>Orbinia latreillii</i>	Dépositoire	Sédiment + Herbier
45	<i>Schistomerings rudolphi</i>	Omnivore	Sédiment + Herbier
46	<i>Syllis garciai</i>	Omnivore	Sédiment + Herbier
47	<i>Capitella capitata</i>	Dépositoire	Sédiment + Herbier

N°	specie	Status trophique	habitat
48	Terebellides spp.	Dépositivore	Herbier
49	Pygospio elegans	Dépositivore	Sédiment + Herbier
50	Scoletoma spp.	Prédateur/Charognard	Sédiment + Herbier
51	Prionospio spp.	Dépositivore	Sédiment + Herbier
52	Eunice vittata	Prédateur/Charognard	Sédiment + Herbier
53	Macroclymene santanderensis	Dépositivore	Sédiment + Herbier
54	Lagis koreni	Dépositivore	Sédiment + Herbier
55	Polydora spp.	Dépositivore	Sédiment + Herbier
56	Protodorvillea kefersteini	Omnivore	Sédiment + Herbier
57	Capitella minima	Dépositivore	Sédiment + Herbier
58	Eupolymnia nebulosa	Dépositivore	Sédiment + Herbier
59	Pista spp.	Dépositivore	Sédiment + Herbier
60	Eumida spp.	Prédateur/Charognard	Sédiment + Herbier
61	Spiophanes bombyx	Dépositivore	Sédiment + Herbier
62	Paradoneis lyra	Dépositivore	Sédiment + Herbier
63	Amphitritides gracilis	Dépositivore	Sédiment + Herbier
64	Parexogone hebes	Omnivore	Sédiment + Herbier
65	Arenicola marina	Dépositivore	Sédiment + Herbier
66	Nicomache lumbricalis	Dépositivore	Sédiment + Herbier
67	Leodice harassii	Prédateur/Charognard	Herbier
68	Microspio spp.	Dépositivore	Sédiment + Herbier
69	Mysta picta	Prédateur/Charognard	Sédiment + Herbier
70	Cirratulidae spp.	Dépositivore	Sédiment + Herbier
71	Phyllodoce lineata	Prédateur/Charognard	Sédiment + Herbier
72	Nephtys spp.	Prédateur/Charognard	Sédiment + Herbier
73	Hypereteone foliosa	Prédateur/Charognard	Sédiment
74	Petaloproctus terricolus	Dépositivore	Sédiment + Herbier
75	Malacoceros spp.	Dépositivore	Herbier
76	Malacoceros fuliginosus	Dépositivore	Sédiment + Herbier
77	Paraonidae spp.	Dépositivore	Sédiment + Herbier
78	Syllis spp.	Omnivore	Herbier
79	Sphaerosyllis spp.	Omnivore	Sédiment + Herbier
80	Malmgrenia spp.	Prédateur/Charognard	Sédiment + Herbier
81	Phyllodoce laminosa	Prédateur/Charognard	Sédiment + Herbier
82	Maldanidae spp.	Dépositivore	Sédiment + Herbier
83	Sabellida pavonina	Suspensivore	Sédiment + Herbier
84	Nereis spp.	Omnivore	Sédiment + Herbier
85	Psamathe fusca	Prédateur/Charognard	Sédiment + Herbier

N°	specie	Status trophique	habitat
86	Chaetozone zetlandica	Dépositivore	Sédiment + Herbier
87	Eumida parva	Prédateur/Charognard	Sédiment + Herbier
88	Orbinia sertulata	Dépositivore	Sédiment + Herbier
89	Goniada spp.	Prédateur/Charognard	Sédiment + Herbier
90	Glycera spp.	Prédateur/Charognard	Sédiment + Herbier
91	Malmgrenia andreapolis	Prédateur/Charognard	Sédiment + Herbier
92	Spirobranchus spp.	Suspensivore	Sédiment + Herbier

