

Année universitaire : 2017 - 2018

Spécialité :

SML-Biologie

Parcours (et option éventuelle) :

Sciences Halieutiques et Aquacoles (REA)

Mémoire de Fin d'Études

- d'Ingénieur de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage
- de Master de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage
- d'un autre établissement (étudiant arrivé en M2)

Identification des espèces cibles et accessoires dans la pêche mixte démersale du golfe de Gascogne.

Par : Timothée JULIEN



Source : © Ifremer - Olivier DUGORNAY

Soutenu à Rennes le 12/09/2018

Devant le jury composé de :

Président (enseignant référent) : Marie-Pierre ETIENNE

Maître de stage : Anne-Sophie CORNOU, Stéphanie MAHEVAS et Youen VERMARD

Enseignant référent : Jérôme GUITTON

Autres membres du jury : Olivier LE PAPE (AGROCAMPUS OUEST, Rennes)

Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celle d'AGROCAMPUS OUEST

Ce document est soumis aux conditions d'utilisation
« Paternité-Pas d'Utilisation Commerciale-Pas de Modification 4.0 France »
disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr>

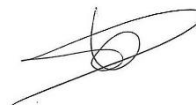


Confidentialité

Non Oui si oui : 1 an 5 ans 10 ans

Pendant toute la durée de confidentialité, aucune diffusion du mémoire n'est possible ⁽¹⁾.

Date et signature du maître de stage ⁽²⁾ : 28/08/2018
(ou de l'étudiant-entrepreneur)



A la fin de la période de confidentialité, sa diffusion est soumise aux règles ci-dessous (droits d'auteur et autorisation de diffusion par l'enseignant à renseigner).

Droits d'auteur

L'auteur⁽³⁾ **JULIEN Timothée**

autorise la diffusion de son travail (immédiatement ou à la fin de la période de confidentialité)

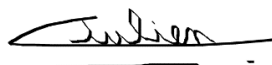
Oui Non

Si oui, il autorise

- la diffusion papier du mémoire uniquement⁽⁴⁾
- la diffusion papier du mémoire et la diffusion électronique du résumé
- la diffusion papier et électronique du mémoire (joindre dans ce cas la fiche de conformité du mémoire numérique et le contrat de diffusion)

(Facultatif) accepte de placer son mémoire sous licence Creative commons CC-BY-NC-Nd (voir Guide du mémoire Chap 1.4 page 6)

Date et signature de l'auteur : 28/08/2018



Autorisation de diffusion par le responsable de spécialisation ou son représentant

L'enseignant juge le mémoire de qualité suffisante pour être diffusé (immédiatement ou à la fin de la période de confidentialité)

Oui Non

Si non, seul le titre du mémoire apparaîtra dans les bases de données.

Si oui, il autorise

- la diffusion papier du mémoire uniquement⁽⁴⁾
- la diffusion papier du mémoire et la diffusion électronique du résumé
- la diffusion papier et électronique du mémoire

Date et signature de l'enseignant :

(1) L'administration, les enseignants et les différents services de documentation d'AGROCAMPUS OUEST s'engagent à respecter cette confidentialité.

(2) Signature et cachet de l'organisme

(3) Auteur = étudiant qui réalise son mémoire de fin d'études

(4) La référence bibliographique (= Nom de l'auteur, titre du mémoire, année de soutenance, diplôme, spécialité et spécialisation/Option) sera signalée dans les bases de données documentaires sans le résumé

Remerciements

En premier lieu, je remercie Anne-Sophie CORNOU, Stéphanie MAHEVAS et Youen VERMARD, encadrants de mon stage, qui ont accepté ma candidature et m'ont aidé en me fournissant de précieuses informations dans mes recherches. Ils ont su être présents et patient pour répondre à mes interrogations tout au long de la réalisation de cette étude.

Je souhaite remercier mes 2 camarades et amies Marine et Solène qui ont partagé cette aventure à mes côtés, sources de nombreux fous rires au quotidien. Encore merci pour m'avoir aidé et soutenu.

Je remercie également mon maître en matière blagues en tous genres (et occasionnellement R). Sans Olivier je n'aurais pas autant rigolé mais avant tout je n'oublie pas que ton aide m'a été précieuse.

Je ne peux que remercier l'équipe des mots fléchés Mathieu et Pierre pour les bons moments mais également pour l'aide et les discussions qui ont su m'éclairer.

Je remercie Arthur sans qui j'aurais dormi sous les ponts au début et à la fin de ce stage. Je te remercie également pour ton aide et les discussions autour de mon travail.

Cette année encore, je remercie la team cerise, Camille et Pauline pour m'avoir soutenu toujours dans la bonne humeur.

Je remercie Lucas, Andréa, Floriane, Louise, Camille, Vincent, et bien d'autres pour la bonne humeur.

Je souhaite également remercier l'ensemble des personnes au sein de l'équipe EMH avec qui j'ai partagé des pauses café enrichissantes.

Enfin, je remercie Fiona, qui m'a soutenue tous les jours et a corrigé mes rapports, ma famille, qui m'a aidé financièrement et affectueusement depuis le début.

Table des matières

Introduction	1
Matériels et méthodes	4
1. Zone d'étude.....	4
2. Données	4
2.1. OBSMER	4
2.2. SACROIS	6
3. Méthodes.....	6
3.1. Différentes échelles	7
3.1.1. Opérations de pêche	8
3.1.2. Séquences de pêche.....	10
3.2. Définition <i>a priori</i> du statut des individus statistiques	13
3.3. Caractérisation du statut des espèces	14
3.3.1. Méthodes d'analyse avec les rejets.....	14
3.3.2. Détermination des seuils	15
3.3.3. Méthodes d'analyse sans les rejets.....	15
3.3.4. Résumé de la démarche	15
3.3.5. Méthodes d'apprentissage sur SACROIS.....	16
Résultats	17
1. Classification à l'échelle de l'opération de pêche	17
1.1. Analyse de la classification des données de captures	17
1.2. Analyse de la classification des données de débarquements.....	20
2. Classification à l'échelle de la séquence de pêche	21
2.1. Analyse de la classification des données de captures	21
2.2. Analyse de la classification des données de débarquements.....	22
2.2.1. Méthode des k-means.....	22
2.2.2. Apprentissage par les arbres de classification.....	23
3. Comparaison des méthodes	24
4. Analyse de la classification supervisée des données SACROIS	26
Discussion	27
1. Statut des espèces	27
2. Limites des données	28
3. Limites des méthodes	29
Conclusion	31
Bibliographie	32
Annexes	36

Liste des figures

Figure 1. Composition des captures de chalutiers et senneurs démersaux décrivant les différentes catégories (encadrées et en italique) des rejets et les prises accidentelles sont retournées à la mer. MLS (Minimal Landing Size) est la taille minimale de débarquement. Ici est considérée la gestion dans le cadre de la Politique Commune de la Pêche avant la mise en place de l'obligation de débarquement (Fernandes et al., 2011).....	2
Figure 2. Position géographique des opérations de pêche échantillonnées pour la pêche mixte de langoustine et merlu du golfe de Gascogne	4
Figure 3. Protocole du programme OBSMER (SIH, 2017b)	5
Figure 4. Schéma de la structure des données OBSMER (D : Débarquements ; R : Rejets) ; les séquences sont créées artificiellement en suivant la réglementation en vigueur pour la déclaration des débarquements	7
Figure 5. Distribution du nombre d'individus par classe de taille de 1cm de Merlu (<i>Merluccius merluccius</i>) capturés pour la pêche mixte langoustine merlu du golfe de Gascogne pour l'année 2014 (D : Débarquements ; R : Rejets). La ligne verticale noire correspond à la Taille Minimum de Débarquement (TMD). Les intervalles correspondent à la gamme de taille sur laquelle la proportion est calculée (données issues de OBSMER, 2014).....	9
Figure 6. Schéma des différents cas pour le passage de l'échelle de l'opération de pêche à la séquence de pêche	11
Figure 7. Graphique de la variabilité des moyennes des fractions de rejetées (à gauche). Graphique du nombre de données disponibles (à droite) en fonction de différentes combinaisons de variables pour le Merlu (<i>Merluccius merluccius</i>).....	12
Figure 8. Schéma de la démarche générale des classifications d'espèces	16
Figure 9. Graphique des individus de la Classification Ascendante Hiérarchique obtenue après ACP. Le niveau de séparation permet d'obtenir 3 clusters (à gauche). Graphique des individus sur les dimensions 1 et 2, avec les 3 clusters issus de la CAH (à droite).....	17
Figure 10. Graphique représentant les moyennes et les écarts types de chaque variable pour les 3 clusters.	Erreur ! Signet non défini.
Figure 11. Barplot représentant les proportions d'appartenances aux différents clusters des individus statistiques d'une même espèce.....	19
Figure 12. Graphique du nombre d'espèces classées dans la classe « mixte » selon les différents intervalles de proportions d'appartenances.	20
Figure 13. Arbre de classification pour les données avec rejets (à gauche) et les données sans rejets (à droite).....	21
Figure 14. Barplot représentant les proportions d'appartenances aux différents clusters des individus statistiques d'une même espèce.....	21
Figure 15. Contribution de l'espèce dans les captures en fonction des séquences étudiées.	22
Figure 16. Arbre de classification pour les données avec rejets (à gauche) et les données sans rejets (à droite).....	23
Figure 17. Arbre de classification du modèle prédictif dont les données d'entraînement sont 50 % des observations des séquences sans rejet.	23

Liste des tableaux

Tableau 1. Définition des classes d'espèces pour la caractérisation de l'individu statistique (Espèce _ Opération de Pêche) à l'échelle d'un métier	13
Tableau 2. Matrice de confusion de classification des individus statistiques.	20
Tableau 3. Matrice de confusion de classification des individus statistiques.	22
Tableau 4. Matrice de confusion de classification des individus statistiques.	24
Tableau 5. Tableau récapitulatif de la classification des espèces en fonction des données et de la méthode utilisée.....	25
Tableau 6. Matrice de confusion de classification des individus statistiques.	26

Liste des annexes

Annexe I. Tableau récapitulatif des méthodes de classification non supervisée	36
Annexe II. Descriptions des variables quantitatives	41
Annexe III. Résultats complémentaires	46
Annexe IV. Profil des variables pour les espèces sous gestion	65

Introduction

Le golfe de Gascogne (divisions CIEM 27.8.abd) est un système productif où de nombreuses espèces d'intérêt commercial élevé réalisent tout ou partie de leur cycle de vie. Plus de 550 navires de France, d'Espagne et de Belgique opèrent dans cette zone et ciblent principalement les espèces benthiques et démersales telles que la langoustine (*Nephrops norvegicus*), le merlu (*Merluccius merluccius*), la baudroie (*Lophius sp.*) et la sole (*Solea solea*) avec divers chaluts et filets maillants (STECF, 2016). Les chevauchements des distributions de ces espèces dans le temps et dans l'espace font qu'elles sont généralement capturées conjointement par les différentes flottilles qui sont alors considérées comme des pêcheries multi-spécifiques et sont sources d'interactions techniques (Drouineau et al., 2006 ; Prellezo et al., 2016). La partie française de la pêche mixte démersale a généré un rendement brut de 200 millions d'euros en 2009. La pêche mixte de langoustine et merlu est suivie depuis longtemps (Charuau, 1988 ; Gourguet et al., 2012 ; Guillen et al., 2013 ; Macher, 2008) et représente un défi pour la gestion des pêches. En effet il est souvent difficile de limiter la capture d'une espèce sans diminuer la capture des autres et cela peut induire potentiellement un manque à gagner pour les pêcheurs.

La principale mesure de gestion des pêches en Atlantique est aujourd'hui les TAC (Totaux Autorisés de Capture) mono-spécifiques. Ces TACs ont été mis en place pour un certain nombre de stocks reconnus comme commercialement importants et soumis à des pressions de pêche importantes. Ces TACs sont définis en fonction des niveaux d'exploitations estimés et des points de références définis. Ils sont pour le moment définis de manière mono-spécifique sans tenir compte des interactions techniques pouvant exister dans les pêcheries mixtes et ont été jusqu'à présent uniquement limitants en termes de débarquement dans la mesure où les rejets étaient autorisés. Avec la mise en place de l'Obligation de Débarquement par la Politique Commune des Pêches depuis 2015 pour les pélagiques puis progressivement en mer Baltique et sur les autres régions, toutes les captures des espèces sous TAC en Atlantique et ayant une taille minimale de conservation en Méditerranée seront systématiquement débarquées. Les individus ayant une taille inférieure à la taille minimale de débarquement seront débarqués afin d'être valorisés pour autre chose que la consommation humaine (Article 15 du Journal officiel de l'Union européenne N°1380, 2013). Si cette mesure est strictement appliquée cela pourrait conduire à un arrêt de la pêche pour les flottilles capturant une espèce dès lors que son TAC est atteint. Dans certains cas, une espèce bien que accessoire avec un TAC faible pourrait devenir restrictive et conduire à l'arrêt de la pêche pour certaines flottilles très rapidement dans l'année (exemple : le lieu noir (*Pollachius virens*) et le cabillaud (*Gadus morhua*) en Mer du Nord, Mortensen et al., 2018). Dans le but d'éviter de tels blocages sans remettre en question les objectifs de gestion durable il est nécessaire d'identifier pour chaque pêche et chaque flottille ces espèces bloquantes et de comprendre les raisons de ces blocages.

Les captures accessoires (*bycatch en anglais*) sont considérées comme les captures non intentionnelles d'espèces non ciblées (Kumar & Deepthi, 2006). Ces dernières peuvent être non commerciales et/ou interdites, dans ce cas, les individus appartenant à différents taxons (tortues, mammifères, cnidaires, poissons, ...) sont directement rejetés à la mer (vivants ou morts). Si l'espèce présente un intérêt commercial, elle peut être débarquée puis vendue ou rejetée si la taille de capture est inférieure à la taille minimale de débarquement ou si le TAC est atteint ou encore si un rejet sélectif est effectué (seuls les spécimens de bonne qualité sont conservés) (Fernandes et al., 2011, Figure 1).

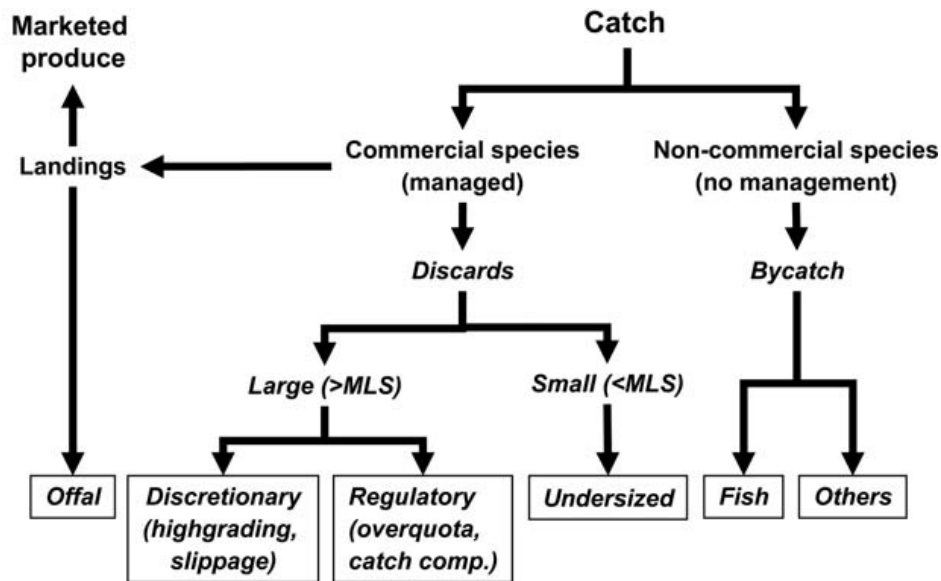


Figure 1. Composition des captures de chalutiers et senneurs démersaux décrivant les différentes catégories (encadrées et en italique) des rejets et les prises accidentelles sont retournées à la mer. MLS (Minimal Landing Size) est la taille minimale de débarquement. Ici est considérée la gestion dans le cadre de la Politique Commune de la Pêche avant la mise en place de l'obligation de débarquement (Fernandes et al., 2011).

Le plan de gestion des eaux occidentales est principalement guidé par la gestion des espèces cibles au travers de règles de fixation des niveaux de TAC prenant en compte les fourchettes de mortalité par pêche autour du point de référence F_{RMD} , (Mortalité par pêche permettant d'assurer un rendement maximum durable) (Rindorf et al., 2017). La possibilité de gérer indirectement les espèces accessoires par l'intermédiaire de l'encadrement opéré sur les espèces ciblées est un enjeu de gestion des prochaines années pour garantir le bon état écologique des écosystèmes marins (Ulrich et al., 2016 ; Garcia et al., 2017). Comme c'est le cas pour le plan de gestion en vigueur en Mer Baltique et en voie d'acceptation en Mer du Nord, les valeurs de mortalité par pêche (F) par espèces pourraient être fixées par groupes d'espèces capturées mêlant des espèces cibles et accessoires (European Commission, 2016). Ainsi, les règles de gestion seront différentes pour les espèces cibles et accessoires. La définition du statut des espèces est donc cruciale.

L'identification d'une espèce cible prend en compte de nombreux paramètres présentant de fortes variations spatio-temporelles. Par exemple, le profil de la pêcherie (mono-spécifique ou mixte) et les forces du marché sont des indicateurs essentiels. Les dynamiques de ventes fluctuent au jour le jour dans une situation de marché volatile, entraînant des rejets instables dans le temps et dans l'espace. Les déchets d'un jour et les prises indésirables peuvent être la cible de demain (Cluclas, 1997).

Identifier et comprendre le statut d'espèces cible ou accessoire passe donc par l'analyse jointe des débarquements et des rejets.

Depuis 2000, en France, un programme d'observations à la mer (OBSMER) a été mis en place pour observer l'ensemble de la capture et estimer les rejets par métier (Cornou et al., 2017). Des observateurs embarquent sur des navires volontaires selon un plan d'échantillonnage par métier standardisé à l'échelle Européenne (SIH, 2017a). Bien que non exhaustives (tous les bateaux et toutes les sorties en mer (marées) ne sont pas observées) ces observations ouvrent la possibilité de comprendre et d'identifier le statut d'espèce cible ou accessoire.

L'objectif de ce travail est de définir une démarche d'analyse et de mettre en œuvre une méthode statistique rigoureuse afin d'identifier les groupes d'espèces cibles et accessoires pour la pêche mixte démersale du golfe de Gascogne. Afin de réaliser ces analyses, deux jeux de données très différents sont disponibles, i) les données OBSMER, issues des campagnes d'observation à la mer qui permettent d'observer la capture totale mais ne représentent qu'un faible pourcentage des marées réalisées et ii) les données SACROIS qui sont exhaustives sur les débarquements mais n'ont pas d'information sur les rejets et agrègent en général plusieurs opérations de pêche qui peuvent être très différentes en terme de composition spécifique de la capture.

Les questions traitées seront donc de deux types : est-il possible de définir les espèces cibles/accessoires en étudiant la capture complète à l'échelle la plus fine de l'opération de pêche (données OBSMER) ? ; Quelles sont les conséquences sur la classification de ces espèces dès lors que l'on ne considère que les débarquements (captures sans les rejets) et que plusieurs opérations de pêche sont regroupées dans une séquence de pêche (données SACROIS) ?

Dans un premier temps, afin de répondre à ces questions, on s'intéressera à clarifier les définitions du statut des espèces et comprendre les différences de catégorisation à partir du jeu de données d'observation de la capture complète par opération de pêche (OBSMER). Dans un deuxième temps, on explorera les capacités à retrouver ces statuts d'espèce en dégradant progressivement les observations OBSMER jusqu'aux données SACROIS et de mesurer la faisabilité de la définition du statut des espèces avec les données de débarquement uniquement (SACROIS).

Ce stage s'inscrit dans le cadre du projet Européen PROBYFISH (Protecting bycaught species in mixed fisheries).

Matériels et méthodes

Cette partie présente le cas d'étude, les données et la méthodologie appliquées afin de catégoriser les espèces.

1. Zone d'étude

Le golfe de Gascogne est un milieu exploité par la pêche avec une grande diversité de navires, d'engins de capture, de métiers et d'espèces commerciales (Macher, 2008). L'étude s'intéresse principalement aux métiers suivants : chaluts jumeaux à panneaux (OTT) et chalut de fond à panneaux (OTB) ciblant les crustacés (OTT_CRU et OTB_CRU) représentant la pêche mixte de langoustines et merlus (Charreau, 1998).

Les captures de langoustines sont essentiellement réalisées sur la Grande Vasière. Cette zone est caractérisée par des fonds sablo-vaseux et par une profondeur qui varie entre 60 et 120 mètres. La zone d'étude est incluse dans les zones de gestion 27.8ab du Conseil International pour l'Exploration de la Mer (Figure 2).

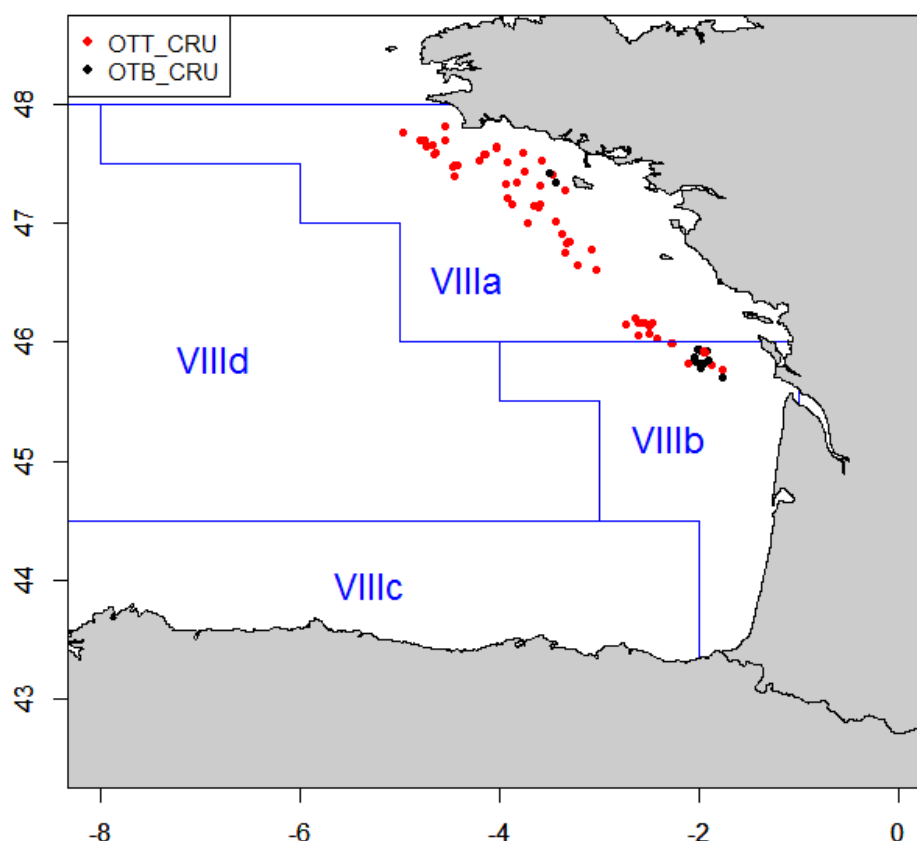


Figure 2. Position géographique des opérations de pêche échantillonnées pour la pêche mixte de langoustine et merlu du golfe de Gascogne

2. Données

2.1. OBSMER

Le programme OBSMER vise à observer *in situ* l'activité de pêche et l'ensemble de la capture, particulièrement la partie non retenue, en identifiant les espèces capturées, les tonnages par

espèce et en effectuant des mesures de tailles. Ce travail s'inscrit dans le cadre d'un partenariat entre le Comité National des Pêches Maritimes et des Elevages Marins (CNPMM), la Direction des Pêches Maritimes et de l'Aquaculture (DPMA), du Ministère de l'agriculture et de l'alimentation, et l'Ifremer. Un échantillon de marées le plus représentatif possible de l'activité de pêche est choisi aléatoirement dans l'ensemble des marées d'un métier de pêche puis les captures de cet échantillon sont observées en détail. Ensuite, une extrapolation des observations de l'échantillon à l'ensemble des navires de pêche est réalisée afin d'obtenir des taux de rejets pour chaque flottille.

En 2016, les données OBSMER ont été utilisées par les groupes d'experts du Conseil International pour l'exploration de la Mer (CIEM) pour l'évaluation quantitative d'une quinzaine de stocks (sole, anchois, chinchard, baudroies, cabillaud, merlan, églefin, plie, merlu, ...) pour ces derniers l'estimation des captures historiques inclut une estimation des rejets des flottilles françaises (Cornou et al., 2017).

Le programme repose sur un protocole pour l'observation en mer permettant de collecter les données (SIH, 2017b, Figure 3). L'observateur doit tout d'abord embarquer sur un bateau avec l'accord du patron de pêche. Au cours d'une marée, une ou plusieurs opérations de pêche ont lieu, l'observateur doit tirer au sort celles qui vont être échantillonnées ou non. Au minimum une opération de pêche sur trois doit être échantillonnée. Dans ce cas, l'observateur identifie, mesure, dénombre et pèse les débarquements, les rejets ainsi que les captures accidentelles (ici, considérées comme les tortues et mammifères marins). A l'inverse, si une opération n'est pas échantillonnée l'observateur identifie et pèse uniquement les débarquements. Toutefois, si des prises accidentelles sont réalisées toutes les mesures (identification, dénombrement, poids et mesure) seront réalisées (Figure 3).



Figure 3. Protocole du programme OBSMER (SIH, 2017b)

Les données utilisées dans cette étude sont stockées dans une base de données selon un certain format (FREE2, Le blond, 2018) qui se structure en plusieurs tables : Marées,

Opération de pêche, Engins, Stratégie, Détail, Captures et Mesures. Après la jointure de ces tables, 87 variables sont renseignées. Un filtre est appliqué sur l'année étudiée (2014) ainsi que la zone de l'étude (Golfe de Gascogne). Seules les données validées et avec un bon déroulement de l'opération de pêche sont conservées. De plus, les animaux inertes et végétaux ne sont pas pris en compte. Au total pour l'ensemble des métiers étudiés en 2014, 30 marées dont 68 opérations de pêche échantillonnées sont renseignées.

L'inconvénient de ces données est qu'elles ne représentent qu'un faible pourcentage des marées totales (<1-3% en fonction des métiers) et la représentativité des bateaux observés repose sur le volontariat des patrons de pêche.

2.2. SACROIS

Afin de disposer d'une information fiable pour le suivi de l'activité de pêche et pour les travaux de recherche scientifique, l'application SACROIS a pour objet de fournir un ensemble de données de production et d'effort de pêche élaboré, validé et consolidé, sur la base du croisement de l'ensemble des sources d'informations disponibles (Hervin & Begot, 2013).