## Réseaux reconstruits

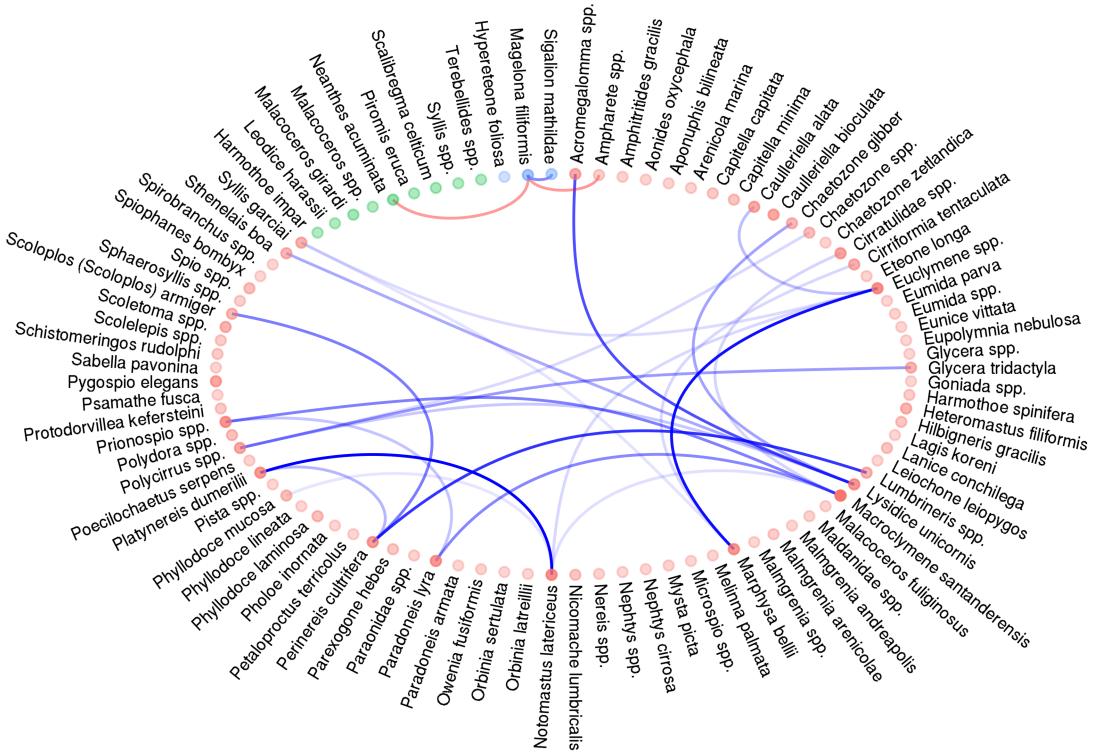


Figure 13 : Réseau reconstruit présentant les probabilités d'interaction inférées par *EMtree* sur la base des corrélations résiduelles associées à l'effet aléatoire du modèle *GLLVM*. Les points rouges représentent les taxa de polychètes retrouvés dans les deux habitats, les verts retrouvés uniquement dans les herbiers et les bleus dans les sédiments meubles. Seules les arrêtes ayant une probabilité supérieure à 0,2 sont affichées. L'opacité des arrêtes est proportionnelle à leur probabilité. L'opacité des points est proportionnelle à leur importance dans le réseau