Pour les données SACROIS, les débarquements sont issus d'un algorithme de croisement entre des logbooks renseignés par les pêcheurs (de manière réglementaire à l'échelle de la séquence telle que définie plus bas), des données VMS (Vessel Monitoring System) et des données de ventes en criée. Ces données décrivent exhaustivement la partie débarquée des captures des flottilles françaises. Contrairement au programme OBSMER, les données SACROIS ne sont pas renseignées à l'échelle de l'opération de pêche mais à l'échelle de la séquence de pêche (Figure 4). Une séquence de pêche au sein d'une marée est définie par un jour, un secteur de pêche (rectangle statistique CIEM) et un engin. Tout changement de l'un de ses paramètres dans une marée donne lieu à une nouvelle séquence de pêche (Hervin & Begot, 2013). Le nombre d'observations des données issues de l'algorithme de croisement est important. En effet, pour l'année 2014, pour l'ensemble des métiers étudiés, 14 974 marées constituant 28 247 séquences de pêche ont été créées avec 55 variables renseignées.

3. Méthodes

La démarche consiste à proposer une méthode statistique basée sur l'analyse des captures d'une espèce pour déterminer si cette espèce est cible ou accessoire. Les espèces accessoires ont été classées en 3 catégories : « espèces accessoires valorisées », « espèces accessoires mixtes » et « espèces accessoires non valorisée ». Ces catégories ayant été choisies à partir de la description du programme PROBYFISH (Protecting bycaught species in mixed fisheries). Idéalement, il faudrait connaître la capture totale de cette espèce à l'échelle de l'opération de pêche et sa partie commercialisée et non commercialisée. Seules les données OBSMER permettraient donc cette analyse. Cependant, le jeu de données OBSMER ne représente qu'un faible échantillon des marées réalisées par les différents navires de pêche (entre 1 et 3%). Même si le protocole d'échantillonnage est adapté à l'étude des fractions débarquées et rejetées des principaux métiers et des espèces commerciales importantes, il ne permet pas d'avoir une vision représentative de l'ensemble des espèces capturées au cours de l'année et sur l'ensemble de la zone.

Les données SACROIS quant à elles, se veulent exhaustives mais présentent l'inconvénient de ne renseigner que les débarquements.

Nous proposons donc de raisonner de la façon suivante : apprendre avec OBSMER ce que les rejets apportent (et donc ce que l'on perd sans les rejets) à la définition du statut des espèces pour les espèces, les mois et les zones observés. Puis « dégrader » progressivement les données OBSMER en retirant la fraction rejetée puis en agrégeant les opérations de pêche en séquences de pêche afin de se rapprocher des déclarations de pêche réglementaires. Cette démarche permettra de comprendre les forces et les faiblesses du jeu de données SACROIS à la définition du statut des espèces. L'idée sous-jacente à cette démarche est d'exploiter le jeu de données OBSMER pour apprendre sur le statut des espèces et d'utiliser cette connaissance pour classer plus globalement, toutes les espèces débarquées, tous les mois et sur toute la zone du jeu de données SACROIS.

Dans cette étude nous nous sommes concentrés sur l'année 2014 et le métier crustacé au chalut de fond jumeau (OTT-CRU) et au chalut de fond simple (OTB-CRU).

3.1. Différentes échelles

Il est important de noter qu'il existe plusieurs échelles et types d'observations de la capture pour une marée. Une marée est composée de manière réglementaire de séquences à l'échelle de laquelle les pêcheurs doivent déclarer leurs débarquements (unité de temps (jour), espace (rectangle statistique CIEM) et engin) : Les captures des opérations de pêche réalisées le même jour dans le même rectangle statistique en utilisant le même engin de pêche sont agrégées et c'est à cette échelle que les données SACROIS sont renseignées. Dans le cadre du programme OBSMER, toutes les opérations de pêche ne sont pas échantillonnées (Figure 4). La dégradation des données OBSMER pour retrouver les échelles spatio-temporelles de SACROIS, sera réalisée en sommant les débarquements des opérations de pêche à l'échelle de la séquence de pêche.

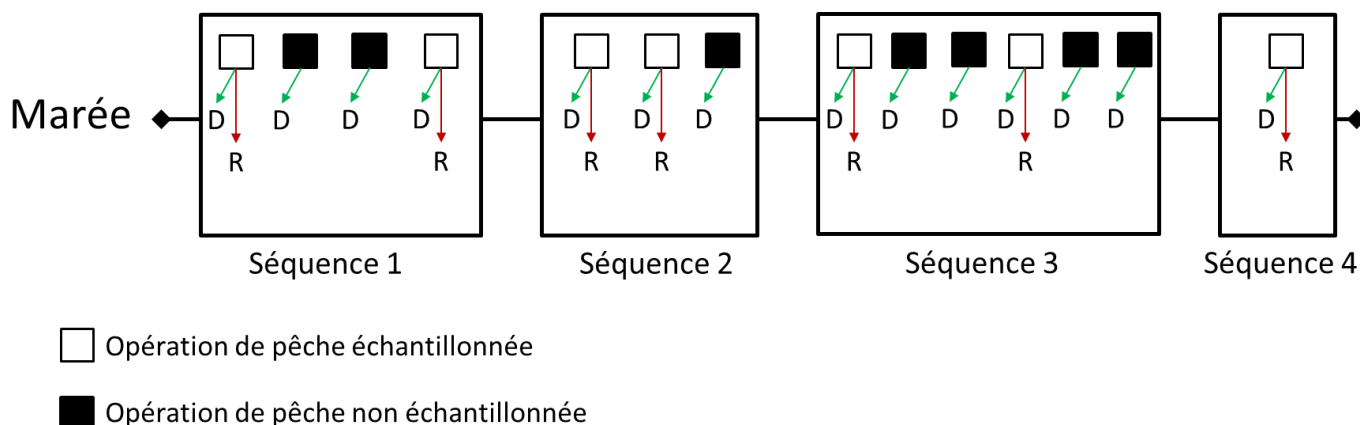


Figure 4. Schéma de la structure des données OBSMER (D : Débarquements ; R : Rejets) ; les séquences sont créées artificiellement en suivant la réglementation en vigueur pour la déclaration des débarquements

L'individu statistique de cette étude est une espèce capturée à l'échelle étudiée (opération de pêche ou séquence de pêche). Autrement dit, une espèce dans une opération de pêche constituera l'individu statistique. Nous présentons dans cette partie les variables permettant de caractériser un individu statistique selon l'échelle.

3.1.1. Opérations de pêche

Différentes notations ont été définies, i pour espèce, j pour opération de pêche, m pour marée, l pour séquence et k pour mois.

I. Contribution de l'espèce dans les captures : $c_{i,j}$

La contribution de l'espèce (i) dans les captures dans l'opération de pêche (j) est classiquement utilisée, notamment pour analyser les rejets (Fernandes et al., 2011). Cette variable représente la part du poids de l'espèce par rapport au poids total des captures. Autrement dit, le rapport entre le poids des captures pour l'espèce et le poids des captures totales pour une opération de pêche (Annexe II, équation 1). Intuitivement, les espèces représentant une part importante de la capture devraient être classées comme cibles et inversement, les espèces minoritaires plutôt comme accessoires.

II. Fraction rejetée : $RC_{i,j}$

La fraction rejetée est une variable également très utilisée lors des études sur les rejets. Les rapports OBSMER les utilisent afin de comparer la fraction rejetée entre les différents métiers, engins, flottilles et pêcheries (Cornou et al., 2017). Le rapport entre le poids rejeté et le poids de la capture est calculé (Annexe II, équation 2). Une forte fraction rejetée peut s'expliquer de deux manières ; i) l'espèce n'est pas désirée (accessoire) et seuls quelques individus sont gardés pour être débarqués, ii) l'espèce est sous mesure de gestion incluant des tailles minimales de conservation (TMC) et l'opération de pêche a conduit à une forte capture d'individus sous la TMC.

Dans la démarche d'analyse des rejets incluant la mise en place de l'obligation de débarquement, la création de variables tenant compte de la gestion et des tailles minimales de débarquement (TMD) (ou encore appelées tailles minimales de conservations) est nécessaire (Catchpole et al., 2017). Ces dernières sont extraites des différents arrêtés (JORF, 2013 ; Journal officiel de l'Union européenne N°120, 2018), les deux prochains indicateurs tiennent compte de la TMD.

III. Fraction rejetée supérieure ou égale à la Taille Minimale de Débarquement : $RCT_{i,j}$

La fraction rejetée supérieure ou égale à la taille minimale de débarquement (TMD) correspond au rapport entre le poids des rejets supérieurs à la TMD et le poids des captures supérieures à la TMD. Cela équivaut à calculer une proportion de rejet pour les captures supérieures ou égales à la TMD (Annexe II, équation 3).

L'objectif est de pouvoir identifier les espèces qui ont une fraction de rejet élevée due à leur faible taille. Une proportion de rejet des captures supérieure ou égale à la TMD permet de différencier une espèce qui a une fraction rejetée (RC) forte et une fraction supérieure ou égale à la TMD (RCT) forte (espèce accessoire quelle que soit la taille), d'une espèce présentant une fraction rejetée (RC) forte et une fraction supérieure ou égale à la TMD (RCT) faible. Cet indicateur peut mettre en évidence certains rejets sélectifs ou des dépassements de quotas. Ces phénomènes sont importants car ils peuvent biaiser l'interprétation du statut d'espèce. Par exemple, si au cours de l'année, une espèce est débarquée puis fortement rejetée elle sera catégorisée comme accessoire à l'échelle annuelle alors que celle-ci peut être une espèce cible dont le quota est dépassé.

IV. Fraction de rejet supérieure ou égale à la taille minimale de débarquement : $RRT_{i,j}$

La fraction de rejet supérieure ou égale à la taille minimale de débarquement est représentée par le rapport entre le poids rejeté des individus ayant une taille supérieure ou égale à la taille minimum de débarquement (TMD) et le poids rejeté. Cela revient à calculer la proportion de rejet supérieure ou égale à la TMD (Annexe II, équation 4).

Par rapport à la variable précédente, celle-ci renseigne uniquement sur les rejets et non les captures, par exemple si une majorité d'individus supérieurs ou égaux à la TMD sont rejetés alors les valeurs des deux indicateurs seront différentes. Si le ratio est proche de 1 l'espèce sera qualifiée comme accessoire non valorisée. Ainsi, le ratio utilisé ici est complémentaire du précédent.

La Figure 5, met en évidence les différences entre les 3 variables prenant en compte les rejets. La variable RC est calculée pour les captures sur l'ensemble du spectre de taille, contrairement à la variable RCT qui est calculée pour les individus ayant une taille supérieure ou égale à la TMD. Enfin, la dernière variable est représentée à l'échelle des rejets tout en tenant compte de la TMD. Pour les variables tenant compte de la TMD (RCT et RRT) si l'espèce ne présente pas de TMD dans la réglementation, le résultat de RC est alors attribué.

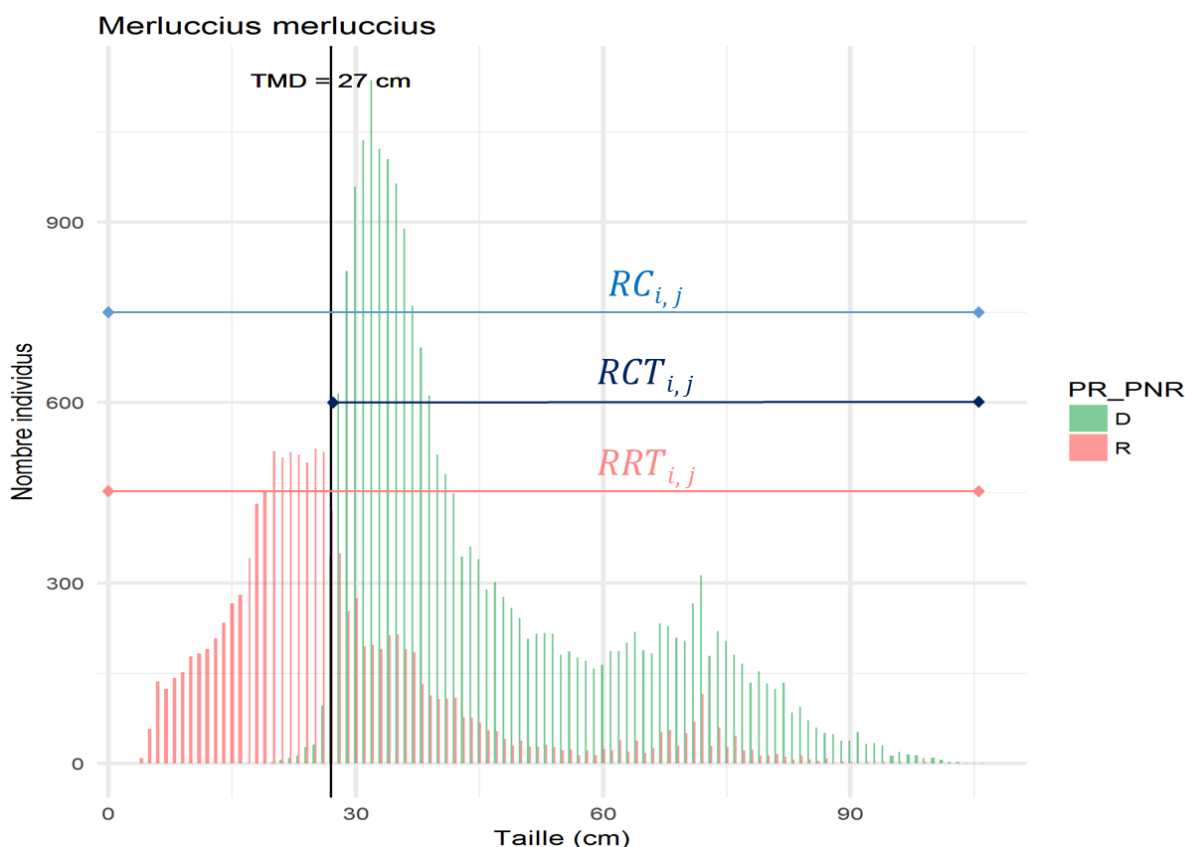


Figure 5. Distribution du nombre d'individus par classe de taille de 1cm de Merlu (*Merluccius merluccius*) capturés pour la pêche mixte langoustine merlu du golfe de Gascogne pour l'année 2014 (D : Débarquements ; R : Rejets). La ligne verticale noire correspond à la Taille Minimum de Débarquement (TMD). Les intervalles correspondent à la gamme de taille sur laquelle la proportion est calculée (données issues de OBSMER, 2014).

Les boxplots des différents indicateurs de rejets calculés à l'échelle de l'opération de pêche sont représentés (Annexe III.a).

V. Valeur : $V_{i,j,k}$

La valeur (en euros) de l'espèce peut fortement influencer sur son statut (Davies et al., 2011 ; Catchpole et al., 2017 ; Fernandes et al., 2011). Cette variable représente la valeur du poids débarqué de l'espèce. Autrement dit, le poids débarqué de l'espèce multiplié par son prix/kg/mois (issue de SACROIS avec une sélection sur la zone du golfe du Gascogne) (Annexe II, équation 5).

Cette variable, de par sa construction, va être corrélée avec le poids, les espèces représentant un fort revenu vont être considérées comme cible alors que des espèces avec un revenu faible ou nul vont être considérées comme accessoires (et si nul, non valorisable).

VI. Fréquence d'occurrence des espèces dans les opérations de pêche échantillonnées : $ne_{m,i}$

La fréquence d'occurrence des espèces dans les opérations de pêche échantillonnées représente la fréquence à laquelle une espèce est capturée dans une marée échantillonnée (Annexe II, équation 6).

En lien avec les autres variables si une espèce est fréquente, représente une forte contribution dans les captures et présente un fort revenu elle sera classée cible. Si l'inverse est observé, l'espèce sera catégorisée comme accessoire.

VII. Fréquence d'occurrence des espèces débarquées dans l'ensemble des opérations de pêche : $n_{m,i}$

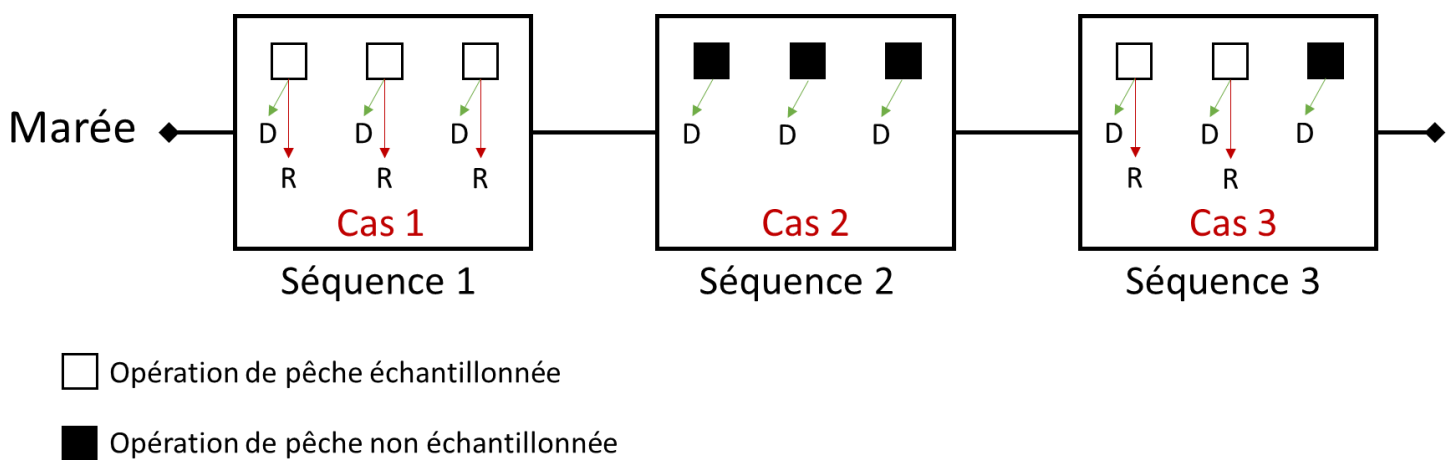
En comparaison avec la variable précédente, la fréquence est calculée à partir des débarquements. Cela permet de prendre en compte les opérations de pêche échantillonnées et non échantillonnées. Ainsi, l'ensemble des opérations de pêche sont prises en compte dans la marée afin de calculer la variable (Annexe II, équation 7).

Cet indicateur est en lien avec les autres variables : si une espèce est fréquente, représente une forte contribution dans les captures et présente un fort revenu alors, elle sera classée comme cible. Si l'inverse est observé, l'espèce sera catégorisée comme accessoire. La différence avec la variable précédente est le nombre d'opérations de pêche pris en compte. Si les deux fréquences sont très différentes et que l'espèce est débarquée il faut privilégier celle-ci.

3.1.2. Séquences de pêche

Le passage à l'échelle de la séquence suite aux analyses sur les opérations de pêche échantillonnées nécessite une agrégation des poids et une élévation des rejets (si on souhaite les conserver). Ce changement d'échelle conduit à créer un identifiant séquence pour les données OBSMER. Cet identifiant est construit en agrégeant un jour, un secteur de pêche et un engin au sein d'une même marée. Les débarquements (disponibles pour chaque opération de pêche) sont toujours sommés à cette échelle. Pour connaître les rejets (uniquement disponibles pour les opérations de pêche échantillonnées) à l'échelle de la séquence, il est nécessaire de proposer une stratégie d'estimation. Différents cas de figure sont possibles. Le premier cas est une séquence constituée uniquement d'opérations de pêche échantillonnées,

les poids débarqués et rejetés sont renseignés. Ces poids sont donc agrégés afin de calculer les variables présentées précédemment (Figure 6, Cas 1). Un cas opposé peut avoir lieu, en effet, seules des opérations de pêche non échantillonnées sont présentes dans la séquence. Les seules informations disponibles sont les poids débarqués. Dans le but d'estimer des poids rejetés à l'échelle de la séquence, la moyenne de la fraction rejetée (RC) par espèce de la marée est attribuée (Figure 6, Cas 2). La même procédure est réalisée pour les autres variables de rejets prenant en compte (RCT et RRT). Enfin, le dernier cas est une séquence constituée d'opérations de pêche échantillonnées et non échantillonnées. Dans ce cas l'approche est analogue au cas précédent excepté que ce sont les moyennes de la fraction par espèces des opérations de pêche échantillonnées au sein de la séquence qui sont utilisées (Figure 6, Cas 3).



*Figure 6. Schéma des différents cas pour le passage de l'échelle de l'opération de pêche à la séquence de pêche
(D = poids des Débarquements ; R= poids des Rejets)*

Pour les cas 2 et 3, il est possible que certaines espèces soient uniquement présentes dans les opérations de pêche non échantillonnées. Cela se produit dans de rares cas (< 5%). Pour le cas 3, une fraction rejetée moyenne est recherchée à l'échelle de la marée. Si aucune valeur n'est trouvée alors un processus est mis en place afin de trouver une valeur la plus précise possible. Pour cela les fractions rejetées par espèce sont agrégées en considérant une succession de combinaisons particulières des différentes variables disponibles. L'identifiant du navire est pris en compte il est le proxy de l'effet du patron sur les rejets, l'engin est également utilisé, ainsi que des indicateurs spatio-temporels (rectangle, zone, mois). Les représentations des variabilités des moyennes des fractions de rejets permettent de définir les combinaisons avec lesquelles les données sont disponibles (et donc plus précises). L'analyse de ces graphiques a permis de définir une succession de combinaisons de variables à agréger afin d'attribuer une fraction rejetée. Dans la pratique, toutes les combinaisons de variables sont calculées, classées dans un ordre décroissant de précision. Dès qu'une valeur est disponible elle est attribuée sinon la recherche continue vers les autres combinaisons possibles (Figure 7). Pour le cas 2, il est important de noter que si une espèce n'est présente que dans des opérations de pêche non échantillonnées elle est retirée de l'analyse.

Cette dernière étape mène à une forte hypothèse pour cette méthode : les espèces systématiquement rejetées sont sous-estimées. En effet, ces dernières sont prises en compte uniquement pour les opérations de pêche échantillonnées.

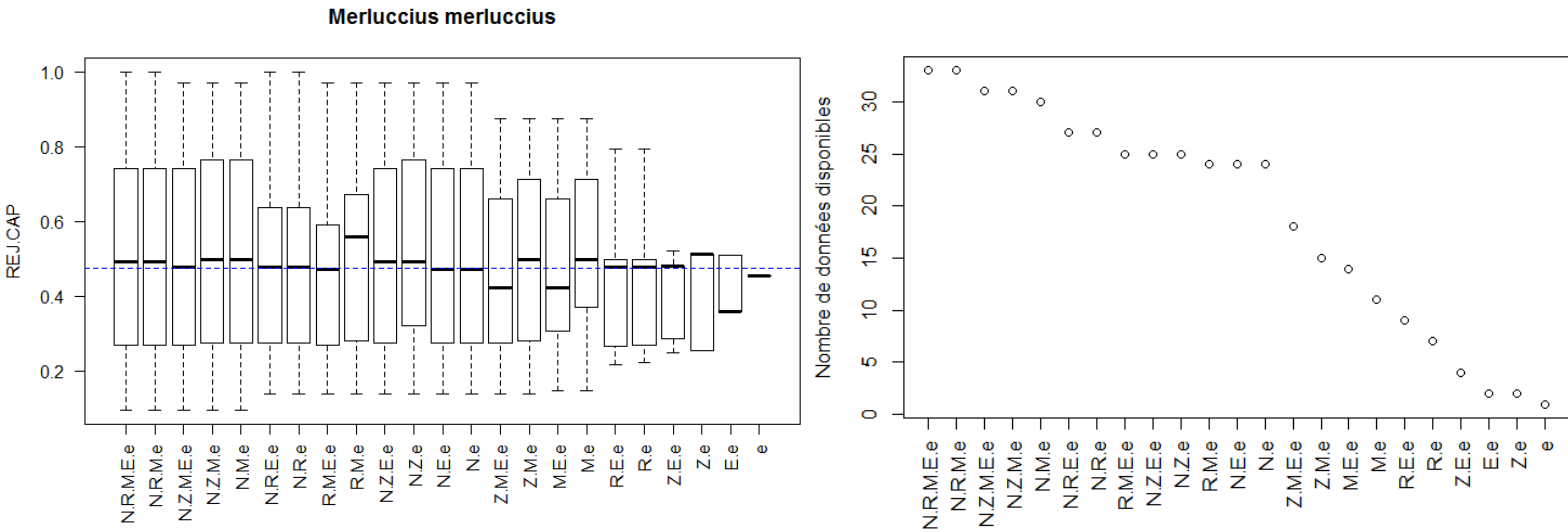


Figure 7. Graphique de la variabilité des moyennes des fractions de rejetées (à gauche). Graphique du nombre de données disponibles (à droite) en fonction de différentes combinaisons de variables pour le Merlu (*Merluccius merluccius*).

Le trait pointillé bleu représente la moyenne de la fraction rejetée pour l'ensemble des combinaisons de variable.
 N=Navire, R=Rectangle, Z=Zone, M=Mois, E=Engin, e=espèce
 Les variables spatiales : rectangle et zone se recouvrent amenant à les séparer.

Une séquence peut être constituée de plusieurs opérations de pêche pratiquant différents métiers, exemple : OTT_CRU (chalut de fond jumeau ciblant les crustacées) et OTT_DEF (chalut de fond jumeau ciblant les espèces démersales). Environ 20% des séquences présentent cette configuration. Dans l'objectif de rapprocher les séquences OBSMER de celles de SACROIS, il est nécessaire d'attribuer un seul métier par séquence. Dans l'algorithme SACROIS, l'espèce majoritaire en valeur associée à l'engin de pêche est choisie afin d'attribuer un métier (Hervin & Begot, 2013). Nous utiliserons la même méthode pour définir le métier de la séquence OBSMER.

Ainsi, une fraction rejetée estimée avec le poids débarqué est obtenue permettant de calculer un poids rejeté estimé avec la formule ci-dessous :

$$\widehat{R}_{i,l} = \text{poids rejeté estimé (Espèce}_i, \text{Séquence de pêche}_l)$$

$$D_{i,l} = \text{poids débarqué agrégé (Espèce}_i, \text{Séquence de pêche}_l)$$

$$\widehat{RC}_{i,l} = \text{fraction rejetée estimée (Espèce}_i, \text{Séquence de pêche}_l)$$

$$\widehat{R}_{i,l} = \frac{D_{i,l} * \widehat{RC}_{i,l}}{1 - \widehat{RC}_{i,l}}$$

Ainsi, l'ensemble des autres variables, analogue à celles présentées précédemment, peuvent être calculée (Annexe II, équation 8 à 10)

3.2. Définition *a priori* du statut des individus statistiques

L'ensemble des variables doit *a priori* permettre de catégoriser les espèces. Dans cette étude, 4 principales classes d'espèces sont étudiées. Les espèces cibles correspondent à celles que les pêcheurs souhaitent capturer. Cette première classe va être définie par des proportions importantes dans les captures, ainsi que des valeurs et fréquences élevées. Une seconde classe représente les espèces valorisées, ce sont des individus qui ne sont pas ciblés mais sont débarqués et vendus. La classe des espèces valorisées est caractérisée par de faibles taux de rejets ainsi que des valeurs élevées. A l'inverse, des valeurs faibles ou nulles associées à de forts taux de rejets constituent la classe des espèces non valorisées. Enfin, la classe mixte peut être définie par une variabilité de tous les indicateurs (Tableau 1). La définition de cette dernière classe peut changer en fonction de l'échelle considérée.

Tableau 1. Définition des classes d'espèces pour la caractérisation de l'individu statistique (Espèce _ Opération de Pêche) à l'échelle d'un métier

Variables	Espèces accessoires			
	Cible	Valorisée	Non Valorisée	Mixte
$C_{i,j} \left(\frac{\text{Poids des captures de l'espèce}}{\text{Poids total des captures de l'OP}} \right)$	Important	Variable	Variable	Variable
$RC_{i,j} \left(\frac{\text{Rejets}}{\text{Captures}} \right)$	Variable	~ 0	~ 1	Variable
$RCT_{i,j} \left(\frac{\text{Rejets} \geq TMD}{\text{Captures}} \right)$	Variable	~ 0	~ 1	Variable
$RRT_{i,j} \left(\frac{\text{Rejets} \geq TMD}{\text{Rejets}} \right)$	Variable	~ 0	~ 1	Variable
$V_{i,j,k}$ (Valeur)	++	++	~ 0	Variable
$n_{m,i}$ ou $ne_{m,i}$ (Fréquence d'occurrences)	~ 1	Variable	Variable	Variable

Des variables qualitatives sont également disponibles et pourraient s'avérer utiles pour caractériser le statut des captures puis des espèces. Par exemple, le port est considéré car il est possible que les stratégies de pêche diffèrent en fonction des ports. L'identifiant du navire permet de prendre en compte « un effet bateau » (et indirectement l'effet « pêcheur »). Des classes de longueur de bateau peuvent souligner des comportements différents entre les grandes et petites unités. L'engin et le métier permettent de mettre en évidence des captures

différentes en fonction du type de pêche réalisée. Enfin, le maillage, l'utilisation ou non de dispositif de sélectivité et/ou répulsif peuvent mettre en évidence une sélectivité engendrant une composition spécifique des captures différente.

3.3. Caractérisation du statut des espèces

La démarche de cette étude consiste dans un premier temps à identifier les espèces cibles et accessoires à partir des données de captures (OBSMER) à l'échelle de l'opération de pêche. L'étape suivante consistera à réaliser la même analyse en passant à l'échelle de la séquence de pêche. Puis, nous appliqueront cette méthode en abandonnant les variables liées aux rejets.

3.3.1. Méthodes d'analyse avec les rejets

Une Analyse en Composante Principale a été réalisée. Cette analyse multivariée réduit la dimensionnalité de l'ensemble de données et identifie l'existence ou non de liens entre les variables ainsi que leur pertinence (Davie & Lordan, 2011). Les composantes sélectionnées seront les axes accumulant 70% des inerties (Deporte et al., 2012). Ces dernières sont présentées par ordre d'importance, avec la plus grande variation décrite par la première composante (Fowler et al., 1998). Pour cela, certaines variables pourront être centrées réduites. En effet, la majorité des variables sont comprises entre 0 et 1 mais les caractéristiques d'activité de pêche (durée d'opérations de pêche, vitesse de l'engin de pêche et profondeur) ainsi que la valeur des débarquements présentent des valeurs et variabilités importantes et très différentes justifiant leur transformation (Annexe III.b). En s'inspirant des différentes méthodes utilisées dans le but de catégoriser les métiers, une Analyse en Composante Principale (ACP) combinée à une Classification Ascendante Hiérarchique (CAH) a été réalisée (Davie & Lordan, 2011 ; Deporte et al., 2012 ; Marchal, P., 2008 ; Pelletier & Ferraris, 2000). La méthode de Classification Ascendante Hiérarchique (CAH) sur les résultats de l'analyse précédente a été choisie pour regrouper les individus. Elle est très visuelle et classe chaque individu dans un cluster. La méthode CAH attribue initialement chaque individu statistique à son propre cluster, alors l'algorithme procède itérativement, à chaque étape rejoignant les deux groupes les plus similaires, jusqu'à ce qu'il n'y ait qu'un seul groupe (Deporte et al., 2012). La similarité est calculée en utilisant une fonction de distance (par défaut, la distance euclidienne) et une méthode de clustering (par défaut, le critère de Ward) (Ward, 1963). Le nombre de clusters est déterminé par la suite, une fois que toutes les combinaisons ont été calculées, en utilisant un Scree plot (graphique d'inertie). Le package FactoMineR est utilisé pour réaliser l'ACP et la CAH (Lê et al., 2008). Les résultats de la CAH sont ensuite confrontés aux définitions *a priori* des classes d'espèces afin de définir les clusters obtenus.

Une comparaison des moyennes et écarts types des variables de chaque cluster permettent l'analyse des différents groupes issus de la CAH. De plus, l'importance des variables au sein de chaque cluster peut être décrite en utilisant la valeur-*t* (*v*-test). Plus *v*-test est grande (en valeur absolue) et plus la probabilité est petite, plus la variable caractérise la classe. Une valeur test négative indique que la moyenne de cette variable dans la classe est plus faible que la moyenne dans tout l'échantillon. Dans les graphiques de résultats, nous ne présenterons que les modalités significativement importantes pour la $|v.test| < 1.96$ (Lemerrier, et al., 2010).

3.3.2. Détermination des seuils

Une fois ces individus statistiques catégorisés dans des clusters bien définis (« cible », « valorisée », « non valorisée »), un rassemblement au niveau de l'espèce est nécessaire dans la mesure où le but de l'analyse est de décrire les espèces. Une proportion d'appartenance à chaque cluster par espèce est ainsi calculée. Ce résultat permet de déterminer une catégorisation à l'échelle de l'espèce. Plusieurs cas sont possibles, tout d'abord, si tous les individus statistiques d'une même espèce appartiennent à la même classe alors l'espèce est catégorisée dans le cluster correspondant. Pour une même espèce ayant des individus statistiques classés dans plusieurs clusters il est nécessaire de définir des seuils. Arbitrairement, nous avons considéré qu'une espèce ayant les individus statistiques appartenant pour une proportion égale à deux classes (50% pour les deux classes) est définie comme mixte. L'absence de seuils dans la littérature oblige à développer une méthode afin de fixer des limites. Une analyse des changements de classe en créant des intervalles des proportions à partir de 50 % avec des pas de 5% peut aider à fixer un seuil pour aboutir à quatre classes : « cible », « valorisée », « non valorisée » ou « mixte ». Les différents seuils possibles seront conservés afin de pouvoir comparer les résultats des différentes échelles étudiées.

3.3.3. Méthodes d'analyse sans les rejets

Dans le but de comprendre les changements suite à la perte d'information des rejets et dans la perspective de travailler sur les données SACROIS, la classification a été mise en place en n'utilisant que les profils de débarquements, valeurs, ainsi que les fréquences d'occurrences de chaque individu statistique à l'aide d'une classification des k-means (Macquenn, 1967). La classification par la méthode des nuées dynamiques (classification k-means) est une méthode itérative qui, quel que soit son point de départ converge vers une solution. La solution obtenue n'est pas nécessairement la même quel que soit le point de départ. En effet, un objet peut être affecté à une classe au cours d'une itération puis changer de classe à l'itération suivante. Pour la première itération un point de départ est choisi le principe est d'associer le centre des k classes à k objets (pris au hasard ou non). Ensuite la distance entre les objets et les k centres est calculée et les objets sont affectés aux centres dont ils sont les plus proches. Puis les centres sont redéfinis à partir des objets qui ont été affectés aux différentes classes. Par la suite, les objets sont réaffectés en fonction de leur distance aux nouveaux centres. Et ainsi de suite jusqu'à ce que la convergence soit atteinte (Macquenn, 1967). Pour se rapprocher d'une stratégie basée sur de l'apprentissage, les centroïdes des classes ont été initialisés en utilisant les valeurs moyennes pour chaque classe préalablement définie avec la classification exploitant les données avec rejets. Le but étant une comparaison avec la démarche prenant en compte les rejets, le nombre de classes a été fixée à 3 (« espèce cible », « espèce accessoire valorisée » et « espèce accessoire non valorisée »). Le nombre de classes étant similaire entre la CAH et les k-means, une matrice de confusion entre les catégorisations des individus statistiques issues des deux méthodes mettra en évidence les différences de classification.

3.3.4. Résumé de la démarche

Les différentes classifications issues de diverses méthodes pourront être comparées (Figure 8). De plus, afin de se rapprocher de la catégorisation sur le jeu de donnée final, une classification sera réalisée à partir d'un modèle d'arbre de décision avec une sélection aléatoire de 50% des données, utilisées comme entraînement. Cette approche sur OBSMER permet de comprendre comment la classification évolue si l'on possède qu'une portion des données.

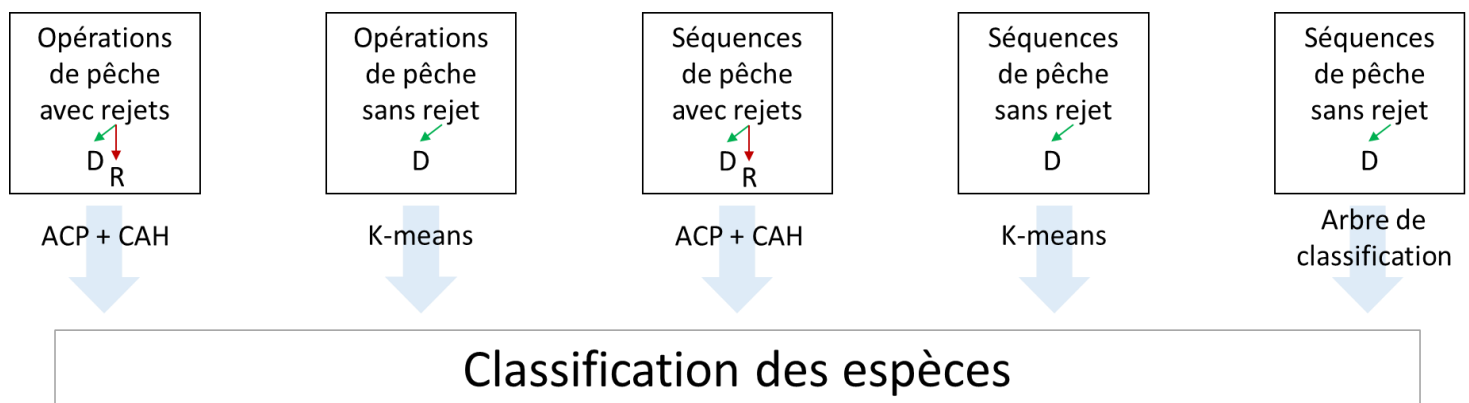


Figure 8. Schéma de la démarche générale des classifications d'espèces

3.3.5. Méthodes d'apprentissage sur SACROIS

L'objectif principal est d'utiliser l'ensemble des données OBSMER comme données d'entraînement afin de classer les espèces à partir de SACROIS. Une analyse de la faisabilité de cette démarche ainsi que la sélection d'une méthode appropriée a été réalisée. Pour cela, différentes approches de classifications supervisées (arbre de classification, forêt aléatoire) ont été explorées (Hastie et al., 2008 ; Venables & Ripley, 2002). Seuls les arbres de classification ont été retenus car les forêts aléatoires sont pertinentes pour sélectionner des variables explicatrices et que le jeu de données dans sa forme actuelle ne permet pas d'envisager un grand nombre de variables explicatrices. Les arbres de classification peuvent être utilisés notamment en fouille de données ou en apprentissage automatique. Le principe est de catégoriser une variable à expliquer qualitative à partir des variables explicatives quantitatives. La construction d'un arbre de discrimination consiste à déterminer une séquence de nœuds, l'algorithme CART (Classification And Regression Trees) est utilisé (Breiman et al., 1984). Dans ce dernier, l'indice de diversité de Gini permet de caractériser l'homogénéité des sous-ensembles. Le package rpart (Therneau & Atkinson, 2018) permet de mettre en œuvre cette méthode avec le logiciel R.

Les arbres de classification ont été utilisés à plusieurs reprises dans cette étude. Premièrement, en fouille de données sur les différentes classifications (jeu de données OBSMER à l'échelle de l'opération de pêche et à l'échelle de la séquence de pêche) afin de comprendre comment les variables quantitatives définissent les classes d'espèces. Deuxièmement, une classification supervisée a été mise en œuvre dans le but catégoriser les espèces à partir de SACROIS. La liaison entre les séquences OBSMER et SACROIS est nécessaire afin de définir les données d'entraînement. Deux classifications supervisées seront testées avec des données d'entraînement différentes. Premièrement, les données d'entraînement seront les résultats de la classification à l'échelle de la séquence prenant en compte les rejets. Dans une seconde analyse, les résultats de la classification à l'échelle de la séquence traitant uniquement des débarquements serviront comme données d'entraînement. Dans les deux cas, seulement les variables quantitatives disponibles sous SACROIS ont été sélectionnées (contribution de l'espèce dans les débarquements, fréquence et valeur). Enfin, une comparaison des résultats des analyses avec les différentes données d'entraînements sera effectuée à l'aide d'une matrice de confusion. L'influence des données d'entraînement sur la classification supervisée pourra donc être étudiée.

Résultats

1. Classification à l'échelle de l'opération de pêche

1.1. Analyse de la classification des données de captures

La variabilité des différentes variables est représentée avant de réaliser une Analyse en Composante Principale (ACP) (Annexe III.b). Les graphiques des résultats de l'ACP montrent que les 3 variables de rejets sont fortement corrélées à l'axe 1. Les variables de fréquence, contribution de l'espèce dans les captures et valeur sont elles corrélées à l'axe 2. Les autres variables liées à « l'activité de pêche » sont le mieux représentées sur l'axe 3, la profondeur est anti corrélée à la vitesse de l'engin et la durée de l'opération de pêche (Annexe III.c). A l'issue des résultats de l'ACP, une Classification Ascendante Hiérarchique (CAH) est réalisée. Le graphique de la perte d'inertie lors du changement de classe renseigne sur le nombre de classes optimal. La séparation en trois classes apparaît comme la plus judicieuse (Figure 9, à droite). Une séparation assez claire apparaît sur le graphique des individus entre l'axe 1 et 2 de l'ACP (Figure 9, à gauche).

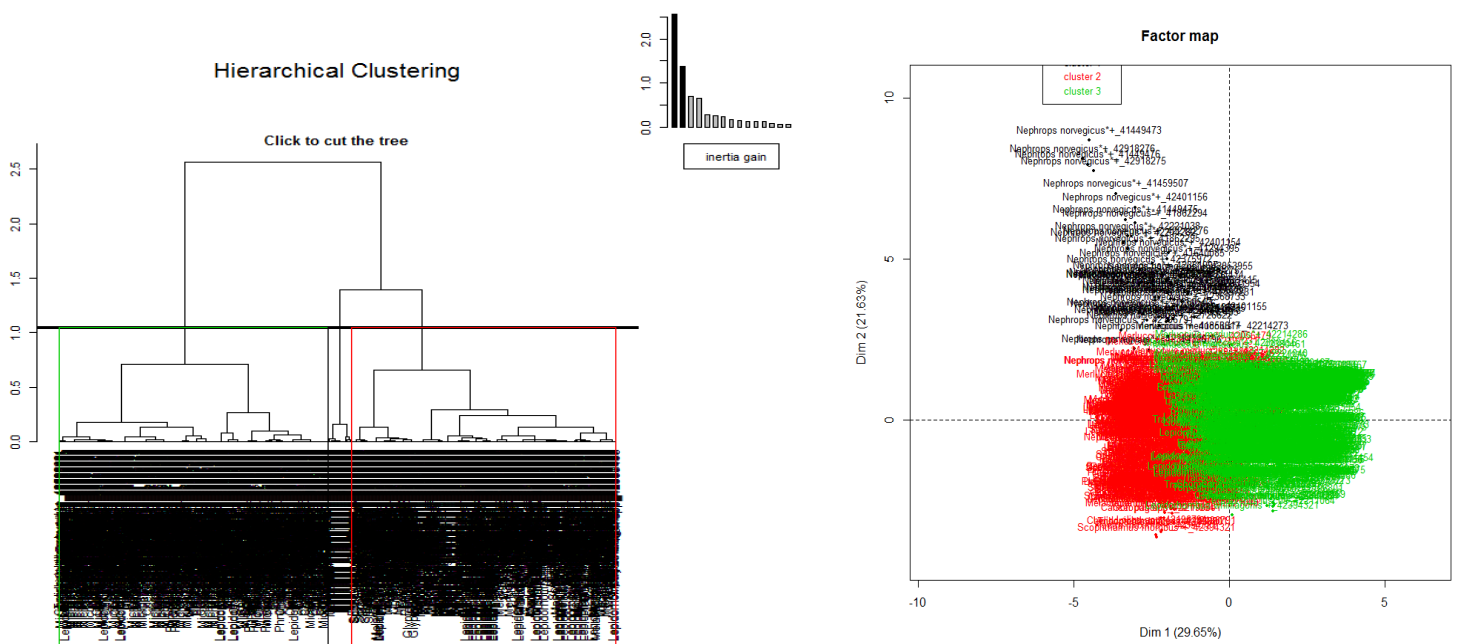


Figure 9. Graphique des individus de la Classification Ascendante Hiérarchique obtenue après ACP. Le niveau de séparation permet d'obtenir 3 clusters (à gauche). Graphique des individus sur les dimensions 1 et 2, avec les 3 clusters issus de la CAH (à droite).

La caractérisation des classes est nécessaire afin de confirmer ou infirmer la correspondance avec la définition *a priori* des catégories présentes.

Une analyse par cluster et par variable est possible :

Le cluster 1 présente une moyenne de fréquence élevée ainsi qu'une contribution de l'espèce dans les captures bien supérieure aux autres catégories. Une moyenne de valeur bien supérieure est mise en évidence ainsi que des rejets inférieurs au cluster 3 sont observés.

Le cluster 2 est caractérisé par la moyenne des contributions de l'espèce dans les captures

ainsi que des rejets et la fréquence faible. Alors que la moyenne de la valeur est plus élevée que pour le cluster 3.

Les cluster 3 présente des moyennes de rejets élevées, des prix et des contributions de l'espèce dans les captures faibles.

Les variables d'activités de pêche et profondeur présentent de grands écarts types. Il apparaît compliqué de discriminer clairement des catégories à partir de ces indicateurs (**Erreur ! S**

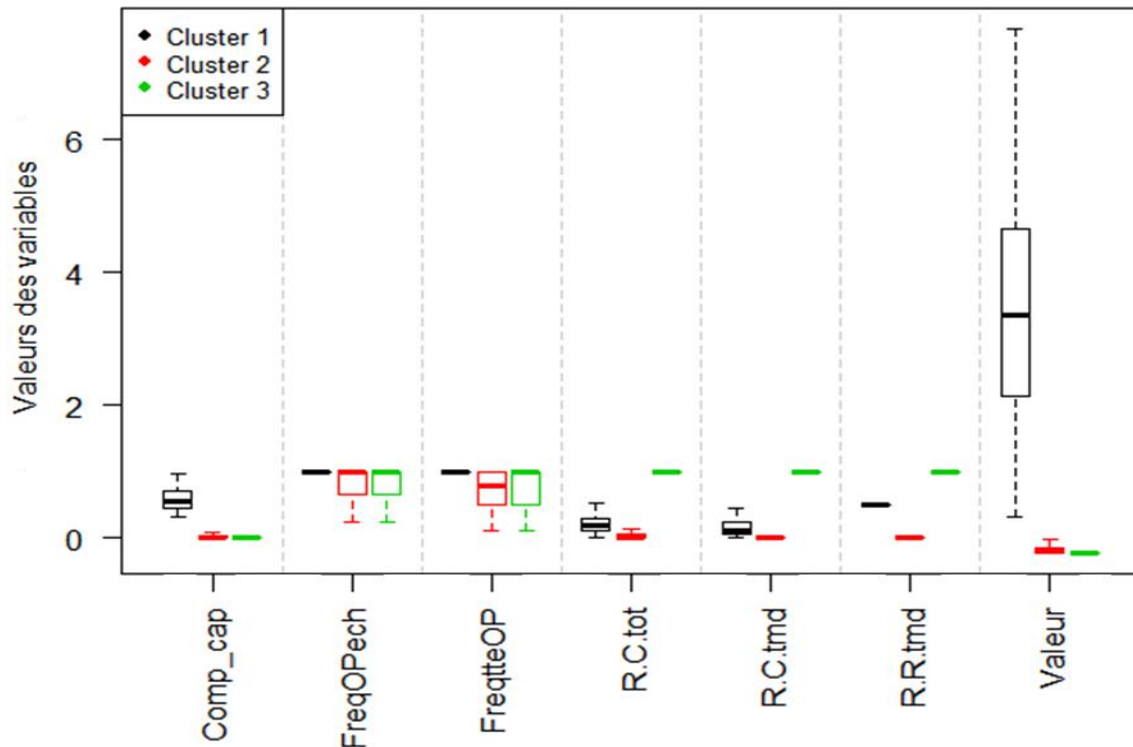


Figure 10. Boxplots de chaque variable pour les 3 clusters.

ource du renvoi introuvable.).

Une analyse des v.test par cluster montre que pour la classe 1, les variables contributions de l'espèce dans les captures et valeurs sont surreprésentées. Pour le second cluster, il apparaît que les variables de rejets sont sous-représentées. Tandis que les variables de rejets sont surreprésentées pour le cluster 3 (Annexe III.d).

Si les résultats précédents sont comparés à la définition *a priori*, alors il est possible de définir les différents clusters :

- la **classe 1** correspond à des **espèces cibles** à l'échelle de l'opération de pêche (avec une contribution de l'espèce dans les captures importante, une valeur importante ainsi que des espèces fréquentes associées à des rejets raisonnables)
- la **classe 2** correspond à des **espèces valorisées** à l'échelle de l'opération de pêche (avec une contribution de l'espèce dans les captures importante variable ou faible, une valeur importante ainsi que des espèces avec une fréquence variable ou faibles avec des rejets très faibles ou nuls)
- les **classes 3** correspond à des **espèces non valorisées** à l'échelle de l'opération de pêche (contribution de l'espèce dans les captures importante variable ou faible, une valeur faible ou nulle ainsi avec des rejets très élevés ou maximum).

Un autre résultat est à prendre en compte, aucune différence bien distincte n'est observée entre les 3 variables de rejets qui essaient de prendre en compte la taille minimale de

débarquement (Annexe III.a).

Dans le but de classifier les espèces, les proportions d'appartenance à chaque cluster des individus statistiques ont été calculées par espèce. Les proportions égales à 1 conduisent à une classification de l'espèce dans le cluster correspondant. Pour les autres espèces, leurs proportions d'appartenances en fonction des clusters peuvent être représentées (Figure 11). Attention les espèces avec une proportion = 1 dans une classe ne sont pas représentées. Ces espèces ont été classées dans la catégorie correspondante (Tableau 5).

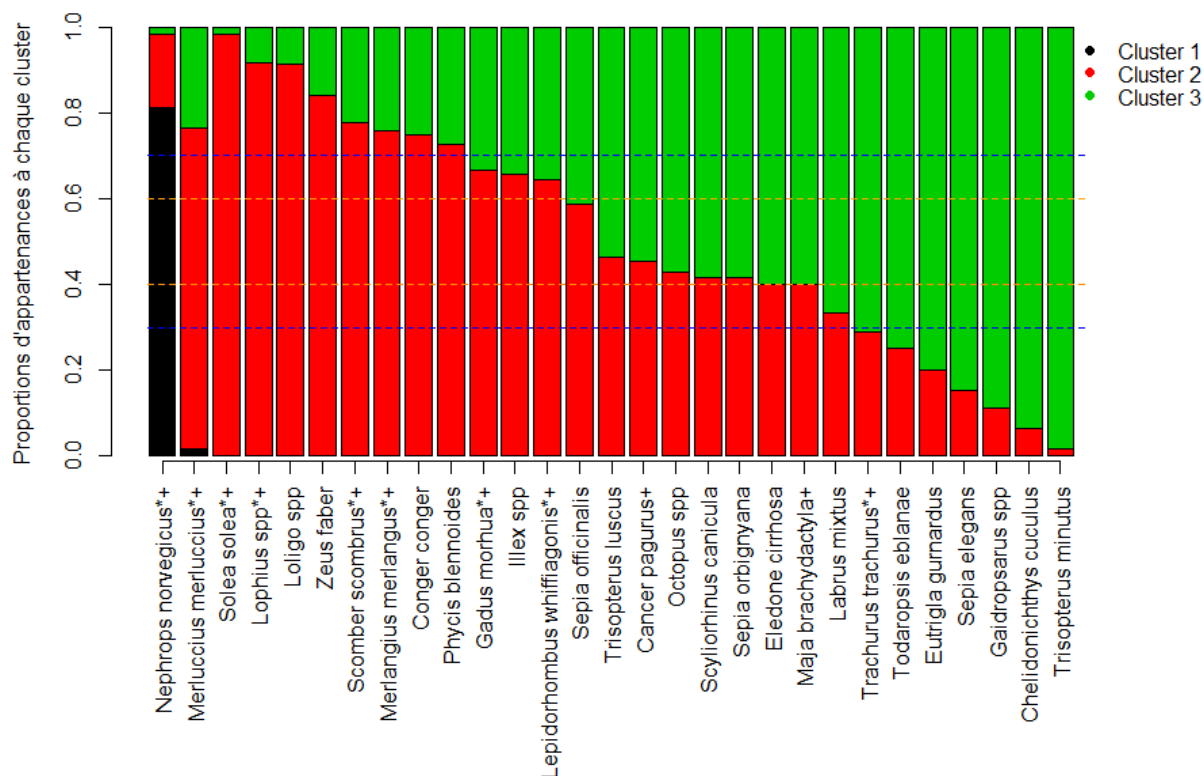


Figure 11. Barplot représentant les proportions d'appartenances aux différents clusters des individus statistiques d'une même espèce.

Les traits pointillés sont les seuils pressentis afin de définir les espèces mixtes.

Les pointillés orange sont pour les seuils 0.4 et 0.6. Les pointillés bleus sont pour les seuils 0.3 et 0.7.

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD.

Correspondance des couleurs des clusters : Noir = cible, rouge = valorisé, vert = non valorisé.

Pour définir le statut espèce accessoire mixte, il est nécessaire de définir un intervalle de mixité. Ce dernier correspond à une proportion d'appartenance à deux classes différentes (ex : 0.5 pour la classe valorisée et 0.5 pour la classe non valorisée). Pour cela, le nombre d'espèces dans la classe mixte est représenté en fonction de la largeur de l'intervalle (Figure 12). L'absence de saut ne permet pas de fixer un seuil évident. La décision de ce seuil sera donc fortement arbitraire. La gamme seuil choisie est : 0.3-0.7. Elle induit les définitions suivantes :

- Espèce cible : une espèce dont la proportion d'appartenance à la classe cible (noire) ≥ 0.2
- Espèce valorisée : une espèce dont la proportion d'appartenance à la classe valorisée (rouge) ≥ 0.7
- Espèce non valorisée : une espèce dont la proportion d'appartenance à la classe non valorisée (vert) ≥ 0.7
- Espèce mixte (bleu) : une espèce dont les proportions d'appartenance aux classes valorisée et non valorisée sont comprises entre 0.3 et 0.7.