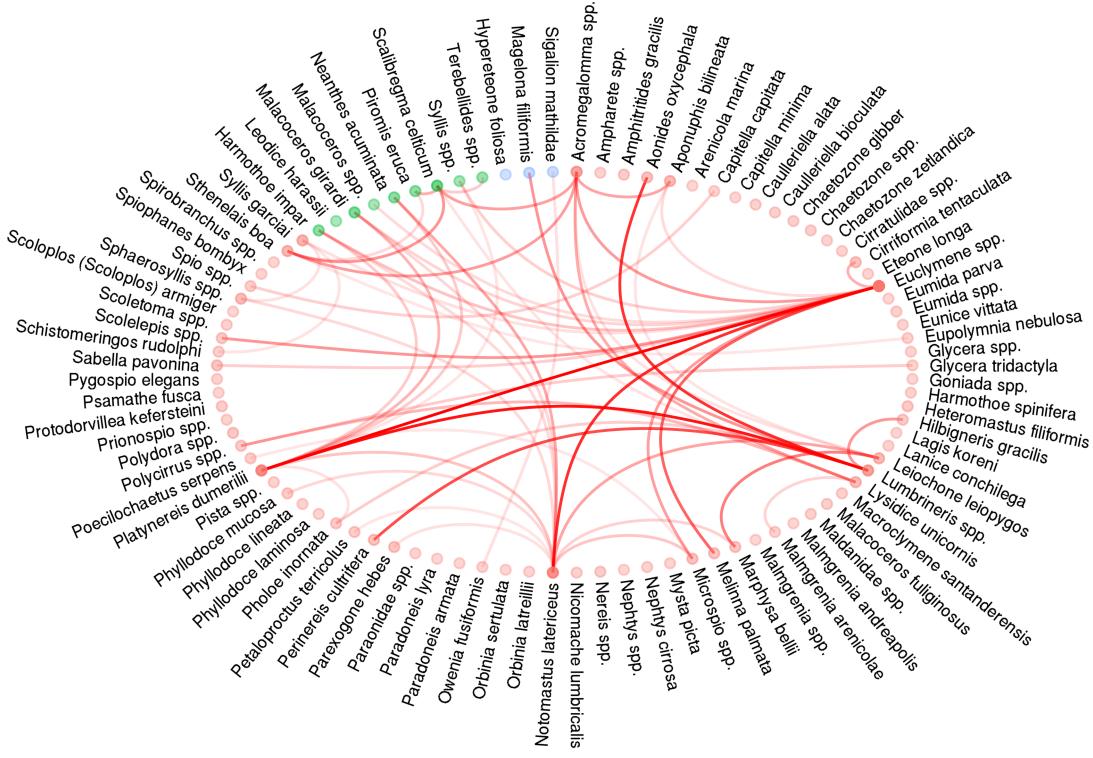


Figure 14 : Réseau présentant les probabilités d’interaction inférées par *EMtree* sur la base des corrélations résiduelles associées à l’effet aléatoire du modèle *PLN*. Les points rouges représentent les taxa de polychètes retrouvés dans les deux habitats, les verts retrouvés uniquement dans les herbières et les bleus dans les sédiments meubles. Seules les arrêtes ayant une probabilité supérieure à 0,2 sont affichées. L’opacité des arrêtes est proportionnelle à leur probabilité. L’opacité des points est proportionnelle à leur importance dans le réseau.