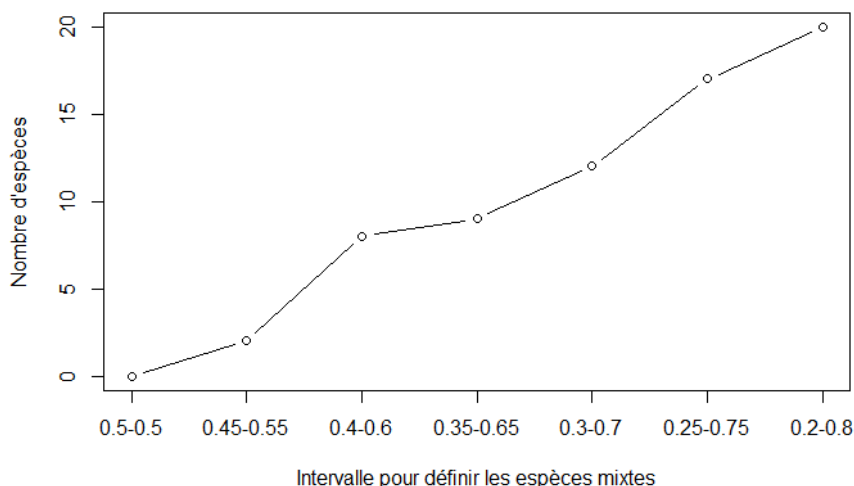


Figure 12. Graphique du nombre d'espèces classées dans la classe « mixte » selon les différents intervalles de proportions d'appartenances.

1.2. Analyse de la classification des données de débarquements

Afin de caractériser les différences entre les classes issues des données avec rejets et sans rejets. Une classification à l'aide de la méthode des k-means est utilisée (Annexe III.e). Une comparaison des résultats de cette approche et de la méthode précédente est effectuée pour les individus statistiques en commun.

Les résultats des k-means sur les débarquements variables issus des débarquements ont tendance à être catégorisés vers des ensembles moins valorisés que la CAH (incluant les rejets). En effet, 21 des 53 individus classés en cible par la CAH sont catégorisés valorisée par la méthode des k-means. De plus, pour les individus classés dans les catégories valorisée et non valorisée, environ 50% sont catégorisés différemment entre les méthodes (Tableau 2).

Tableau 2. Matrice de confusion de classification des individus statistiques.

En ligne, catégorisation à partir des rejets (ACP + CAH). En colonne, classification sans les rejets (k-means)

	Cible	Non Valorisée	Valorisée
Cible	32	0	21
Non Valorisée	0	44	47
Valorisée	0	259	268

La construction d'arbre de décision peut permettre de comprendre la distinction entre les classes valorisées et non valorisées pour les 2 méthodes. La variable quantitative séparant les classes valorisée et non valorisée n'est pas la même. Pour la CAH, ce sont les rejets alors que pour la méthode avec les k-means la fréquence discrimine les 2 classes. Ensuite, une différence des bornes de la variable valeur est visible, elle est plus importante pour les k-means (Figure 13).

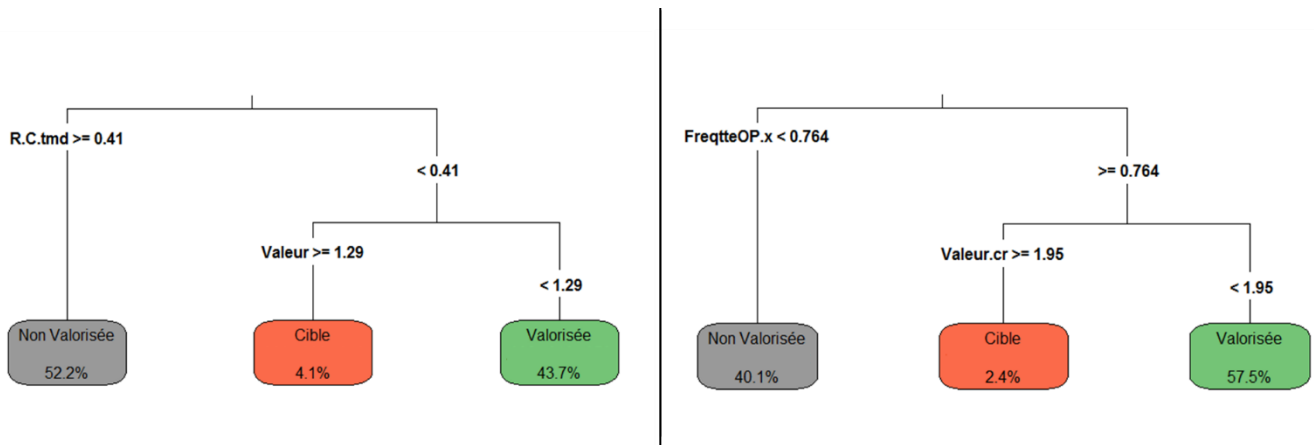


Figure 13. Arbre de classification pour les données avec rejets (à gauche) et les données sans rejets (à droite). Pour les variables Valeur et Valeur.cr, les quantités représentées sont centrées réduites.

2. Classification à l'échelle de la séquence de pêche

2.1. Analyse de la classification des données de captures

Une démarche identique à celle de l'opération de pêche est réalisée suite à l'élévation des rejets à l'échelle de la séquence. L'Analyse en Composante Principale ainsi que la CAH amènent aux mêmes conclusions que précédemment (Annexe III.f). Après avoir caractérisé les classes obtenues, le graphique des proportions d'appartenance à chaque cluster par espèces est réalisé. Des différences sont visibles avec l'échelle de l'opération de pêche notamment pour le merlu (*Merluccius merluccius*, Figure 14) pour lequel un grand nombre d'individus sont maintenant classés dans le cluster cible.

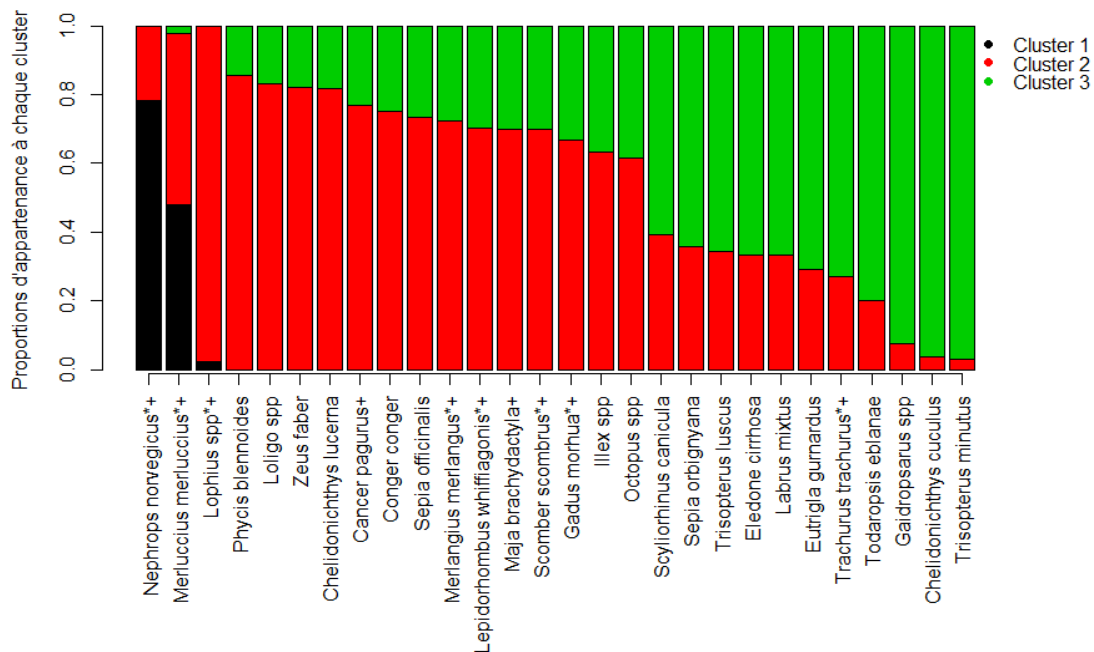


Figure 14. Barplot représentant les proportions d'appartenances aux différents clusters des individus statistiques d'une même espèce.

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD. Correspondance des couleurs des clusters : Noir = cible, rouge = valorisé, vert = non valorisé.

La création de séquence sur OBSMER est essentielle pour avoir des données comparables à SACROIS. Lors de cette agrégation, le métier attribué est celui correspondant à SACROIS. Le métier renseigné dans OBSMER, à l'échelle de l'opération de pêche, permet d'identifier des séquences constituées d'une mixité de métiers, par exemple, OTT_CRU (Chalut jumeaux à panneaux visant les crustacées) et OTT_DEF (Chalut jumeaux à panneaux visant les espèces démersales). Ces séquences sont qualifiées comme « non pures » (np). La représentation de la contribution de l'espèce dans les captures en fonction des différentes séquences pour le merlu met en évidence que l'ensemble des séquences non pures appartiennent à la classe cible (Figure 15).

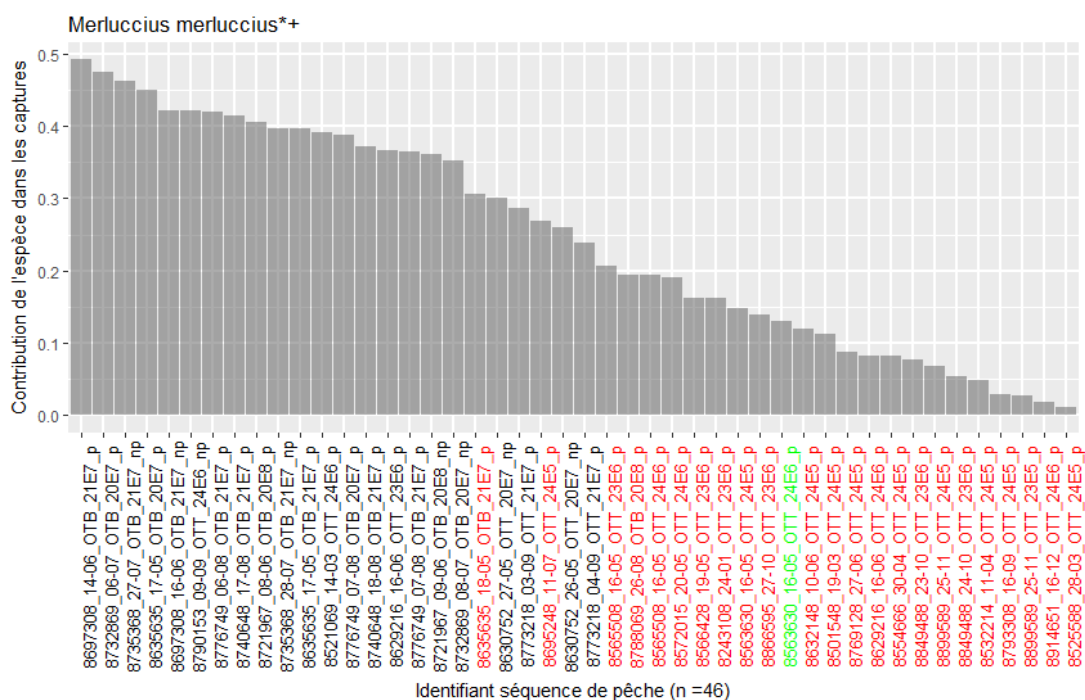


Figure 15. Contribution de l'espèce dans les captures en fonction des séquences étudiées. Les acronymes p et np, à la fin de l'identifiant de séquence signifient pure et non pure. Les identifiants en noir indiquent une appartenance à la classe cible ; rouge pour la classe valorisée et vert pour non valorisée.

2.2. Analyse de la classification des données de débarquements

2.2.1. Méthode des k-means

Les résultats des classifications de CAH et k-means sont comparés à l'aide d'une matrice de confusion (Annexe III.g). Cette dernière présente de fortes différences de classification pour la classe valorisée de la CAH qui est catégorisée non valorisée par les k-means. De plus, les individus classés cibles par la CAH sont catégorisés dans les 3 classes par les k-means (Tableau 3).

Tableau 3. Matrice de confusion de classification des individus statistiques. En ligne, catégorisation à partir des rejets (ACP + CAH). En colonne, classification sans les rejets (k-means)

	Cible	Non Valorisée	Valorisée
Cible	14	10	35
Non Valorisée	0	99	0
Valorisée	0	451	21

En analogie avec les résultats tenant compte des rejets à l'échelle de l'opération de pêche, les variables de rejets discriminent les classes valorisées et non valorisées. De plus, les limites des cibles et valorisées sont définies par la variable de contribution de l'espèce dans les captures estimées. Tandis que pour la méthode avec les k-means, la valeur discrimine toutes les classes (Figure 16). Pour les k-means, la variable discriminant les catégories valorisée et non valorisée est différente pour l'opération de pêche (fréquence) et la séquence de pêche (valeur).

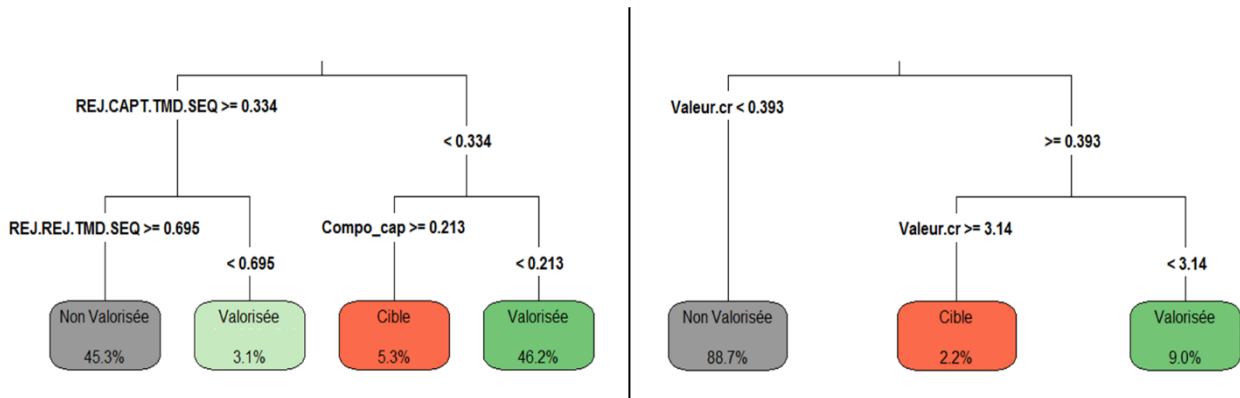


Figure 16. Arbre de classification pour les données avec rejets (à gauche) et les données sans rejets (à droite). Pour la variable Valeur.cr, les quantités représentées sont centrées réduites.

2.2.2. Apprentissage par les arbres de classification

Dans le but de construire un modèle prédictif, 50% des données OBSMER ne prenant pas en compte les rejets à l'échelle de la séquence ont été sélectionnées aléatoirement. Le résultat de ce modèle induit des différences avec l'arbre de classification prenant en compte l'ensemble des données. Le pourcentage d'individus statistiques appartenant à chaque classe, ainsi que la variable discriminante sont identiques. Cependant, les bornes de la variable explicative (valeur) sont différentes (Figure 17 et Figure 16, à droite). L'erreur globale des nœuds est approximativement égale à 10%, d'autres graphiques d'explorations du modèle ont été réalisés (Annexe III.h).

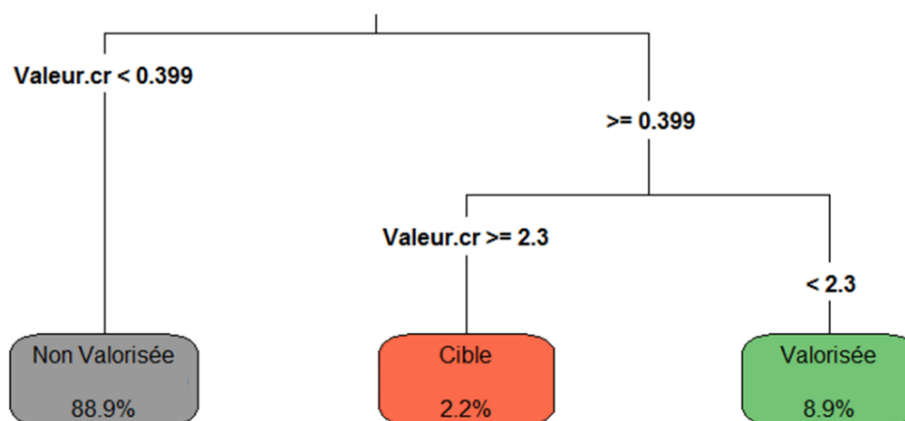


Figure 17. Arbre de classification du modèle prédictif dont les données d'entraînement sont 50 % des observations des séquences sans rejet. Pour la variable Valeur.cr, les quantités représentées sont centrées réduites.

Les prédictions issues du modèle précédent présentent de faibles différences avec la classification issues des k-means. Seulement 3 individus sur 325 sont classés différemment entre cible (pour k-means) et valorisée (pour arbre de classification), dû aux limites quantitatives de la variables discriminantes (Tableau 4).

Tableau 4. Matrice de confusion de classification des individus statistiques.
En ligne, prédictions du modèle avec l'arbre de classification. En colonne, classification sans les rejets (k-means)

	Cible	Non Valorisée	Valorisée
Cible	9	0	3
Non Valorisée	0	279	0
Valorisée	0	0	24

3. Comparaison des méthodes

Un tableau récapitulatif contenant l'ensemble des espèces catégorisées permet de réaliser une comparaison entre les différentes approches (Tableau 5).

Certaines espèces ont un statut différent selon l'approche. De plus, le nombre d'espèces est variable en fonction des données étudiées. En effet, la prise en compte ou non des espèces systématiquement rejetées explique cette variation.

De nombreuses différences sont observées entre les résultats de classification avec et sans rejets. Les espèces appartiennent à la même classe ou sont déclassées vers une catégorie moins valorisée. Quatre exceptions à l'échelle de l'opération de pêche, *Chelidonichthys cuculus*, *Lepidorhombus whiffiagonis*, *Sepia elegans* et *Todaropsis eblanae*, classées dans des ensembles plus valorisés. Il est important de rappeler que les méthodes sans les rejets ne tiennent pas compte des 18 espèces systématiquement rejetées (NA, Tableau 5).

Assez peu de différences sont observées pour le changement de catégorie d'espèces prenant en compte les rejets aux deux échelles (opération de pêche et séquence de pêche). Premièrement, le passage à la séquence induit l'introduction de nouvelles espèces telles que *Dicentrarchus labrax* et *Trachinus draco*. La mixité des métier au sein des séquences explique l'apparition de nouvelles espèces. Cinq espèces ont été classées différemment entre les deux échelles (*Cancer pagurus*, *Lepidorhombus whiffiagonis*, *Maja brachydactyla*, *Merluccius merluccius* et *Sepia officinalis*) appartenant toutes à une catégorie plus valorisée à l'échelle de la séquence.

La comparaison des approches sans rejet (k-means) présente des classifications différentes. En effet, le changement d'échelle induit, à l'échelle de la séquence, une catégorisation non valorisée pour l'ensemble des espèces (excepté : *Nephrops norvegicus*, classée en cible).

Enfin, l'utilisation de la méthode CART en utilisant uniquement la moitié des données pour apprendre entraîne une perte de trois espèces : *Gaidropsarus spp*, *Liocarcinus holsatus*, *Torpedo marmorata* (Tableau 5) qui ne peuvent donc être prédites.

Tableau 5. Tableau récapitulatif de la classification des espèces en fonction des données et de la méthode utilisée en utilisant le jeu de données OBSMER.

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD.
Noir = espèce cible, rouge = espèce valorisée, vert = espèce non valorisée, bleu = espèce mixte

Espèce	Opération de pêche avec rejets (ACP+CAH)	Opération de pêche sans rejet (K-means)	Séquence avec rejets (ACP+CAH)	Séquence sans rejet (K-means)	Séquence sans rejet (Arbre de classification)
Alloteuthis spp	Non Valorisée	NA	Non Valorisée	NA	NA
Alosa alosa+	Non Valorisée	NA	Non Valorisée	NA	NA
Argentina sphyraena	Non Valorisée	NA	Non Valorisée	NA	NA
Arnoglossus imperialis	Non Valorisée	NA	Non Valorisée	NA	NA
Arnoglossus laterna	Non Valorisée	NA	Non Valorisée	NA	NA
Callionymus lyra	Non Valorisée	NA	Non Valorisée	NA	NA
Callionymus maculatus	Non Valorisée	NA	Non Valorisée	NA	NA
Cancer pagurus+	Mixte	Mixte	Valorisée	Non Valorisée	Non Valorisée
Capros aper*	Non Valorisée	NA	Non Valorisée	NA	NA
Cepola macrophthalma	Non Valorisée	NA	Non Valorisée	NA	NA
Chelidonichthys cuculus	Non Valorisée	Mixte	Non Valorisée	Non Valorisée	Non Valorisée
Chelidonichthys lucerna	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Conger conger	Valorisée	Mixte	Valorisée	Non Valorisée	Non Valorisée
Dicentrarchus labrax+	NA	NA	Valorisée	Non Valorisée	Non Valorisée
Eledone cirrhosa	Mixte	Non Valorisée	Mixte	Non Valorisée	Non Valorisée
Eutrigla gurnardus	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée
Gadus morhua*+	Mixte	Non Valorisée	Mixte	Non Valorisée	Non Valorisée
Gaidropsarus spp	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée	NA
Galathee spp	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée
Glyptocephalus cynoglossus	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Gobiidae	Non Valorisée	NA	Non Valorisée	NA	NA
Illex spp	Mixte	Mixte	Mixte	Non Valorisée	Non Valorisée
Labrus bergylta	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Labrus mixtus	Mixte	Non Valorisée	Mixte	Non Valorisée	Non Valorisée
Lepidorhombus boscii*	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Lepidorhombus whiffiagonis*+	Mixte	Valorisée	Valorisée	Non Valorisée	Non Valorisée
Liocarcinus holsatus	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée	NA
Loligo spp	Valorisée	Mixte	Valorisée	Non Valorisée	Non Valorisée
Lophius spp*+	Valorisée	Valorisée	Valorisée	Non Valorisée	Non Valorisée
Maja brachydactyla+	Mixte	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Melanogrammus aeglefinus*+	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Merlangius merlangus*+	Valorisée	Valorisée	Valorisée	Non Valorisée	Non Valorisée
Merluccius merluccius*+	Valorisée	Valorisée	Cible	Non Valorisée	Non Valorisée
Microchirus variegatus	Non Valorisée	NA	Non Valorisée	NA	NA
Micromesistius poutassou*	Non Valorisée	NA	Non Valorisée	NA	NA
Microstomus kitt	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Molva molva*+	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Mullus surmuletus	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Mustelus asterias	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Necora puber	Non Valorisée	NA	Non Valorisée	NA	NA
Nephrops norvegicus*+	Cible	Cible	Cible	Cible	Cible
Octopus spp	Mixte	Non Valorisée	Mixte	Non Valorisée	Non Valorisée
Palinurus mauritanicus+	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Phrynorhombus norvegicus	Non Valorisée	NA	Non Valorisée	NA	NA
Phycis blennoides	Valorisée	Mixte	Valorisée	Non Valorisée	Non Valorisée
Pleuronectes platessa*+	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Pollachius pollachius*+	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Psetta maxima*+	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Raja sp	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Scomber scombrus*+	Valorisée	Mixte	Valorisée	Non Valorisée	Non Valorisée
Scylliorhinus canicula	Mixte	Mixte	Mixte	Non Valorisée	Non Valorisée
Scyllarus arctus	Non Valorisée	NA	Non Valorisée	NA	NA

Sepia elegans	Non Valorisée	Mixte	Non Valorisée	Non Valorisée	Non Valorisée
Sepia officinalis	Mixte	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Sepia orbignyana	Mixte	Mixte	Mixte	Non Valorisée	Non Valorisée
Sepiola	Non Valorisée	NA	Non Valorisée	NA	NA
Solea solea*+	Valorisée	Valorisée	Valorisée	Non Valorisée	Non Valorisée
Spondyliosoma cantharus	Valorisée	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée
Sprattus sprattus*	Non Valorisée	NA	Non Valorisée	NA	NA
Todaropsis eblanae	Non Valorisée	Valorisée	Non Valorisée	Non Valorisée	Non Valorisée
Torpedo marmorata	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée	NA
Trachinus draco	NA	NA	Non Valorisée	NA	NA
Trachurus trachurus*+	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée
Trisopterus luscus	Mixte	Mixte	Mixte	Non Valorisée	Non Valorisée
Trisopterus minutus	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée	Non Valorisée

4. Analyse de la classification supervisée des données SACROIS

Deux types de données d'entraînement ont été testés, en prenant en compte ou non les rejets. Les résultats des modèles ainsi que les graphiques associés ont été réalisés (Annexe III.i). La taille du jeu d'entraînement est très faible (198 individus statistiques) par rapport aux données à prédire (331 216 individus statistiques). La faible proportion de données d'entraînement s'explique par la taille du jeu de données OBSMER, mais aussi par l'absence de d'identifiant de séquence commun entre SACROIS et OBSMER entraînant une perte non négligeable de séquences (environ 20 %). La performance du modèle est affectée par la faible proportion des données allouées à l'entraînement. Une forte hétérogénéité des classes est observée quel que soit le jeu de données choisi. Cependant, les données d'entraînement issues de OBSMER prenant en compte les rejets catégorisent plus de 59% des individus statistiques (198 456 individus) dans la classe valorisée. Tandis que pour l'autre jeu de données d'entraînement, la classe non valorisée est composée d'environ 90% des individus statistiques (296 567 individus). La prise en compte ou non des rejets induit une forte différence entre les classes valorisée et non valorisée. De plus, comme vu précédemment (Tableau 3), les données ne prenant pas en compte les rejets ont tendance à catégoriser les individus statistiques dans des classes moins valorisées (Tableau 6).

Tableau 6. Matrice de confusion de classification des individus statistiques.

En ligne, prédictions du modèle avec l'arbre de classification prenant en compte les rejets. En colonne, classification à partir des données d'entraînement sans les rejets

	Cible	Non Valorisée	Valorisée
Cible	7226	18807	18656
Non Valorisée	0	88071	0
Valorisée	361	189689	8406

Un tableau récapitulatif à l'échelle de l'espèce permet également de mettre en évidence les différences de classification en fonction des données d'entraînement choisies (Annexe III.j). L'application de la méthode sur les données SACROIS augmente considérablement le nombre d'espèces à catégoriser. En effet, 103 espèces sont prises en compte dans SACROIS contre seulement 68 dans OBSMER.

Discussion

1. Statut des espèces

Les résultats issus de la caractérisation à l'échelle de l'opération de pêche prenant en compte les rejets peuvent être jugés comme satisfaisants. Certaines espèces permettent de justifier cette énonciation. Par exemple, la langoustine (*Nephrops norvegicus*) est classée comme cible, le merlu (*Merluccius merluccius*), la sole (*Solea solea*) et la lotte (*Lophius spp.*) sont des espèces valorisées, tandis que le sanglier (*Capros aper*), le merlan bleu (*Micromesistius poutassou*) et le dragonnet (*Callionymus lyra*) composent la catégorie non valorisée (Tableau 5). Ces 3 dernières espèces sont connues pour être considérées comme non commerciales (fortement rejetées) ou valorisées par des pêcheries spécifiques (Van Der Veer et al., 1900 ; Olaso et al., 1998 ; White et al., 2011).

L'absence de différences significatives entre les variables de rejets est étudiée uniquement pour les espèces qui sont soumises à une TMD. Des rejets dus à une taille inférieure à la taille réglementaire sont observés. Cependant, l'absence de différences significatives témoigne de pratiques telles que les rejets sélectifs (Annexe III.a). Ces derniers peuvent être expliqués par la détérioration du poisson dans le chalut. L'analyse ne permet pas d'exclure la pratique d'highgrading ou de mettre en évidence des dépassements de quotas. Ces pratiques peuvent avoir lieu, notamment dans les pêcheries mixtes (Batsleer et al., 2015).

La définition des seuils visant à caractériser le statut des espèces est arbitraire. La validité de ces limites est une réelle question qui devra être explorée plus finement.

Les résultats renseignent sur le biais suite à la suppression des rejets ou au changement d'échelle. La suppression des rejets à l'échelle de l'opération de pêche, entraîne une différence de caractérisation des catégories valorisée et non valorisée. La variable discriminante est la fréquence. Donc, en l'absence de rejets, la définition du statut d'espèce change. Une espèce valorisée est définie par des individus fréquemment rencontrés dans les captures (Figure 13). De ce fait, certaines espèces présentant de faibles taux de rejets sont déclassées dans la catégorie non valorisée. Afin de valider cette approche, il est nécessaire d'admettre qu'une espèce valorisée est caractérisée par des individus fréquemment rencontrés dans les captures. A l'échelle de la séquence de pêche, la suppression des rejets, a également pour conséquence un changement de la variable discriminante. La valeur définit l'ensemble des classes (Figure 16). Les individus valorisés sont donc ici considérés comme des individus ayant un intérêt commercial.

L'approche à l'échelle de la séquence présente des différences avec l'opération de pêche. Tout d'abord, l'analyse avec les rejets met en évidence des individus et des espèces ayant changé de classes (Tableau 5) et la variable change (Figure 13 et Figure 16). La différence de classification des individus s'explique par l'augmentation du nombre d'individus statistiques ainsi que par la mixité des métiers dans les séquences (Figure 15). Cette mixité est due à l'hétérogénéité des opérations de pêche. Effectivement, dans notre cas, un pêcheur peut pratiquer deux métiers avec le même engin, dans la même zone, la même journée si sa capture est majoritairement constituée de poissons ou de crustacés. La mixité des métiers va avoir pour conséquence d'augmenter les captures d'espèces démersales au sein de la séquence et donc de biaiser la classification. Le changement de variable discriminante entre les échelles peut s'expliquer par la définition de la fréquence (Annexe II, équation 7 et 10). Cette dernière est définie comme le nombre d'apparition d'une espèce au sein de la marée. Ainsi, par construction la variabilité de la fréquence est différente selon l'échelle (il y a moins

de séquence de pêche par marée que d'opérations de pêche). Au contraire, la variabilité de la valeur, de augmente avec le passage à l'échelle de la séquence de pêche. La construction des variables, leur variabilité ainsi que la transition d'échelle expliquent le changement des variables discriminantes.

A l'échelle de la séquence de pêche traitant uniquement les débarquements, un individu statistique ayant une faible valeur est considéré comme non valorisé. L'inverse définit un individu cible. Cette approche centrée sur la valeur est potentiellement une vision plus globale des métiers étudiés. En effet, les catégories valorisées ou cibles sont composées de quatre espèces (la langoustine (*Nephrops norvegicus*), le merlu (*Merluccius merluccius*), la sole (*Solea solea*) et la lotte (*Lophius spp.*)). Ces dernières constituent les espèces majoritaires en valeur quelle que soit l'approche (Annexe IV). Cependant, l'assemblage d'espèces est le résultat de l'action de pêche, mais ne peut refléter avec précision l'intention réelle de ciblage des pêcheurs (Ulrich et al., 2012).

La comparaison sur le jeu de données OBSMER des méthodes de classification (avec et sans apprentissage) donne des indices sur la possibilité d'exploiter les sorties d'analyse d'OBSMER pour travailler sur les données SACROIS (Figure 17). Les résultats sont dépendants des données d'entraînement. La classification des espèces en utilisant SACROIS a mis en évidence le même constat. La jointure entre les séquences OBSMER et SACROIS entraîne une perte de données d'entraînement. Cependant, si l'ensemble des données OBSMER étaient prises en compte, la proportion des données d'entraînement par rapport aux données de prédictions (SACROIS) serait inférieure à 0,2%. Cette faible proportion affecte la performance du modèle. En effet, plus les données d'entraînement sont importantes, plus les performances du modèle sont satisfaisantes (Hastie et al., 2008). Les taux d'erreur des deux modèles (avec et sans rejets) sont élevés (Annexe III.i). Le plus faible observé est pour les données d'entraînement ne prenant pas en compte les rejets (0,47 contre 0,86 avec les rejets). Cependant, la suppression des rejets induit un biais dans la classification (cf comparaison sur jeu de données OBSMER). L'ensemble des raisons évoquées démontre que cette approche ne donne pas de résultats exploitables en l'état sur le jeu de données SACROIS.

2. Limites des données

Les données OBSMER sont des captures renseignées à très fine échelle jugées comme fiables. Cependant, l'un des principaux problèmes liés aux données d'observateurs embarqués est la grande variation spatiale et temporelle des rejets. Hormis la difficulté évidente d'obtenir des estimations précises pour un phénomène très variable, si le plan d'échantillonnage n'en tient pas compte, cette forte variation pourrait masquer un biais dans l'estimation, qui sera transféré et multiplié lors de l'élévation des estimations au niveau de la flotte entière ou par strate (Allen et al., 2001). Seules 0,2 % des marées ont été échantillonnées en 2014 pour le métier étudié (Cornou et al., 2015). De plus, ce programme est basé sur le volontariat des pêcheurs. Cela met en évidence deux problèmes, le biais de « déploiement » résultant d'une distribution non aléatoire d'observateurs parmi les unités d'échantillonnage et les effets d'observateurs dus à des changements de pratiques de pêche ou d'emplacement lorsque les observateurs sont à bord (Benoit & Allard., 2009).

Les données SACROIS sont des débarquements issus d'un algorithme de croisement renseignés à l'échelle de la séquence de pêche. Le nombre important de données explique qu'elles soient considérées comme représentatives à l'échelle des pêcheries. Cependant, les débarquements sont déclarés par les pêcheurs. Certains débarquements n'ayant pas lieu en

criée (exemple : vente directe) ne sont potentiellement pas renseignés par les pêcheurs désirent frauder. Aucune information sur les rejets n'est disponible. Enfin, la correction par l'algorithme n'est pas systématiquement optimale, des erreurs d'identification d'espèces ou de métiers ont été répertoriées.

Les deux jeux de données sont nécessaires dans le but de définir le statut des espèces à partir de données représentatives de la pêcherie. La multiplication des observations embarquées permettra une meilleure compréhension de la variabilité des captures, nécessaire à la classification des espèces.

3. Limites des méthodes

La définition *a priori* du statut des espèces en exploitant les variables disponibles fut mis en place à l'aide de la bibliographie (Annexe II). D'autres définitions pourraient être possibles avec d'autres données disponibles ou à une autre échelle.

La caractérisation du statut des espèces à partir de l'ACP et la CAH donne des résultats intéressants. Cette approche est possible sur les données OBSMER car le nombre d'individus est limité. Si le jeu de données était plus conséquent, plusieurs solutions seraient envisageables. Premièrement, une alternative serait de réaliser un tirage aléatoire des individus (afin de réduire leur nombre), puis, de faire une CAH et d'utiliser un arbre de classification pour prédire pour la classe des individus restants. Une autre alternative serait de choisir une autre méthode de classification non supervisée (Annexe I).

La suppression des rejets aide à comprendre si les données de rejets sont indispensables pour caractériser le statut de l'espèce. L'analyse des données sans rejet requiert une méthode acceptant un faible nombre de variable (peu de variables disponibles). Les résultats montrent que l'ajout d'une partie des données précédemment catégorisées aide à mieux comprendre la classification. Les variations des résultats entre les deux méthodes sont donc plus visibles (Annexe III.e et Annexe III.g). La combinaison des méthodes CAH sur les rejets puis k-means sans les rejets en utilisant les centroïdes des classes des CAH est donc intéressante dans le but de catégoriser les espèces à partir de variables quantitatives.

Le changement d'échelle repose sur une importante hypothèse : les individus systématiquement rejetés sont sous-estimés. Ce choix s'explique par la configuration du jeu de donnée final (SACROIS) dans lequel ces espèces n'apparaîtront pas. Cependant, une alternative aurait pu être d'élever le poids de ces espèces selon l'effort de pêche. En effet, un poids moyen par heure de pêche et par espèce aurait pu être calculé et intégré à l'élévation. L'agrégation à l'échelle de la séquence est nécessaire afin de comprendre le biais introduit. Travailler à plus grande échelle bruite les données par rapport à l'opération de pêche. Les séquences composées de différents métiers en sont un exemple. Malgré les hypothèses et le bruit intégrés, les analyses suggèrent que les résultats de l'élévation sont utilisables.

La suppression des rejets à l'échelle de la séquence change la vision des espèces cibles et accessoires. Le merlan (*Merlangius merlangus*) est une espèce sous quota avec une TMD qui est déclassée comme non valorisée. Au vu des résultats des autres méthodes et des fractions rejetées pour cette espèce, le statut considéré du merlan ne devrait pas être « non valorisée ». Les données de séquences traitant uniquement des débarquements, ne seraient donc pas adaptées pour définir un statut d'espèce visant à améliorer la gestion.

De ce fait, il a été choisi de tester des méthodes d'apprentissage avec des données d'entraînement prenant en compte ou non les rejets.

La méthode choisie pour classer les individus sous SACROIS doit accepter des données d'entraînement avec un faible nombre de variables explicatives. Ces 3 dernières sont quantitatives alors que la variable à expliquer est qualitative. La méthode de random forest a pour but de tirer aléatoirement les différentes variables explicatives afin de créer différents arbres (Hastie et al., 2008). Un nombre trop faible de variables ne rend pas l'utilisation de cette méthode optimale. L'apprentissage à partir d'un arbre de classification a été mis en place. Cette méthode présente des résultats assez peu satisfaisants. La classification obtenue sur les données SACROIS peut être considérée comme non robuste et non exploitable pour cette pêcherie.

Plusieurs solutions pourraient être envisagées afin d'améliorer cette catégorisation. Un changement de méthode d'apprentissage couplé à une augmentation du nombre de variables explicatives engendrerait une amélioration des résultats. Une méthode alternative d'apprentissage peut être mise en œuvre : l'analyse discriminante décisionnelle. L'objectif est également de définir des règles de décisions. Il s'agit de la modélisation d'une variable qualitative à multiples modalités par différentes variables quantitatives. Différents modèles d'analyse discriminante décisionnelle auraient pu être testés : règle linéaire et quadratique de décision dans le cas gaussien ou la règle non paramétrique par k plus proches voisins (Bardos & Zhu, 1997 ; Wikistat, 2016). L'utilisation de cette méthode générerait potentiellement de meilleures performances. De plus, ces dernières pourraient être améliorées suite à l'augmentation du nombre de variables. Un ajout de variables supplémentaires induit une modification de la définition *a priori*. Cette variable doit être commune aux différents jeux de données et préférentiellement quantitative (sinon d'autres méthodes doivent être choisies). La prise en compte de l'état des stocks actuels et de la biomasse disponible des espèces dans le milieu peut être envisagée. Un indicateur de la probabilité de captures pourrait être construit en tenant compte de la sélectivité de l'engin, de l'effort alloué, de la biomasse de l'espèce, de la saison, Une variable tenant compte de la dynamique de vente pourrait également être considérée afin de définir le statut d'espèce.

Cependant, le faible nombre de données d'entraînement reste toujours une difficulté. De plus, il ne faut pas omettre que les celles-ci ont été construites à partir d'un jeu de données biaisé.

Cette étude a pu mettre en évidence des espèces sous quotas et non valorisée (Tableau 5). Les plus préoccupantes sont les espèces systématiquement rejetées. Elles sont identifiées grâce au programme OBSMER, trois espèces potentiellement bloquantes sont répertoriées : le sprat (*Sprattus sprattus*), le sanglier (*Capros aper*) et le merlan bleu (*Micromesistius poutassou*). Le sprat est capturé en faible quantité comme le sanglier qui bénéficie d'un statut particulier. Pour les pêcheries françaises, un TAC 0 est attribué à cette espèce (Journal officiel de l'Union européenne N°120, 2018). Les captures du merlan bleu représentent une proportion dans les captures totales supérieure à 1% avec des taux de rejets maximum (Annexe V). Avec la mise en place de l'obligation de débarquement, cette espèce pourrait être problématique pour cette pêcherie en entraînant une interdiction de pêche pour les bateaux le capturant accessoirement. Le chincharde (*Trachurus trachurus*) pourrait être également problématique. Cependant, son statut d'espèce non valorisée dépend essentiellement du seuil fixé. Il est donc difficile d'affirmer si le chincharde sera une espèce bloquante dans cette pêcherie.

Conclusion

Cette première approche est intéressante, les biais des différentes données et méthodes ont été identifiés. Comprendre et intégrer ces biais aideront à mieux définir le statut des espèces. La compréhension des variabilités de captures et de richesse spécifique dans la pêche mixte démersale du golfe de Gascogne est également essentielle.

Un ensemble de métier ciblant la langoustine a été étudié. En perspective, il serait intéressant de travailler à l'échelle des métiers (combinaison d'un engin de pêche, d'une ou plusieurs espèces cibles et d'une zone de pêche), de la flottille (ensemble de navires présentant des caractéristiques communes et provenant de la même zone géographique) et donc de pouvoir catégoriser les espèces à l'échelle des flottilles pour les pêcheries démersales du golfe de Gascogne. L'étude des liens spatio-temporels permettra de comprendre une partie de la variabilité des captures, pour ensuite, améliorer la classification du statut des espèces. De plus, l'amélioration des informations concernant les rejets est attendue dans les prochaines années. Enfin, une étude pluriannuelle démontrerait si le comportement pêcheurs s'est adapté à la suite de l'annonce de l'obligation de débarquement. Définir les statuts d'espèces est une étape dans le but d'instaurer une gestion plurispécifique dans la pêche mixte démersale du golfe de Gascogne.

Bibliographie

- ALLEN, Michelle, KILPATRICK, David, ARMSTRONG, Mike, BRIGGS, Richard, PÉREZ, Nélide et COURSE, Grant, 2001. Evaluation of sampling methods to quantify discarded fish using data collected during discards project EC 95/094 by Northern Ireland, England and Spain. *Fisheries Research*. janvier 2001. Vol. 49, n° 3, pp. 241-254. DOI [10.1016/S0165-7836\(00\)00207-1](https://doi.org/10.1016/S0165-7836(00)00207-1).
- BARDOS, M et ZHU, W H, 1997. Comparaison de l'analyse discriminante linéaire et des réseaux de neurones. Application à la détection de défaillance d'entreprises. *Revue de statistique appliquée*. 1997. Vol. 45, n° 4, pp. 65-92.
- BATSLEER, J., HAMON, K. G., VAN OVERZEE, H. M. J., RIJNSDORP, A. D. et POOS, J. J., 2015. High-grading and over-quota discarding in mixed fisheries. *Reviews in Fish Biology and Fisheries*. décembre 2015. Vol. 25, n° 4, pp. 715-736. DOI [10.1007/s11160-015-9403-0](https://doi.org/10.1007/s11160-015-9403-0).
- BENOÎT, Hugues P. et ALLARD, Jacques, 2009. Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? *Canadian Journal of Fisheries and Aquatic Sciences*. décembre 2009. Vol. 66, n° 12, pp. 2025-2039. DOI [10.1139/F09-116](https://doi.org/10.1139/F09-116).
- BREIMAN, Leo, FRIEDMAN, Jerome, STONE, Charles J. et OLSHEN, R.A., 1984. *Classification and regression trees*. Wadsworth & Brooks. ISBN 978-0-412-04841-8.
- CATCHPOLE, Thomas L., RIBEIRO-SANTOS, Ana, MANGI, Stephen C., HEDLEY, Christopher et GRAY, Tim S., 2017. The challenges of the landing obligation in EU fisheries. *Marine Policy*. août 2017. Vol. 82, pp. 76-86. DOI [10.1016/j.marpol.2017.05.001](https://doi.org/10.1016/j.marpol.2017.05.001).
- CHARUAU Anatole. 1988. Les pêcheries mixtes de langoustine et merlu du Golfe de Gascogne, modélisation bio-économique et simulation des procédures de gestion. DGXIV/B/1 : 2378 - IFREMER 88/1210774/BF. <http://archimer.ifremer.fr/doc/00000/2569/>
- CLUCAS, I., 1997. *A study of the options for utilization of bycatch and discards from marine capture fisheries* [en ligne]. Food and Agriculture Organization of the United Nations. FAO fisheries circular. <https://books.google.fr/books?id=Yk0YAQAIAAJ>
- CORNOU Anne-Sophie, GOASCOZ Nicolas, SCAVINNER Marion, CHASSANITE Aurore, DUBROCA Laurent, ROCHET Marie-Joelle, 2017. *Captures et rejets des métiers de pêche français. Résultats des observations à bord des navires de pêche professionnelle en 2016*. <http://archimer.ifremer.fr/doc/00418/52945/>
- CORNOU, Anne-Sophie, QUINIO-SCAVINNER, Marion, DELAUNAY, Damien, DIMEET, Joel, GOASCOZ, Nicolas, DUBE, Benoit, FAUCONNET, Laurence, ROCHET, Marie-Joelle, ASTROU, Adele, BISEAU, Alain, METRAL, Luisa et PERE, Anthony, 2015. *Observations à bord des navires de pêche professionnelle. Bilan de l'échantillonnage 2014*. <http://doi.org/10.13155/39722>
- DAVIE, Sarah et LORDAN, Colm, 2011. Definition, dynamics and stability of métiers in the Irish otter trawl fleet. *Fisheries Research*. novembre 2011. Vol. 111, n° 3, pp. 145-158. DOI [10.1016/j.fishres.2011.07.005](https://doi.org/10.1016/j.fishres.2011.07.005).
- DEPORTE, N., ULRICH, C., MAHEVAS, S., DEMANECHÉ, S. et BASTARDIE, F., 2012. Regional métier definition: a comparative investigation of statistical methods using a workflow applied to international otter trawl fisheries in the North Sea. *ICES Journal of Marine Science*. 1 mars 2012. Vol. 69, n° 2, pp. 331-342. DOI [10.1093/icesjms/fsr197](https://doi.org/10.1093/icesjms/fsr197).
- DROUINEAU, Hilaire, MAHÉVAS, Stéphanie, PELLETIER, Dominique et BELIAEFF, Benoît, 2006. Assessing the impact of different management options using ISIS-Fish: the French *Merluccius merluccius* – *Nephrops norvegicus* mixed fishery of the Bay of Biscay. *Aquatic Living Resources*. janvier 2006. Vol. 19, n° 1, pp. 15-29. DOI [10.1051/alr:2006002](https://doi.org/10.1051/alr:2006002).
- EUROPEAN COMMISSION, 2016. *Regulation of the European Parliament and of the Council on establishing a multi-annual plan for demersal stocks in the North Sea and the fisheries exploiting those stocks and repealing Council Regulation (EC) 676/2007 and Council Regulation (EC) 1342/2008*. 3 août 2016.
- FERNANDES, P. G., COULL, K., DAVIS, C., CLARK, P., CATARINO, R., BAILEY, N., FRYER, R. et POUT, A., 2011. Observations of discards in the Scottish mixed demersal trawl fishery. *ICES*

- Journal of Marine Science*. 1 septembre 2011. Vol. 68, n° 8, pp. 1734-1742. DOI [10.1093/icesjms/fsr131](https://doi.org/10.1093/icesjms/fsr131).
- FOWLER, Jim, COHEN, Louis et JARVIS, Phil, 1998. *Practical statistics for field biology*. 2nd ed. Chichester ; New York : Wiley. ISBN 978-0-471-98295-1. QH318.5 .F68 1998.
- GARCÍA, Dorleta, PRELLEZO, Raúl, SAMPEDRO, Paz, DA-ROCHA, José María, CASTRO, José, CERVIÑO, Santiago, GARCÍA-CUTRÍN, Javier et GUTIÉRREZ, María-José, 2016. Bioeconomic multistock reference points as a tool for overcoming the drawbacks of the landing obligation. *ICES Journal of Marine Science: Journal du Conseil*. 21 mars 2016. pp. fsw030. DOI [10.1093/icesjms/fsw030](https://doi.org/10.1093/icesjms/fsw030).
- GOURGUET, S., MACHER, C., DOYEN, L., THÉBAUD, O., BERTIGNAC, M. et GUYADER, O., 2013. Managing mixed fisheries for bio-economic viability. *Fisheries Research*. février 2013. Vol. 140, pp. 46-62. DOI [10.1016/j.fishres.2012.12.005](https://doi.org/10.1016/j.fishres.2012.12.005).
- GUILLEN, Jordi, MACHER, Claire, MERZÉREAUD, Mathieu, BERTIGNAC, Michel, FIFAS, Spyros et GUYADER, Olivier, 2013. Estimating MSY and MEY in multi-species and multi-fleet fisheries, consequences and limits: an application to the Bay of Biscay mixed fishery. *Marine Policy*. juillet 2013. Vol. 40, pp. 64-74. DOI [10.1016/j.marpol.2012.12.029](https://doi.org/10.1016/j.marpol.2012.12.029).
- HASTIE, Trevor, TIBSHIRANI, Robert et FRIEDMAN, Jerome, 2008. *The Elements of Statistical Learning*. Second edition. New York, NY : Springer New York. Springer Series in Statistics. ISBN 978-0-387-84857-0.
- HERVIN, Paul-Antoine et BEGOT, Eric, 2013. *Spécifications logicielles Sacrois V3*. Direction des Pêches Maritimes et de l'Aquaculture (DPMA). Référence du document : BRE.SACV4.SFG.13.0006
- Journal Officiel de l'Union Européenne, 2013. RÈGLEMENT (UE) N° 1380/2013 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 11 décembre 2013 relatif à la politique commune de la pêche, modifiant les règlements (CE) n° 1954/2003 et (CE) n° 1224/2009 du Conseil et abrogeant les règlements (CE) n° 2371/2002 et (CE) n° 639/2004 du Conseil et la décision 2004/585/CE du Conseil
- Journal Officiel de l'Union Européenne, 2018. RÈGLEMENT (UE) 2018/ 120 DU CONSEIL du 23 janvier 2018 - établissant, pour 2018, les possibilités de pêche pour certains stocks halieutiques et groupes de stocks halieutiques, applicables dans les eaux de l'Union et, pour les navires de pêche de l'Union, dans certaines eaux n'appartenant pas à l'Union et modifiant le règlement (UE) 2017/ 127.
- Journal Officiel de la République Française (JORF), 2013. *Arrêté du 28 janvier 2013 déterminant la taille minimale ou le poids minimal de capture et de débarquement des poissons et autres organismes marins pour la pêche professionnelle*. Version consolidée au 02 mars 2018. 14 février 2013. NOR : TRAM1240494A.
- KUMAR, A Biju et DEEPTHI, GR, 2006. Trawling and by-catch: Implications on marine ecosystem. *Current Science*. 2006. Vol. 90, n° 8, pp. 922-931.
- LE BLOND, Samuel, 2018. PRESTO/DTI-DCD/12-001: *Agrégation FREE2*. Système d'Informations Halieutiques (SIH).
- LÊ, Sébastien, JOSSE, Julie et HUSSON, François, 2008. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*. 2008. Vol. 25, n° 1. DOI [10.18637/jss.v025.i01](https://doi.org/10.18637/jss.v025.i01).
- LEMERCIER, Claire, MILANI, Pauline et SOFIO, Séverine, 2010. *Tutoriel FactoMineR pour l'analyse des correspondances multiples avec une petite annexe sur la classification automatique*. 20 avril 2010.
- MACHER, Claire, 2009. *Productions jointes et mesures de gestion des pêcheries mixtes: application à la pêcherie langoustinière du golfe de Gascogne*. Université de Bretagne occidentale - Brest. Disponible à l'adresse : <https://tel.archives-ouvertes.fr/tel-00361589>
- MACQUEEN, J., 1967. Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1, pp. 281-297.

- MARCHAL, P., 2008. A comparative analysis of métiers and catch profiles for some French demersal and pelagic fleets. *ICES Journal of Marine Science*. 27 février 2008. Vol. 65, n° 4, pp. 674-686. DOI [10.1093/icesjms/fsn044](https://doi.org/10.1093/icesjms/fsn044).
- MORTENSEN, Lars O., ULRICH, Clara, HANSEN, Jan et HALD, Rasmus, 2018. Identifying choke species challenges for an individual demersal trawler in the North Sea, lessons from conversations and data analysis. *Marine Policy*. Janvier 2018. Vol. 87, pp. 1-11. DOI [10.1016/j.marpol.2017.09.031](https://doi.org/10.1016/j.marpol.2017.09.031).
- OLASO, I, VELASCO, F et PÉREZ, N, 1998. Importance of discarded blue whiting (*Micromesistius poutassou*) in the diet of lesser spotted dogfish (*Scylliorhinus canicula*) in the Cantabrian Sea. *ICES Journal of Marine Science*. juin 1998. Vol. 55, n° 3, pp. 331-341. DOI [10.1006/jmsc.1997.0263](https://doi.org/10.1006/jmsc.1997.0263).
- PELLETIER, Dominique et FERRARIS, Jocelyne, 2000. A multivariate approach for defining fishing tactics from commercial catch and effort data. *Canadian Journal of Fisheries and Aquatic Sciences*. 2000. Vol. 57, pp. 51-65.
- PRELLEZO, Raúl, CARMONA, Itsaso et GARCÍA, Dorleta, 2016. The bad, the good and the very good of the landing obligation implementation in the Bay of Biscay: A case study of Basque trawlers. *Fisheries Research*. septembre 2016. Vol. 181, pp. 172-185. DOI [10.1016/j.fishres.2016.04.016](https://doi.org/10.1016/j.fishres.2016.04.016).
- RINDORF, Anna, MUMFORD, John, BARANOWSKI, Paul, CLAUSEN, Lotte Worsøe, GARCÍA, Dorleta, HINTZEN, Niels T., KEMPF, Alexander, LEACH, Adrian, LEVONTIN, Polina, MACE, Pamela, MACKINSON, Steven, MARAVELIAS, Christos, PRELLEZO, Raúl, QUETGLAS, Antoni, TSERPES, George, VOSS, Rüdiger et REID, David, 2017. Moving beyond the MSY concept to reflect multidimensional fisheries management objectives. *Marine Policy*. novembre 2017. Vol. 85, pp. 33-41. DOI [10.1016/j.marpol.2017.08.012](https://doi.org/10.1016/j.marpol.2017.08.012).
- Scientific, Technical and Economic Committee for Fisheries (STECF), 2016. *The 2016 Annual Economic Report on the EU Fishing Fleet* (STECF 16-11); Publications Office of the European Union, Luxembourg, 472. ISBN 978-92-79-64633-1. DOI [10.2788/842673](https://doi.org/10.2788/842673)
- Système d'Informations Halieutiques (SIH), 2017a. *Manuel de l'Observateur à bord des navires de pêche commerciaux*. juillet 2017.
- Système d'Informations Halieutiques (SIH), 2017b. *Embarquez un observateur Obsmer !* août 2017.
- THERNEAU, Terry et ATKINSON, Beth, 2018. *rpart: Recursive Partitioning and Regression Trees*. Disponible à l'adresse : <https://CRAN.R-project.org/package=rpart>
- ULRICH, Clara, VERMARD, Youen, DOLDER, Paul J., BRUNEL, Thomas, JARDIM, Ernesto, HOLMES, Steven J., KEMPF, Alexander, MORTENSEN, Lars O., POOS, Jan-Jaap et RINDORF, Anna, 2016. Achieving maximum sustainable yield in mixed fisheries: a management approach for the North Sea demersal fisheries. *ICES Journal of Marine Science: Journal du Conseil*. 7 août 2016. pp. fsw126. DOI [10.1093/icesjms/fsw126](https://doi.org/10.1093/icesjms/fsw126).
- ULRICH, Clara, WILSON, Douglas C.K., NIELSEN, J. Rasmus, BASTARDIE, Francois, REEVES, Stuart A., ANDERSEN, Bo S. et EIGAARD, Ole R., 2012. Challenges and opportunities for fleet- and métier-based approaches for fisheries management under the European Common Fishery Policy. *Ocean & Coastal Management*. décembre 2012. Vol. 70, pp. 38-47. DOI [10.1016/j.ocecoaman.2012.06.002](https://doi.org/10.1016/j.ocecoaman.2012.06.002).
- VAN DER VEER, H.W., CREUTZBERG, F., DAPPER, R., DUINEVELD, G.C.A., FONDS, M., KUIPERS, B.R., VAN NOORT, G.J. et WITTE, J.I.J., 1990. On the ecology of the dragonet *Callionymus lyra* L. in the southern North Sea. *Netherlands Journal of Sea Research*. octobre 1990. Vol. 26, n° 1, pp. 139-150. DOI [10.1016/0077-7579\(90\)90063-M](https://doi.org/10.1016/0077-7579(90)90063-M).
- VENABLES, W. N. et RIPLEY, B. D., 2002. *Modern Applied Statistics with S*. Fourth edition. New York, NY : Springer New York. Statistics and Computing. ISBN 978-1-4419-3008-8.
- WARD, Joe H., 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*. mars 1963. Vol. 58, n° 301, pp. 236-244.
- WHITE, E., MINTO, C., NOLAN, C. P., KING, E., MULLINS, E. et CLARKE, M., 2011. First estimates of age, growth, and maturity of boarfish (*Capros aper*): a species newly exploited in the Northeast Atlantic. *ICES Journal of Marine Science*. 1 janvier 2011. Vol. 68, n° 1, pp. 61-66. DOI [10.1093/icesjms/fsq150](https://doi.org/10.1093/icesjms/fsq150).