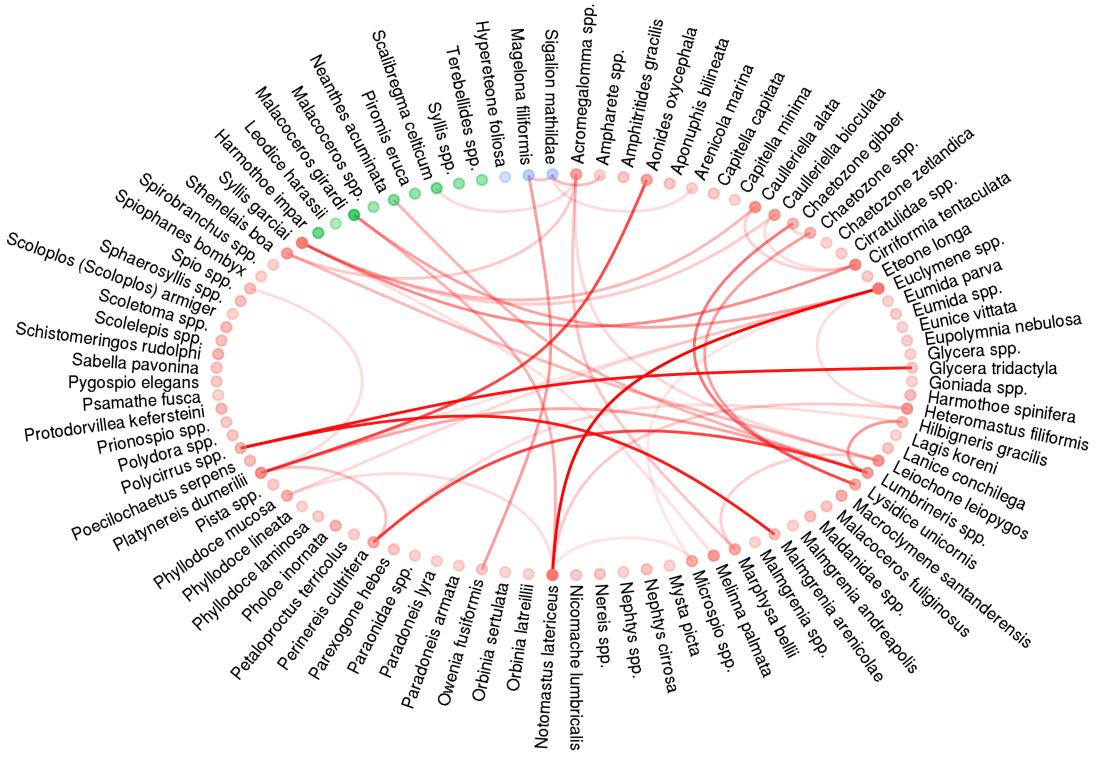


Figure 15 : Réseau reconstruit présentant les probabilités d'interaction inférées par *EMtree* sur la base des corrélations résiduelles associées à l'effet aléatoire spatial du modèle *HMSC\_hier*. Les points rouges représentent les taxa de polychètes retrouvés dans les deux habitats, les verts retrouvés uniquement dans les herbiers et les bleus dans les sédiments meubles. Seules les arrêtes ayant une probabilité supérieur à 0,2 sont affichées. L'opacité des arrêtes est proportionnelle à leur probabilité. L'opacité des points est proportionnelle à leur importance dans le réseau.