Wikistat, 2016. Analyse discriminante décisionnelle. Disponible à l'adresse : <http://wikistat.fr/pdf/st-m-app-add.pdf>

Sources des données :

Direction des Pêches Maritimes et de l'Aquaculture (DPMA), 2017. Paris.
Ifremer - Système d'Informations Halieutiques (SIH), 2017. Plouzané.

Annexes

ANNEXE I. Tableau récapitulatif des méthodes de classification non supervisée

Méthode	Données entrées	Données sorties / Librairie associée	Avantages / Principe	Inconvénients
Classification Ascendante Hiérarchique (CAH)	Matrice similarité. Variables qualitatives et quantitatives (selon la distance choisi l'algorithme utilisé)	Package : Stats Sorties : Structure organisationnelle : dendrogramme Ex algorithme : Agglomératif : SLINK, ALINK, CLINK, CURE, CHAMELEON, ROCK Divisif : DIANA, PDDP	Basé sur les distances. Principe : On regroupe les clusters en minimisant quelque chose, ex : l'inertie intraclasse (critère de Ward) Avantages : Méthode reconnue. Bonne visualisation de l'organisation des données et du processus de clustering	Parfois sorties compliquées à interpréter notamment pour les algorithmes divisifs. Nécessite le calcul des distances entre individus pris deux à deux. Ce qui est très rapidement prohibitif dès que la taille du fichier excède le millier d'individus. Crée toutes les classes possibles, si on veut un nombre de classes limitées privilégier k-means.
Aggrégation autour de centres mobiles (= Algorithmes de réallocation dynamique) : k-means (Mc Queen, 1967)	Données sur répartition convexe. Données décrites par des attributs numériques (quantitatif) Le nombre de clusters désiré + une mesure de dissimilarité sur l'ensemble des objets à traiter	Package : Stats Sorties : Une partition de l'espace des objets. Le but est de comparer plusieurs partitionnements afin de retenir le schéma qui optimise un critère de qualité Algorithme K-means flou existe (Fuzzy k-means, Axial k-means)	Basé sur les distances dans un espace Euclidien. Principe : Trouve les meilleurs centroïdes en prenant les médians des barycentres. Avantages : Méthode reconnue et très utilisée, peut être une méthode qui en initie d'autres (mélange gaussiens). Permet généralement d'avoir une idée du nombre de classes « optimales »	Si on rajoute des dimensions cela peut changer le calcul lié à l'inertie. Il faut normaliser pour éviter ça. La position initial des barycentres peut avoir un impact sur les résultats il faut donc faire différents essais avec différentes positions de départ. Il faut un nombre de classes fixé par avance

<p>Aggrégation autour de centres mobiles (= Algorithmes de réallocation dynamique) : Nuées dynamiques (Diday, 1972)</p>	<p>Données sur répartition convexe. Données qualitatives acceptées. Le nombre de clusters désiré + une mesure de dissimilarité sur l'ensemble des objets à traiter</p>	<p>Sorties : Une partition de l'espace des objets. Package : FactoClass</p>	<p>Basé sur les distances. Principe : ce n'est plus un seul point qui représente une classe mais un noyau de points constitués d'éléments représentatifs de la classe. Avantages : Cela permet de corriger l'influence d'éventuelles valeurs extrêmes sur le calcul du barycentre</p>	<p>Pareil K-means</p>
<p>Aggrégation autour de centres mobiles (= Algorithmes de réallocation dynamique) : Partitionning Around Medoïds (PAM)</p>	<p>Données sur répartition convexe. Données qualitatives acceptées. Le nombre de clusters désiré + une mesure de dissimilarité sur l'ensemble des objets à traiter</p>	<p>Sorties : Partition finale avec affectation de tous les objets aux médoïdes issus du meilleur schéma généré Package : ClusterR ou fonction pam Ex d'algorithme : CLARA (Clustering LARge Applications), CLARANS (CLustering Algorithm based on RANdomized Search) Algorithme K-médoïdes flou existe (Fuzzy c-Medoids)</p>	<p>Basé sur les distances dans un espace Non Euclidien. Principe : Proche des k-means. Utilise des médoïdes plutôt que des centroïdes. Avantages : Besoin uniquement de matrice de distance</p>	<p>Pour les larges ensembles de données car la recherche de médoïdes est plus coûteuse que celle de centroïdes</p>
<p>Le clustering par densité : Density-based spatial clustering of</p>	<p>Travail sur des données ayant une forme non convexe.</p>	<p>Sorties : Une partition de X en k d-clusters. Package : DBScan</p>	<p>Basé sur les distances non Euclidienne et la densité des clusters. Principe : Si 2</p>	<p>Ne va pas bien fonctionner si les densités dans les clusters sont très différentes. Si on a un espace</p>

<p>applications with noise (DBSCAN)</p>	<p>Besoin de données minimales pour constituer un cluster. Un ensemble X de n objets, un rayon et M (minimum d'un nombre d'objet) sont les paramètres à définir.</p>	<p>Exemple d'algorithme alternatif : DBCLASD, GDBSCAN et DENCLUE (partitionnement flou). OPTICS (en complément à DBSCAN, définit un ordre aux objets)</p>	<p>points sont suffisamment proches ils sont associés à la même classe. Même principe k-means sauf que les objets centraux sont des noyaux par lesquels les autres objets sont « d-accessibles » Avantages : permet d'écarter les points suffisamment éloignés de tous les clusters (≠ k-means qui classe tous les points). Juste besoin de définir une distance minimale pour créer une densité et définir un nombre minimal que l'on autorise dans un cluster.</p>	<p>avec des données très proches et certaines éloignées les unes des autres. La topologie de l'espace joue énormément sur ce type d'algorithme. Le choix de paramètre du rayon et M (minimum d'un nombre d'objet) difficile à fixer</p>
<p>Mélange gaussiens</p>	<p>Données doivent avoir une répartition sous forme de loi Gaussienne. On rentre l'ensemble des observations, le nombre de clusters et un seuil de tolérance. Peut prendre en compte des données qualitatives et quantitatives.</p>	<p>Sortie : Ensemble de probabilités d'appartenance à une classe Package : mclust Ex d'algorithme : EM (Expectation Maximization)</p>	<p>Basé sur la loi de Bayes. Principe : Probabilité d'appartenir à une loi probabiliste. Avantages : Evite les limites des approches avec les méthodes utilisant les distances. Proche des k-means mais prend mieux en compte une topologie sous-jacente que l'on pourrait avoir.</p>	<p>Besoin de données qui s'y prête. Sa lenteur due au nombre élevé d'itérations nécessaire à la convergence. Coût important.</p>
<p>Spectral clustering</p>			<p>Principe : Réduction dimensionnelle du problème, enlever des dimensions qui font du bruit et faire un k-means derrière</p>	

Clustering par grille	Vise principalement les données spatiales	<p>Sortie : Partition des données via une partition de l'espace de représentation. Clusters correspondent à un ensemble de cellules denses et connectées.</p> <p>Package : subspace</p> <p>Ex d'algorithme : STING, WaveCluster, GIZMO et CLIQUE</p>	<p>Principe : Construction d'une structure hiérarchique (découpage de l'espace) et sélection d'un ensemble de cellules répondants à des caractéristiques exprimées sous forme d'une requête.</p> <p>Avantages : Peut traiter de très grandes bases de données</p>	Faibles performances dans le cas de données décrites dans un espace à grande dimension (car augmente le nombre de cellules). Privilégier des données à faibles dimensions sinon utiliser l'algorithme CLIQUE (CLustering In QUEst) qui traite en sous problème dans des espaces de dimension plus faible.
Clustering conceptuel	Données décrites par des attributs numériques (quantitatifs) et qualitatifs	<p>Sortie : Une hiérarchie de concepts classifiant l'instance x_i</p> <p>Package : Rweka, cf Weka clusterers</p> <p>Ex algorithmme : COBWEB, CLUSTER/S</p>	<p>Principe : Plutôt que de définir une mesure de similarité l'idée est de générer une structure (hiérarchique) de concepts. Avantages : CODWEB est très utilisé et est considéré comme une référence dans le domaine.</p>	Difficulté à traiter les attributs numériques, l'influence de l'ordre dans lequel les objets sont intégrés à la hiérarchie ou encore le stockage coûteux de l'ensemble des instances dans cette hiérarchie.
Clustering avec « neurones artificiels », Cartes auto-organisatrices	Vecteurs numériques uniquement, de nombreux paramètres à définir (nombre d'itérations, voisinages d'un nœud vainqueur,...)	<p>Sortie : Classes visualisable sur un plan 2D</p> <p>Packages : SOMbrero, yasomi</p> <p>Ex algorithmme : SOM (Self-Organizing Map), cartes auto-organisatrice de Kohonen</p>	<p>Principe : un réseau monocouche, où chaque objet d'entrée (neurone d'entrée) est associé à un neurone de sortie par projection. Ces neurones de sortie sont en fait les représentants des classes, en nombre fixé <i>a priori</i>, et organisés généralement sur une grille 2D. Cet algorithme procède de façon incrémentale : chaque vecteur d'entrée engendre une compétition entre</p>	Les données d'entrées ainsi que la dépendance du résultat avec l'initialisation du modèle.

			neurones pour déterminer le neurone vainqueur (le plus proche), induisant alors la mise à jour des poids synaptiques du vainqueur et de son environnement proche. Avantages : Visualisation du résultat dans un plan	
Clustering basé sur les graphes	Différents graphiques (Nearest Neighbor Graph, Minimum Spanning Tree, Relative Neighborhood Graph, Gabriel Graph et Delaunay Triangulation)		Principe : A partir des graphes d'entrée, on cherche à diviser l'ensemble des sommets du graphe en plusieurs "groupes naturels" par rapport aux relations entre objets, caractérisées par les arêtes.	La complexité d'une telle méthode augmente de façon exponentielle avec le paramètre k. Il est nécessaire de traiter des volumes importants de données pour cette méthode.

Différentes bibliographies nécessaires à la réalisation du tableau :

- BENDHAIBA, Laura, 2013. *Création d'un package R pour des cartes auto-organisatrices*. Stage de fin d'études. Polytech'Lille.
- CLEUZIOU, Guillaume, 2004. *Une méthode de classification non-supervisée pour l'apprentissage de règles et la recherche d'information*. Université d'Orléans. Disponible à l'adresse : <https://tel.archives-ouvertes.fr/tel-00084828>
- DIDAY, E., 1973. The dynamic clusters method in nonhierarchical clustering. *International Journal of Computer & Information Sciences*. mars 1973. Vol. 2, n° 1, pp. 61-88. DOI [10.1007/BF00987153](https://doi.org/10.1007/BF00987153).
- LEBARBIER, E et MARY-HUARD, T, 2008. *Classification non supervisée* [en ligne]. AgroParisTech. Disponible à l'adresse : <http://www2.agroparistech.fr/IMG/pdf/ClassificationNonSupervisee-AgroParisTech.pdf>
- SUBLEMONTIER, Jacques-Henri, 2012. *Classification non supervisée: de la multiplicité des données à la multiplicité des analyses* [en ligne]. Université d'Orléans. Disponible à l'adresse : <https://tel.archives-ouvertes.fr/tel-00801555v1>

ANNEXE II. Descriptions des variables quantitatives

A l'échelle de l'opération de pêche

Variable	Formule
Contribution de l'espèce dans les captures	<p data-bbox="786 496 1787 544">$C_{i,j} = \text{poids des captures (Espèce}_i, \text{Opération de pêche}_j)$</p> $c_{i,j} = \frac{C_{i,j}}{\sum_{i=1}^n C_{i,j}} \quad \text{avec} \quad \sum_i c_{i,j} = 1 \quad \forall_j \quad (1)$
Fraction rejetée	<p data-bbox="797 940 1720 987">$R_{i,j} = \text{poids rejeté (Espèce}_i, \text{Opération de pêche}_j)$</p> <p data-bbox="763 1003 1760 1051">$D_{i,j} = \text{poids débarqué (Espèce}_i, \text{Opération de pêche}_j)$</p> $RC_{i,j} = \frac{R_{i,j}}{C_{i,j}} \quad \text{avec} \quad C_{i,j} = R_{i,j} + D_{i,j} \quad (2)$

<p>Fraction rejetée supérieure ou égale à la Taille Minimale de Débarquement (TMD)</p>	$RT_{i,j} = \text{poids rejeté (Espèce}_i, \text{Opération de pêche}_j) \geq TMD (\text{Espèce}_i)$ $DT_{i,j} = \text{poids débarqué (Espèce}_i, \text{Opération de pêche}_j) \geq TMD (\text{Espèce}_i)$ $RCT_{i,j} = \frac{RT_{i,j}}{CT_{i,j}} \quad \text{avec} \quad CT_{i,j} = RT_{i,j} + DT_{i,j}$ <p style="text-align: right;">(3)</p>
<p>Fraction de rejet supérieure ou égale à la taille minimale de débarquement (TMD)</p>	$R_{i,j} = \text{poids rejeté (Espèce}_i, \text{Opération de pêche}_j)$ $RT_{i,j} = \text{poids rejeté (Espèce}_i, \text{Opération de pêche}_j) \geq TMD (\text{Espèce}_i)$ $RRT_{i,j} = \frac{RT_{i,j}}{R_{i,j}}$ <p style="text-align: right;">(4)</p>
<p>Valeur</p>	$D_{i,j,k} = \text{poids débarqué (Espèce}_i, \text{Opération de pêche}_j, \text{Mois}_k)$ $P_{i,k} = \text{prix/kg (Espèce}_i, \text{Mois}_k)$ $V_{i,j,k} = D_{i,j,k} * P_{i,k}$ <p style="text-align: right;">(5)</p>

<p>Fréquence d'occurrence des espèces dans les opérations de pêche échantillonnées</p>	<p>$NE_m =$ Nombre d'opérations de pêche échantillonnées dans la marée$_m$ $JE_m =$ Opération de pêche échantillonnées$_j$ dans la marée$_m$ $C_i =$ poids des captures (Espèce$_i$)</p> $ne_{m,i} = \frac{NE_{m,i}}{NE_m} \quad \text{avec} \quad NE_{m,i} = \sum_{j=1}^{JE_m} \mathbb{1}_{[C_i > 0]}^{(j)}$ <p style="text-align: right;">(6)</p>
<p>Fréquence d'occurrence des espèces débarquées dans l'ensemble des opérations de pêche</p>	<p>$N_m =$ Nombre d'opérations de pêche dans la marée$_m$ $J_m =$ Opération de pêche$_j$ dans la marée$_m$ $D_i =$ poids des débarquements (Espèce$_i$)</p> $n_{m,i} = \frac{N_{m,i}}{N_m} \quad \text{avec} \quad N_{m,i} = \sum_{j=1}^{J_m} \mathbb{1}_{[D_i > 0]}^{(j)}$ <p style="text-align: right;">(7)</p>

A l'échelle de la séquence de pêche

Rappel, les fractions rejetées sont estimées suite à une élévation puis un poids rejeté estimé est calculé (formule dans le rapport, p.12)

Variable	Formule
Contribution de l'espèce dans les captures estimées	<p data-bbox="616 454 1937 518">$\hat{C}_{i,l} = \text{poids des captures estimé (Espèce}_i, \text{Séquence de pêche}_l)$</p> <p data-bbox="1052 582 1456 662">$\hat{C}_{i,l} = \hat{R}_{i,l} + D_{i,l}$</p> <p data-bbox="761 742 2027 1013"> $\hat{c}_{i,l} = \frac{\hat{C}_{i,l}}{\sum_{i=1}^n \hat{C}_{i,l}} \quad \text{avec} \quad \sum_i \hat{c}_{i,l} = 1 \quad \forall_l \quad (8)$ </p>
Valeur	<p data-bbox="627 1093 1926 1157">$D_{i,l,k} = \text{poids débarqué (Espèce}_i, \text{Séquence de pêche}_l, \text{Mois}_k)$</p> <p data-bbox="907 1189 2027 1252">$P_{i,k} = \text{prix/kg (Espèce}_i, \text{Mois}_k) \quad (9)$</p> <p data-bbox="1008 1292 1500 1364">$V_{i,l,k} = D_{i,l,k} * P_{i,k}$</p>

Fréquence
d'occurrence des
espèces capturées
estimées

$NL_m =$ Nombre de séquence de pêche dans la marée $_m$

$L_m =$ Séquence de pêche $_l$ dans la marée $_m$

$\hat{C}_i =$ poids des captures estimé (Espèce $_i$)

(10)

$$nl_{m,i} = \frac{NL_{m,i}}{N_m} \quad \text{avec} \quad NL_{m,i} = \sum_{l=1}^{L_m} \mathbb{1}_{[\hat{C}_i > 0]}^{(l)}$$

ANNEXE III. Résultats complémentaires

Table des matières annexe III

A. Comparaison des variables de rejets à l'échelle de l'opération de pêche	47
B. Boxplot des variables à l'échelle de l'opération de pêche	48
C. Analyse en Composante Principale à l'échelle de l'opération de pêche	49
D. Description des clusters issus de la Classification Ascendante Hiérarchique à l'échelle de l'opération de pêche	51
E. Description des clusters et appartenance aux classes par espèces issus des kmeans à l'échelle de l'opération de pêche	52
F. Analyse prenant en compte les rejets à l'échelle de la séquence de pêche	53
G. Description des clusters issus des kmeans à l'échelle de la séquence de pêche ...	57
H. Analyse du modèle CART d'apprentissage sur la séquence de pêche sans rejets .	58
I. Analyse du modèle CART d'apprentissage sur SACROIS	59
J. Tableau récapitulatif des classes d'espèces issues de SACROIS	63

a. Comparaison des variables de rejets à l'échelle de l'opération de pêche

Une représentation des variables de rejets fut réalisée, pour l'ensemble des espèces étudiées, aucune différence significative n'est observée (Figure annexe 1).

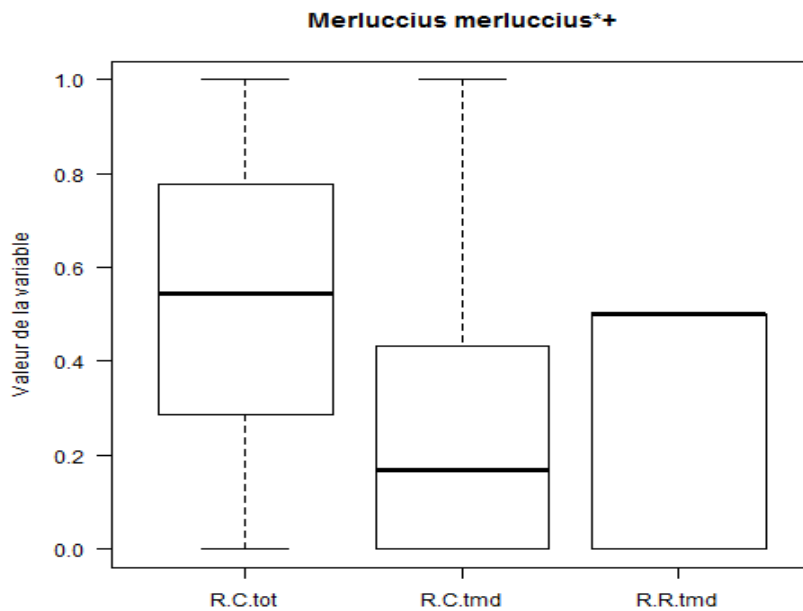


Figure annexe 1. Boxplots des 3 variables de rejets pour le merlu (Merluccius merluccius)

b. Boxplot des variables à l'échelle de l'opération de pêche

Les différentes variables : contribution de l'espèce dans les captures (Comp_cap), les rejets (R.C.tot, R.C.tmd, R.R.tmd), fréquence (FreqOPech, FreqtteOP), valeur, profondeur, durée de pêche (DUREE_OP) et vitesse de l'engin (GEAR_SPEED) sont représentés sous forme de boxplot afin de visualiser leurs variabilités.

L'analyse de ces boxplots justifie la transformation des quatre dernières variables (Valeur, profondeur, durée de l'OP et vitesse de l'engin, Figure annexe 2). Elles sont centrées réduites afin de réduire leurs variabilités (Figure annexe 3).

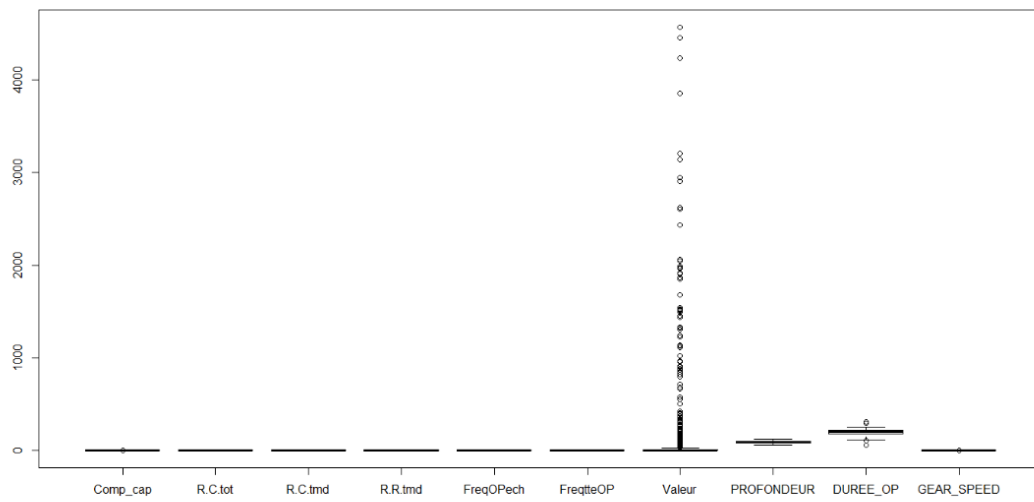


Figure annexe 2. Boxplots des variables brutes à l'échelle de l'opération de pêche

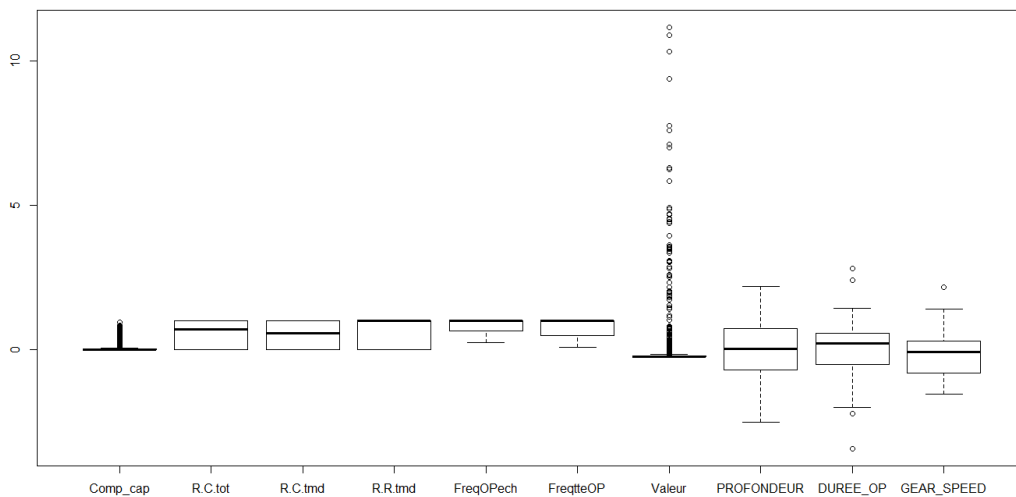


Figure annexe 3. Boxplots des variables centrées réduites à l'échelle de l'opération de pêche

c. Analyse en Composante Principale à l'échelle de l'opération de pêche

Les graphiques ci-dessus montre que pour expliquer plus de 60 % de la variabilité totale, il est nécessaire d'étudier les 3 première dimensions de l'Analyse en Composante Principale, expliquant respectivement 29.6 %, 21.6 % et 13.9 % de la variabilité (Figure annexe 4).

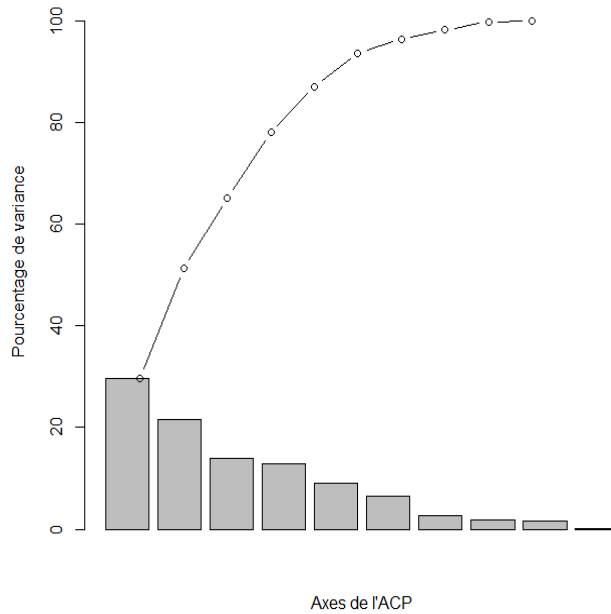


Figure annexe 4. Graphique représentant la variabilité expliquée en fonction de la dimension de l'ACP.

Les graphiques des résultats de l'ACP montre que les 3 variables de rejets sont fortement corrélées à l'axe 1. Les variables de fréquence, composition de capture et valeur sont elles corrélées à l'axe 2 (Figure annexe 5).

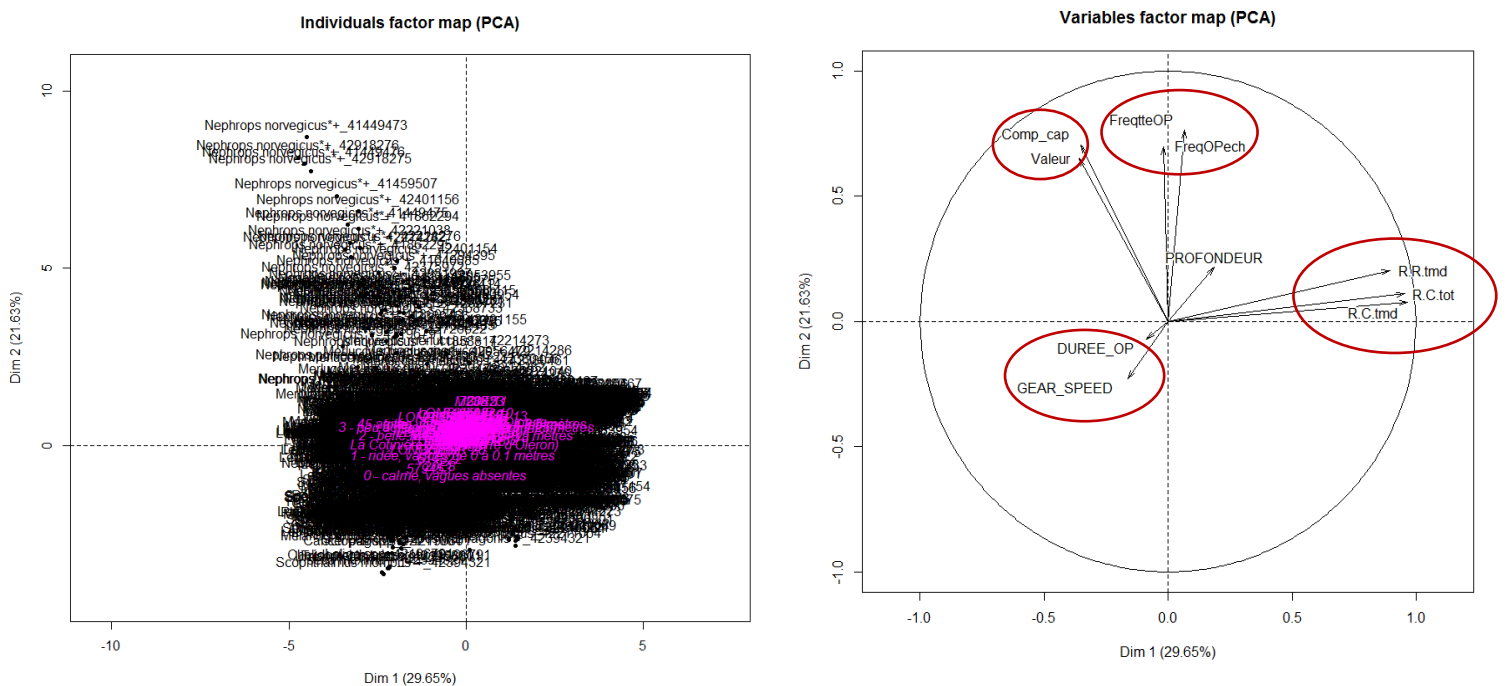


Figure annexe 5. Graphiques des individus et des variables issues des résultats de l'ACP sur les axes 1 et 2.

Il n'y a pas de réelles différences qui sont observés pour les variables qualitatives étudiées. Il semblerait que les rejets ne soient pas lié à une zone ou à un métier (Figure annexe 6).

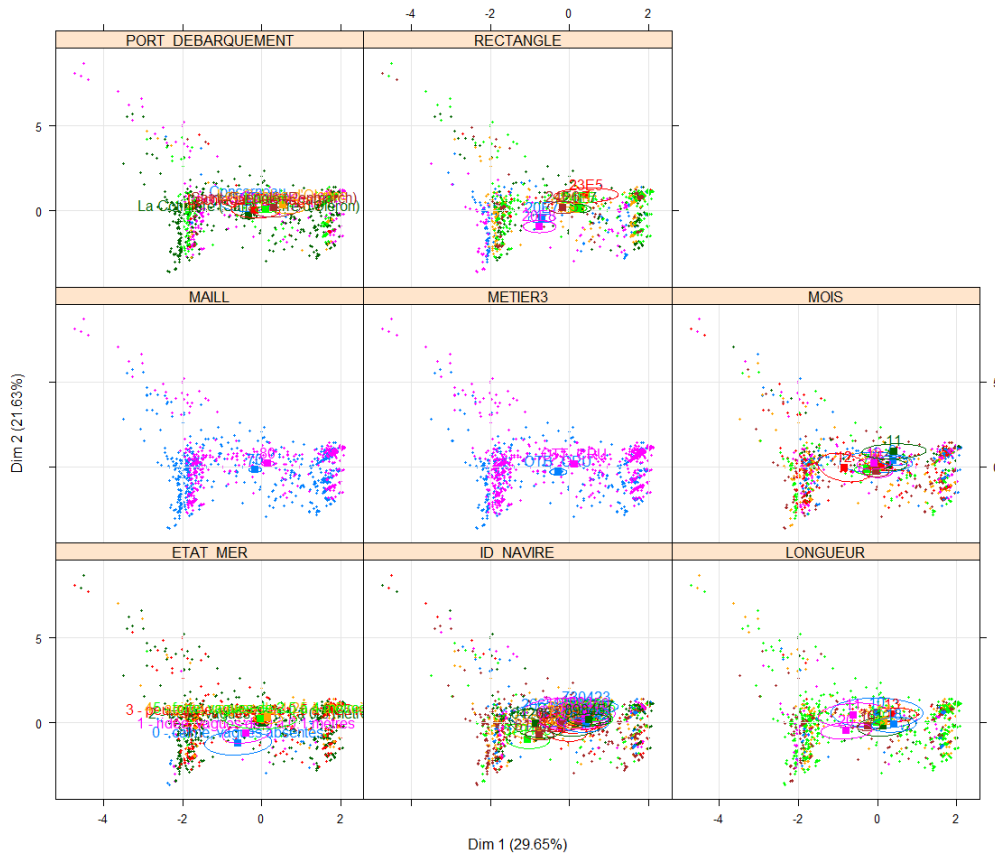


Figure annexe 6. Graphiques des individus issus des résultats de l'ACP sur les axes 1 et 2 en fonction des différentes variables qualitatives supplémentaires.

Les autres variables lié à « l'activité de pêche » sont le mieux représenté sur l'axe 3, la profondeur est anti corrélée à la vitesse de l'engin et la durée de l'opération de pêche (Figure annexe 7).

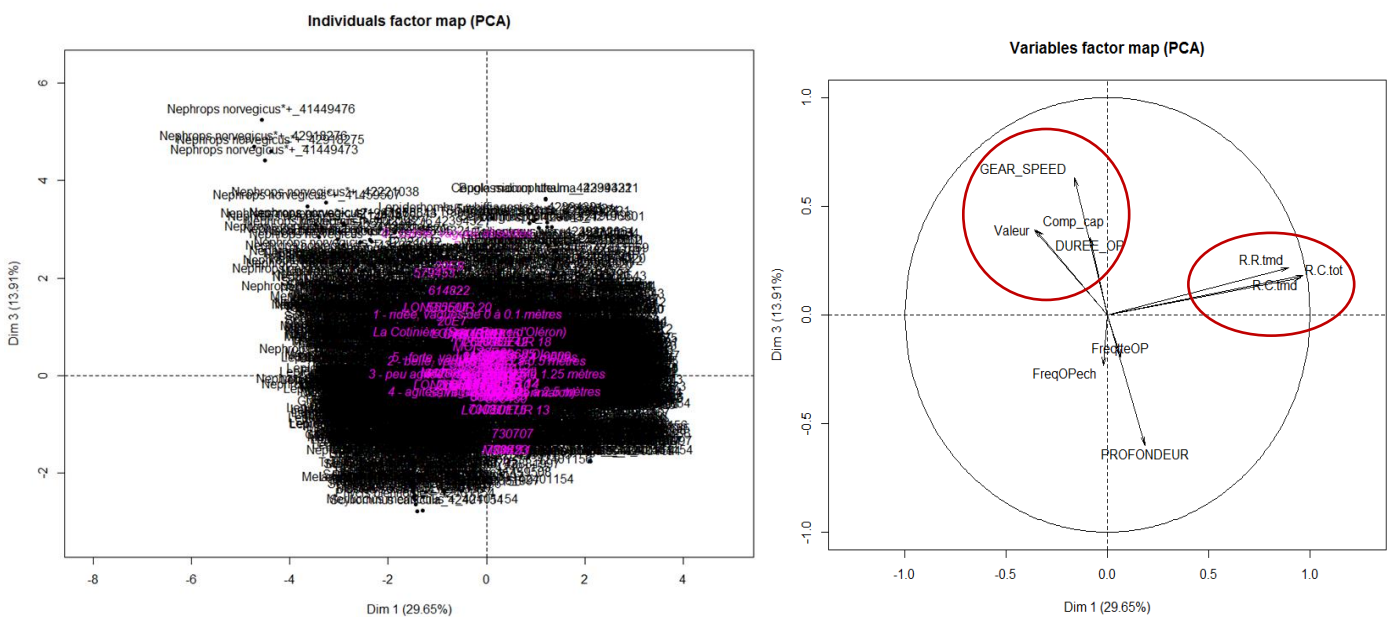


Figure annexe 7. Graphiques des individus et des variables issus des résultats de l'ACP sur les axes 1 et 3.

d. Description des clusters issus de la Classification Ascendante Hiérarchique à l'échelle de l'opération de pêche

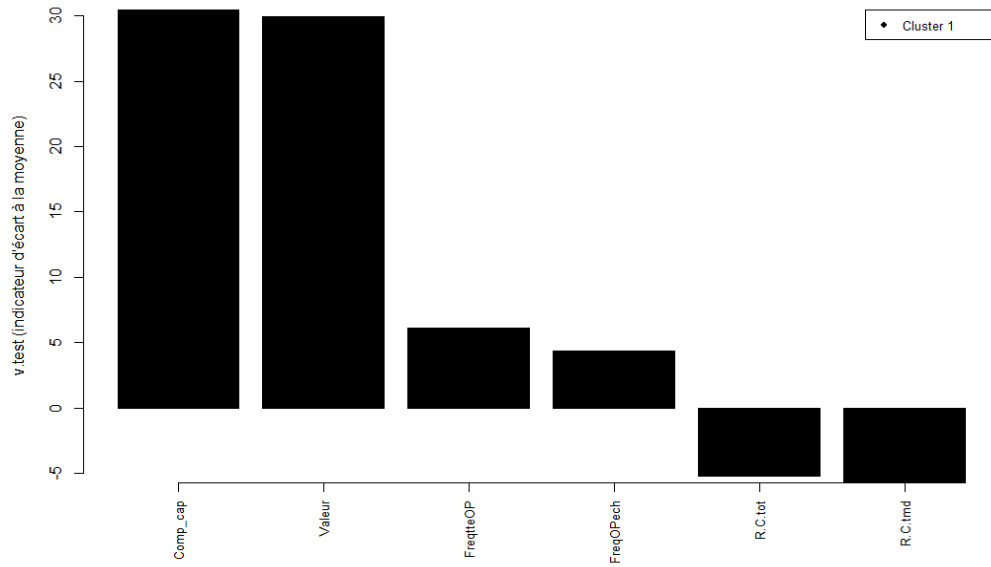


Figure annexe 8. Graphique des valeurs test en fonction des variables significatives pour le cluster 1.

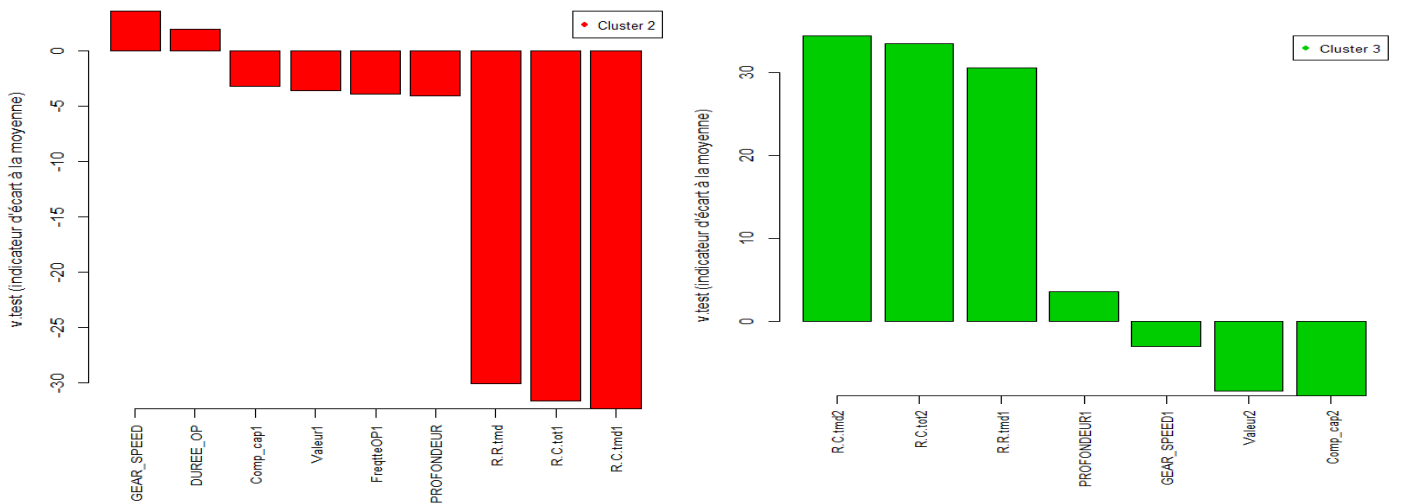


Figure annexe 9. Graphique des valeurs test en fonction des variables significatives pour le cluster 2 (à gauche) et pour le cluster 3 (à droite).

e. Description des clusters et appartenance aux classes par espèces issus des kmeans à l'échelle de l'opération de pêche

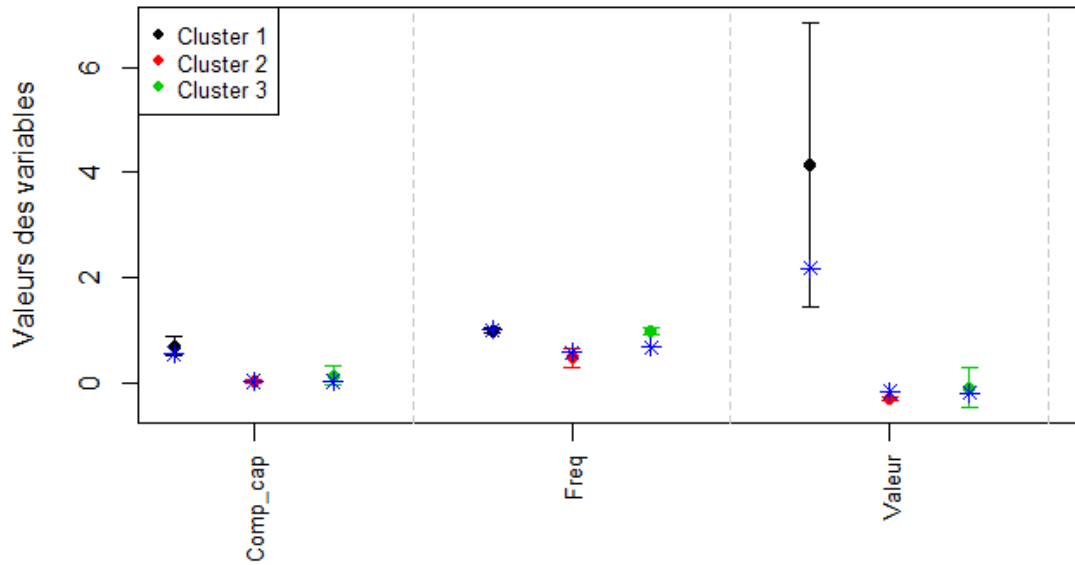


Figure annexe 10. Graphique représentant les moyennes et les écarts types de chaque variable pour les 3 clusters. Les étoiles représentent les valeurs des centroïdes initiaux.

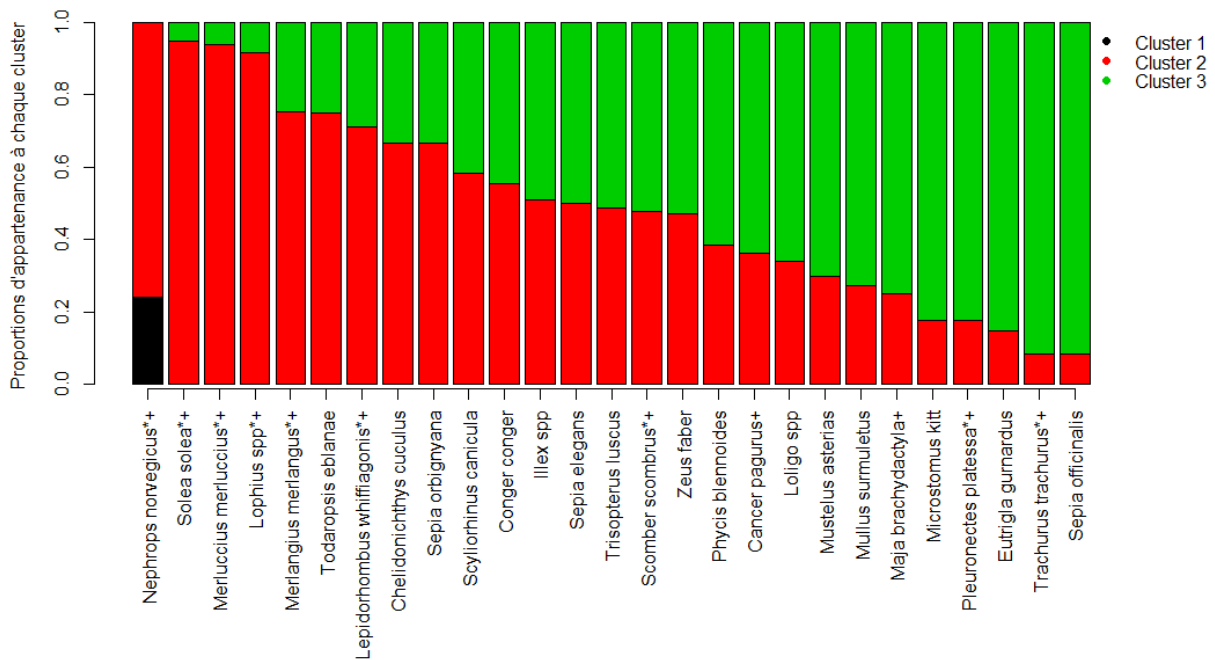


Figure annexe 11. Barplot représentant les proportions d'appartenance aux différents clusters des individus statistiques d'une même espèce.

Les espèces ayant une proportion égale à 1 ne sont pas représentées.

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD.

Correspondance des couleurs des clusters : Noir = cible, rouge = valorisé, vert = non valorisé.

f. Analyse prenant en compte les rejets à l'échelle de la séquence de pêche

- Variabilité des variables

Suite à l'élévation, les différentes variables : contribution de l'espèce dans les captures (Compo_cap), les rejets (REJ.CAPT.SEQ, REJ.CAPT.TMD.SEQ, REJ.REJ.TMD.SEQ), fréquence (FreqSEQ), valeur (PRIXQUANTSEQ) et la durée de pêche (DUREE_SEQ) sont représentés sous forme de boxplot afin de visualiser leurs variabilités (Figure annexe 12).

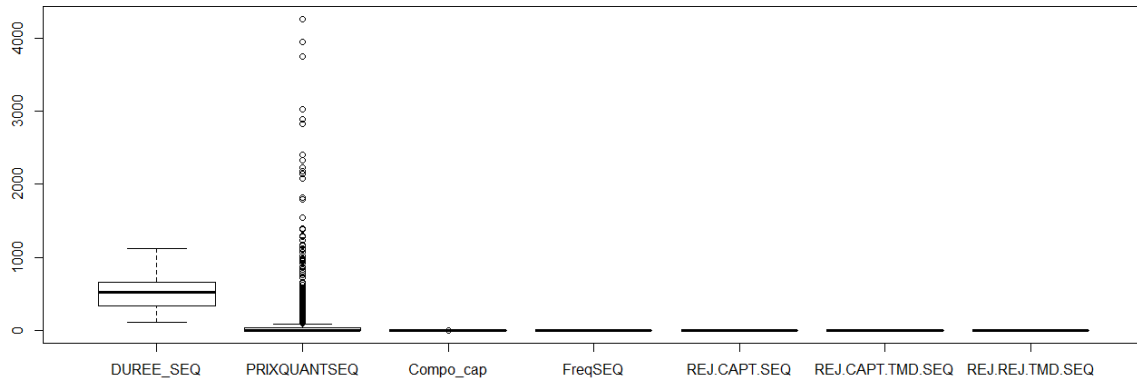


Figure annexe 12. Boxplots des variables brutes à l'échelle de la séquence de pêche

En analogie avec l'échelle d'opération de pêche, deux variables (la valeur et la durée de pêche) sont centrées réduites afin de réduire leurs variabilités (Figure annexe 13).

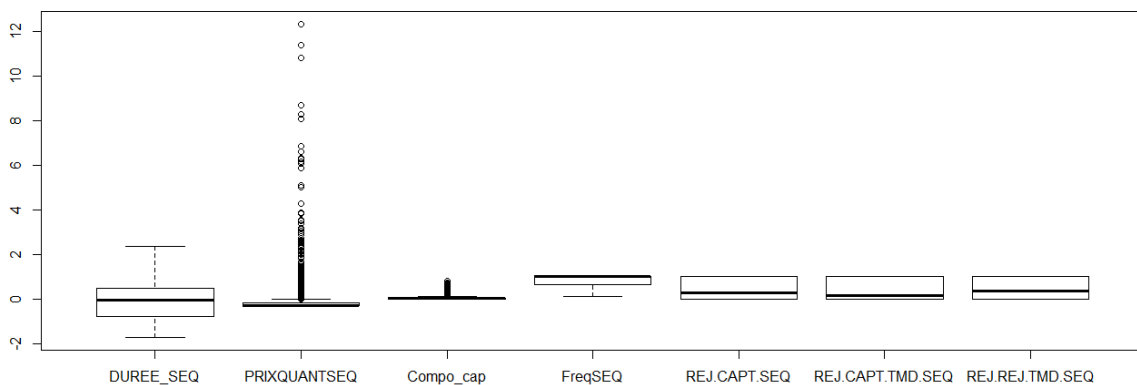


Figure annexe 13. Boxplots des variables centrées réduites à l'échelle de la séquence de pêche

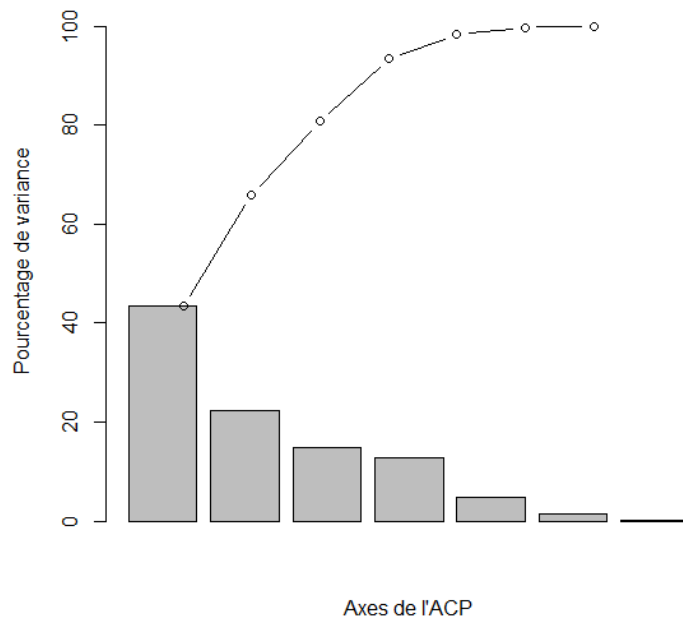


Figure annexe 14. Graphique représentant la variabilité expliquée en fonction de la dimension de l'ACP.

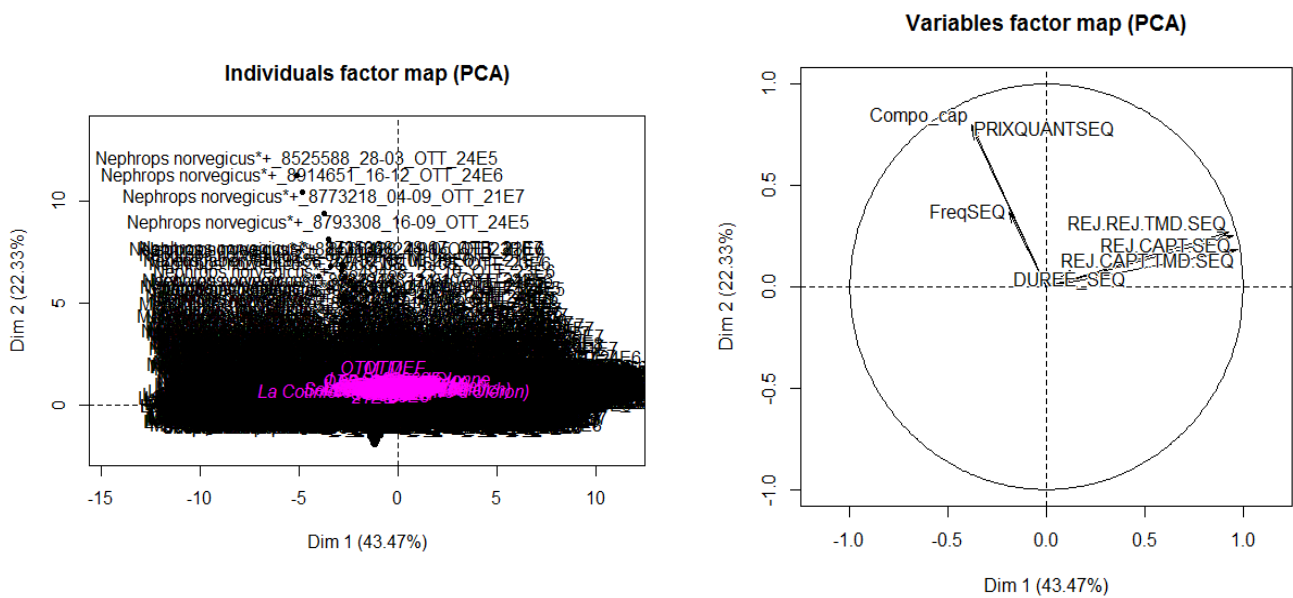


Figure annexe 15. Graphiques des individus et des variables issues des résultats de l'ACP sur les axes 1 et 2.