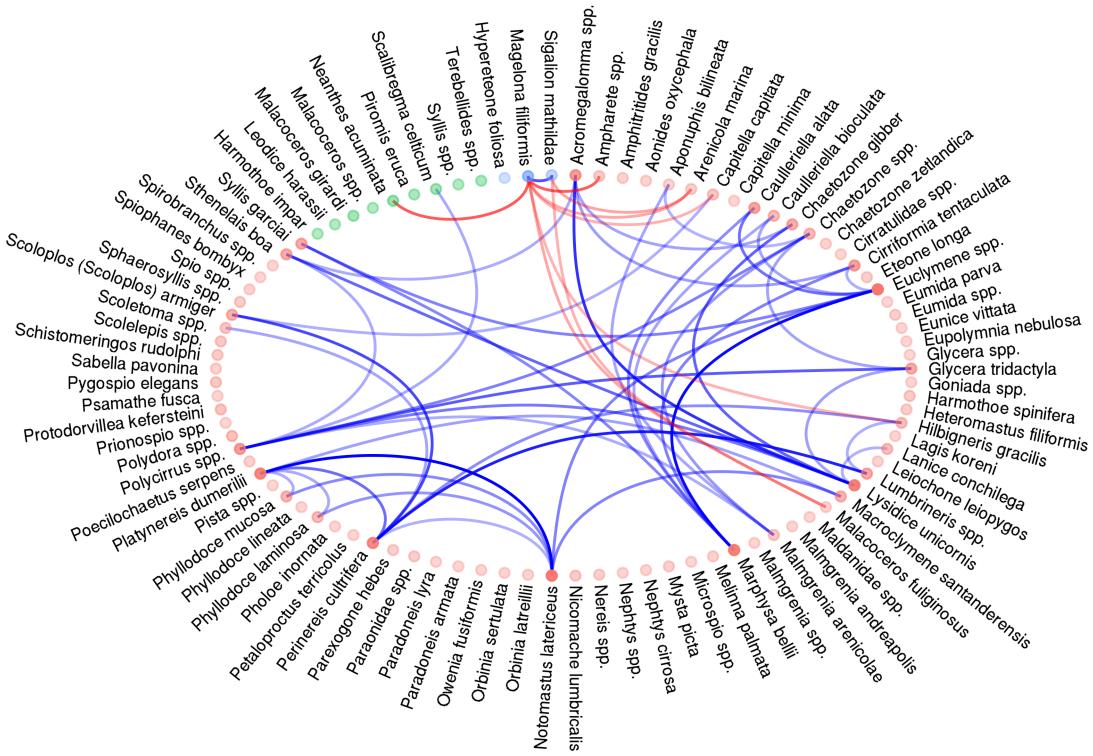


Figure 16 : Réseau reconstruit présentant les probabilités d'interaction inférées par *EMtree* sur la base des corrélations résiduelles associées à l'effet aléatoire site du modèle *HMSC\_hier*. Les points rouges représentent les taxa de polychètes retrouvés dans les deux habitats, les verts retrouvés uniquement dans les herbiers et les bleus dans les sédiments meubles. Seules les arrêtes ayant une probabilité supérieure à 0,2 sont affichées. L'opacité des arrêtes est proportionnelle à leur probabilité. L'opacité des points est proportionnelle à leur importance dans le réseau.

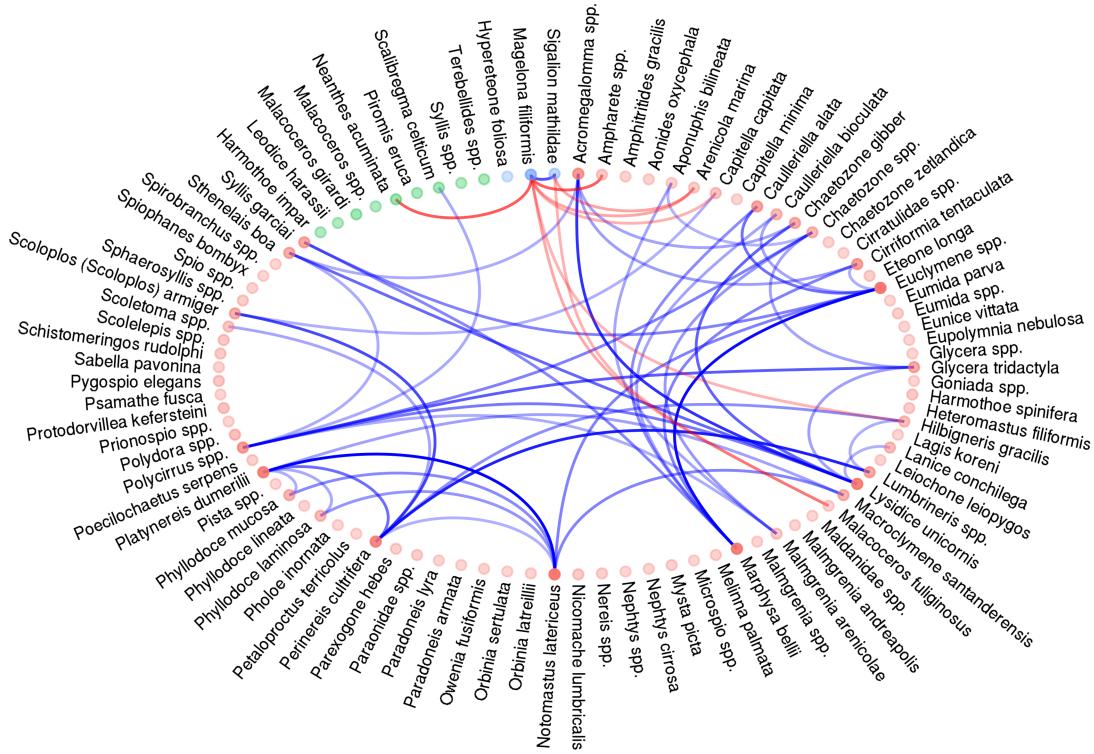


Figure 17 : Réseau reconstruit présentant les probabilités d'interaction inférées par *EMtree* sur la base des corrélations résiduelles associées à l'effet aléatoire habitat du modèle *HMSC\_hier*. Les points rouges représentent les taxa de polychètes retrouvés dans les deux habitats, les verts retrouvés uniquement dans les herbiers et les bleus dans les sédiments meubles. Seules les arrêtes ayant une probabilité supérieur à 0,2 sont affichées. L'opacité des arrêtes est proportionnelle à leur probabilité. L'opacité des points est proportionnelle à leur importance dans le réseau.

## Résumé

Prédire les facteurs qui biotiques et abiotiques qui contribuent à la distribution des espèces au tour du monde est une tâche difficile. Ces dernières années, une nouvelle catégorie de modèles de distribution d'espèces est apparue : les modèles de distribution conjointe d'espèces. En plus de variables climatiques, certains de ces modèles prennent en compte d'autres types de données : comme des données phylogéniques ou de traits fonctionnels. Ces modèles offrent deux perspectives : celle de mieux prédire que les SDM classiques les patrons spatiotemporels de la distribution espèces et de permettre d'accéder aux interactions entre espèces. Observer et quantifier les interactions dans les écosystèmes marins peut-être une tâche difficile, spécifiquement pour les écosystèmes benthiques. L'utilisation de JSDM est un outil intéressant pour pouvoir accéder à cette information. Ainsi, ce travail a eu pour objectif d'explorer le potentiel de trois méthodes différentes de JSDM sur un jeu de donnée d'abondance d'une communauté benthique pour quantifier les effets abiotiques sur les patrons de diversité spatiotemporelle et d'appréhender les interactions au niveau de la communauté. Nos résultats montrent que tous les modèles expliquent correctement une forte proportion d'espèces. Toutefois, les trois mo-

dèles prédisent dans l'ensemble correctement la richesse spécifique, néanmoins aucun d'entre eux n'est capable de prédire l'abondance. Nos modèles montrent un fort effet des effets aléatoires, ce qui laisse à penser que d'autres variables environnementales structurent ces communautés biologiques. Enfin, les réseaux d'interactions reconstruits grâce à ces méthodes ne font pas consensus chez un panel d'experts : toutefois, certaines interactions sont correctement retrouvées par nos modèles. Les JSDM sont donc des outils intéressants à utiliser en écologie des communautés.

Predicting the biotic and abiotic factors that contribute to the distribution of species around the world is a difficult task. In recent years, a new category of species distribution patterns has emerged : joint species distribution patterns. In addition to climate variables, some of these models take into account other types of data : such as phylogenetic data or functional trait data. These models offer two perspectives : to better predict the spatiotemporal patterns of species distribution than conventional MDS and to provide access to interactions between species. Observing and quantifying interactions in marine ecosystems can be a difficult task, specifically for benthic ecosystems. The use of JSDM is an interesting tool to access this information. Thus, the objective of this work was to explore the potential of three different JSDM methods on a benthic community abundance dataset to quantify abiotic effects on patterns of spatiotemporal diversity and to understand interactions at the community level. Our results show that all models correctly explain a high proportion of species. However, all three models generally correctly predict species richness, yet none of them is able to predict abundance. Our models show a strong effect of random effects, suggesting that other environmental variables structure these biological communities. Finally, the networks of interactions reconstructed using these methods do not meet the consensus of a panel of experts : however, some interactions are correctly found by our models. JSDMs are therefore interesting tools to use in community ecology.

%