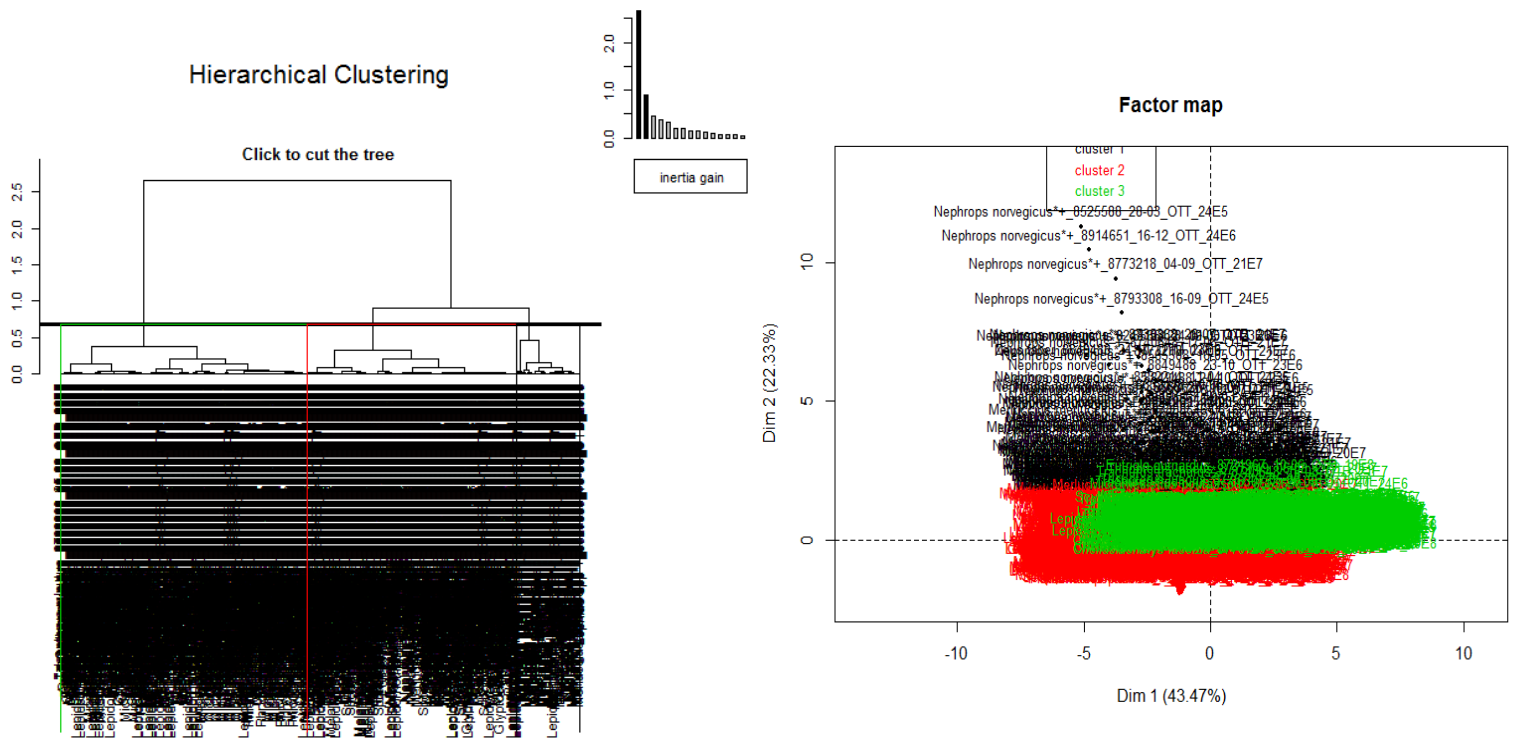


Figure annexe 16. Graphique des individus de la Classification Ascendante Hiérarchique obtenue après ACP. Le niveau de séparation permet d'obtenir 3 clusters (à gauche). Graphique des individus sur les dimensions 1 et 2, avec les 3 clusters issus de la CAH (à droite).

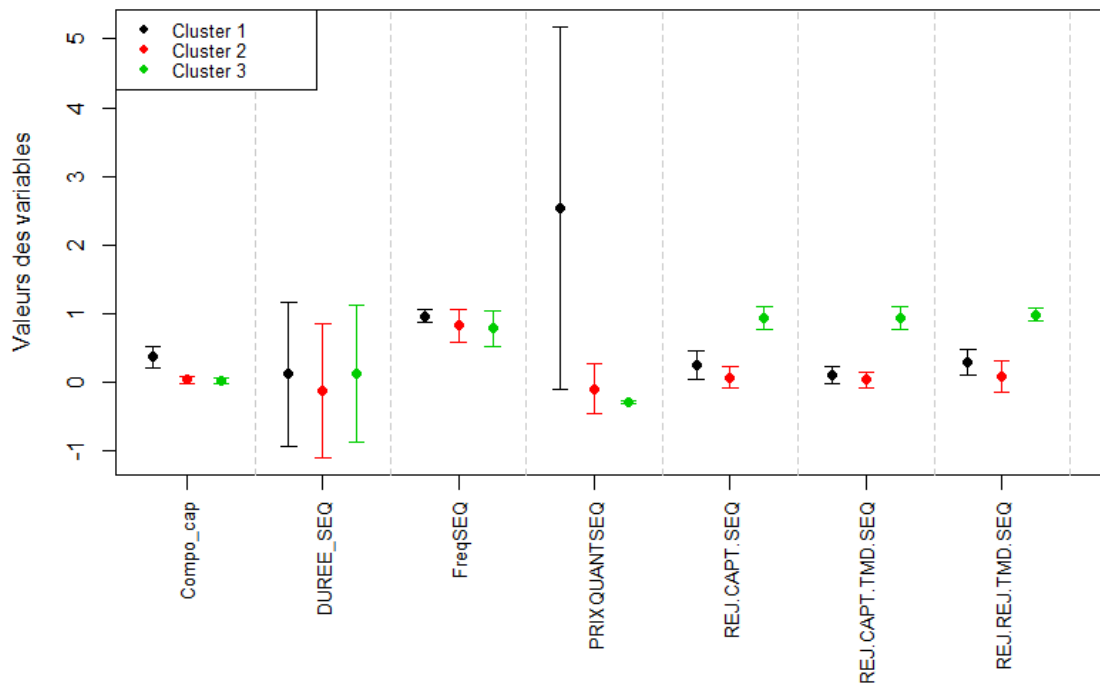


Figure annexe 17. Graphique représentant les moyennes et les écarts types de chaque variable pour les 3 clusters.

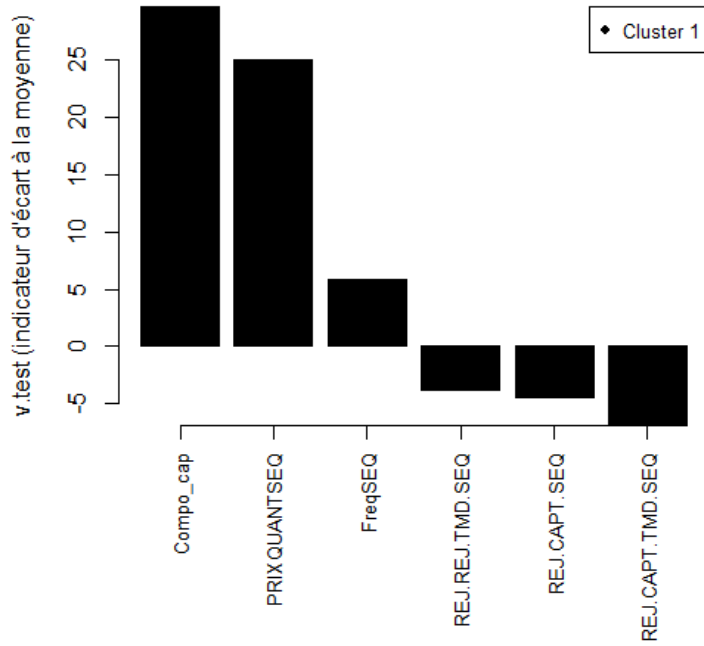


Figure annexe 18. Graphique des valeurs test en fonction des variables significatives pour le cluster 1.

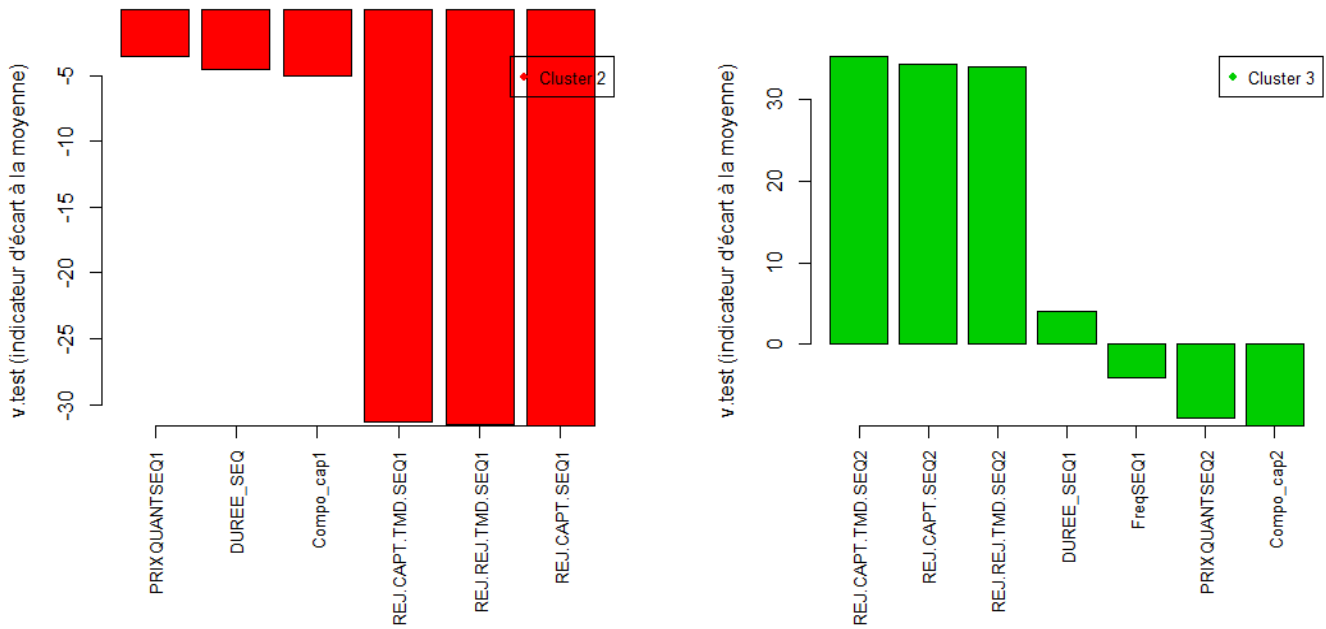


Figure annexe 19. Graphique des valeurs test en fonction des variables significatives pour le cluster 2 (à gauche) et pour le cluster 3(à droite).

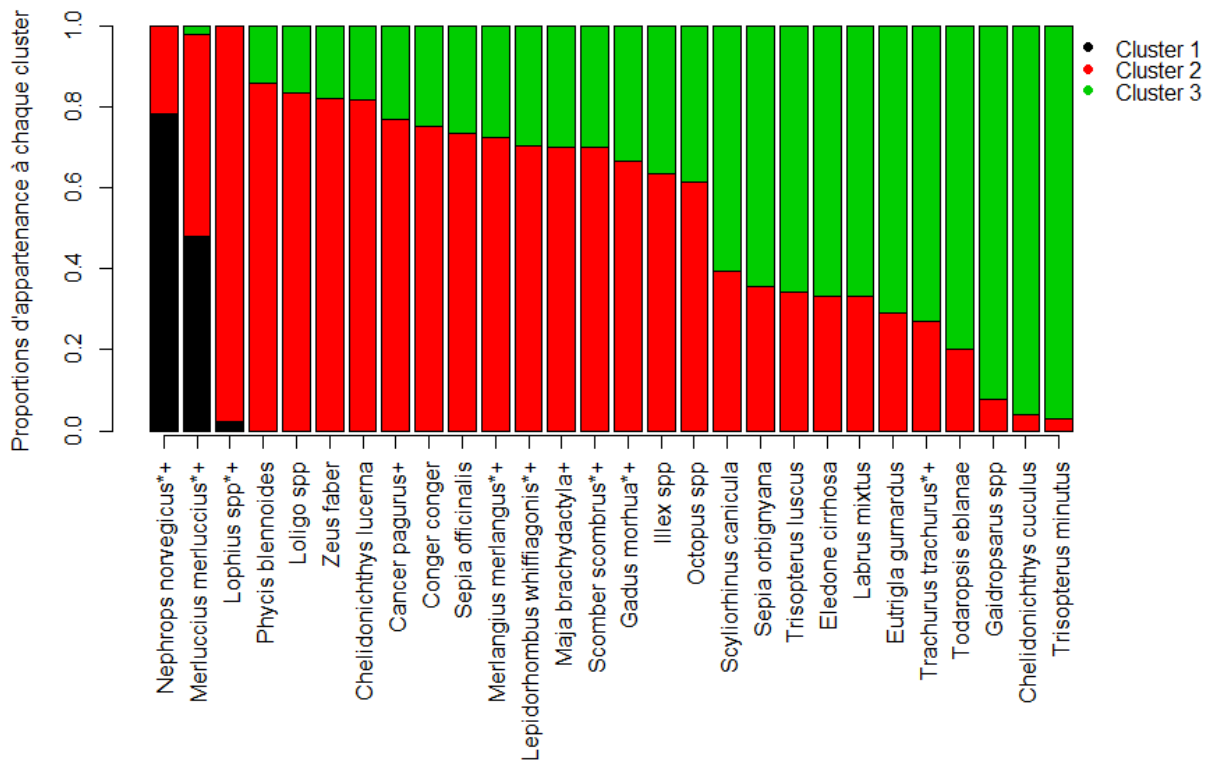


Figure annexe 20. Barplot représentant les proportions d'appartenances aux différents clusters des individus statistiques d'une même espèce.

Les espèces ayant une proportion égale à 1 ne sont pas représentées.

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD.

Correspondance des couleurs des clusters : Noir = cible, rouge = valorisé, vert = non valorisé.

g. Description des clusters issus des kmeans à l'échelle de la séquence de pêche

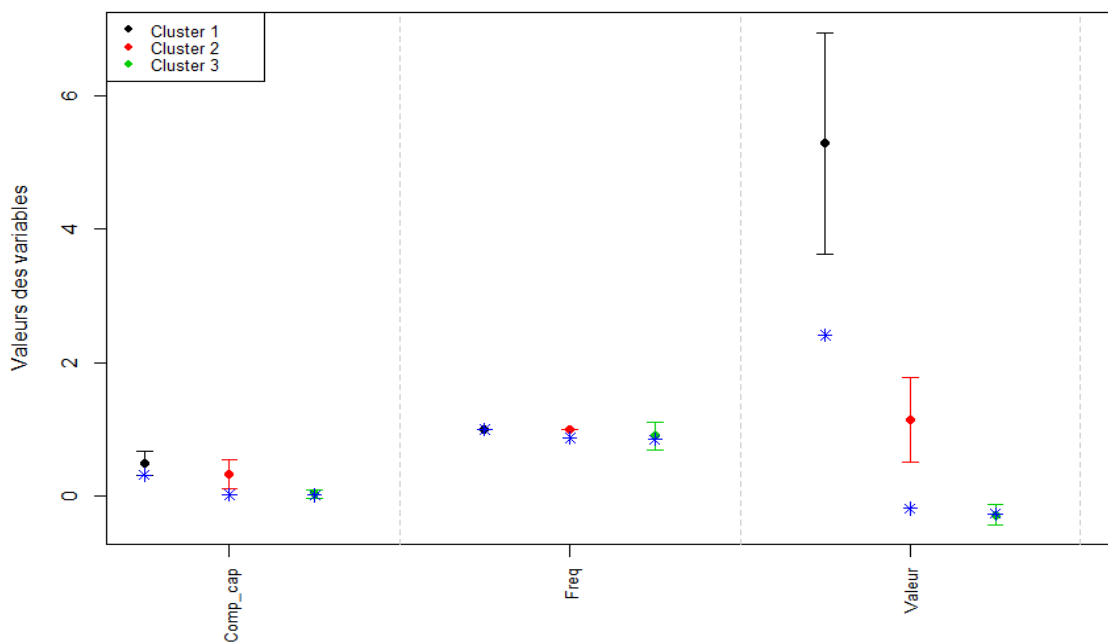


Figure annexe 21. Graphique représentant les moyennes et les écarts types de chaque variable pour les 3 clusters. Les étoiles représentent les valeurs des centroïdes initiaux.

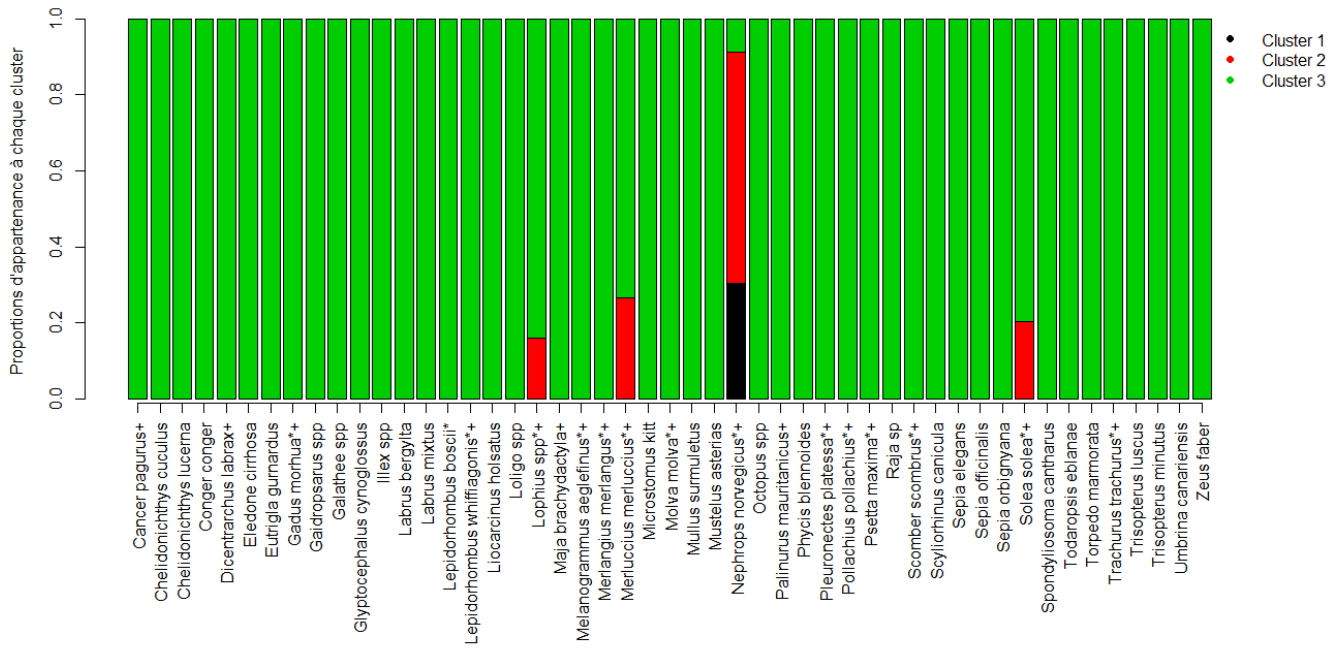


Figure annexe 22. Barplot représentant les proportions d'appartenances aux différents clusters des individus statistiques d'une même espèce.

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD. Correspondance des couleurs des clusters : Noir = cible, rouge = valorisé, vert = non valorisé.

h. Analyse du modèle CART d'apprentissage sur la séquence de pêche sans rejets

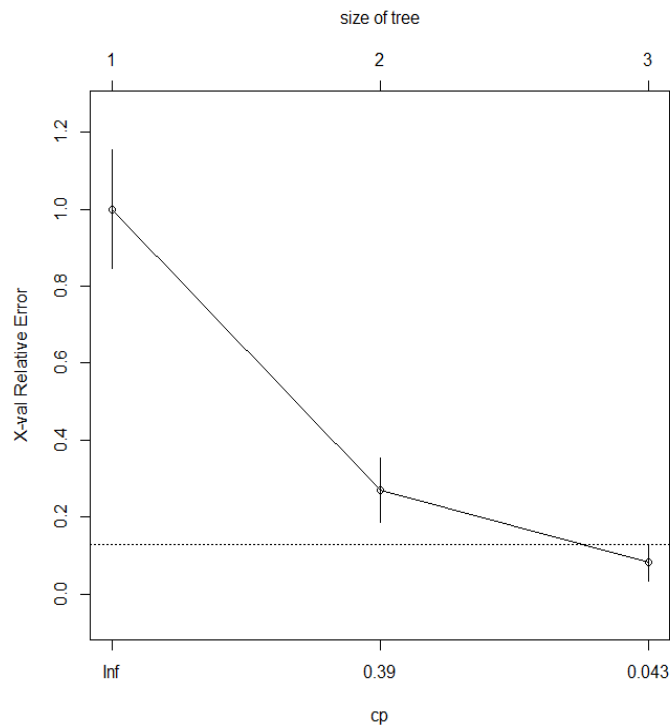


Figure annexe 23. Représentation de la moyenne et de l'écart type de l'erreur relative du modèle en fonction du nombre de nœuds.

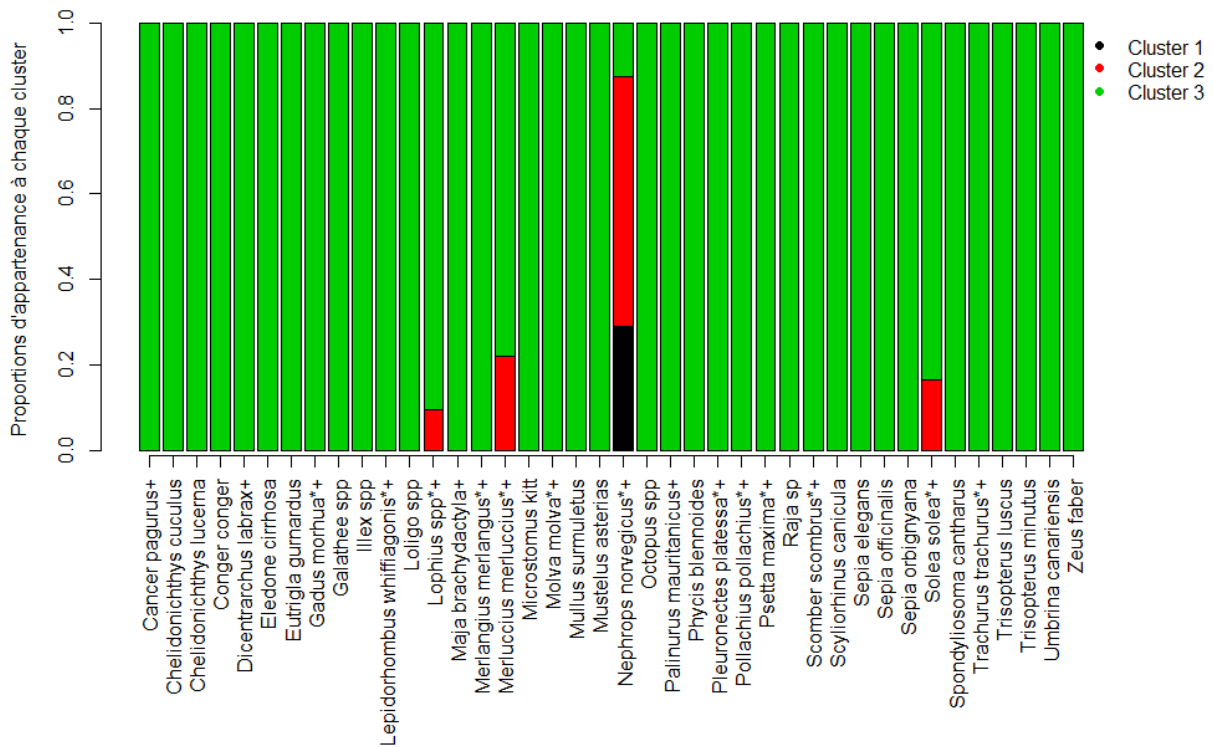


Figure annexe 24. Barplot représentant les proportions d'appartenances aux différents clusters des individus statistiques d'une même espèce.

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD.
Correspondance des couleurs des clusters : Noir = cible, rouge = valorisé, vert = non valorisé.

i. Analyse du modèle CART d'apprentissage sur SACROIS

- Données d'entraînement : séquences prenant en compte les rejets

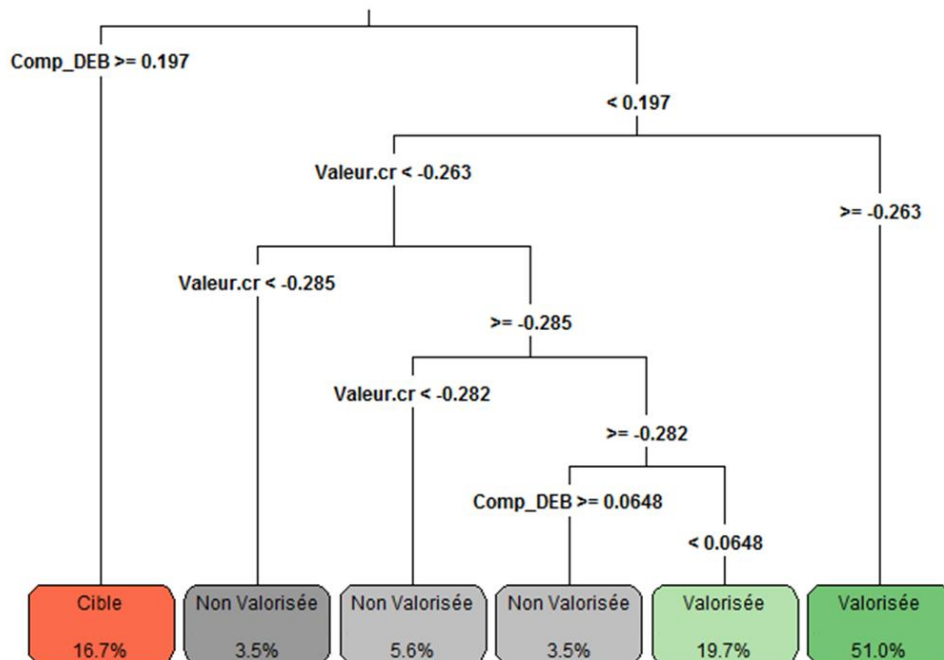


Figure annexe 25. Arbre de classification du modèle prédictif sur SACROIS dont les données d'entraînement des observations à l'échelle de la séquence de pêche avec rejets.
Pour la variable Valeur.cr, les quantités représentées sont centrées réduites.

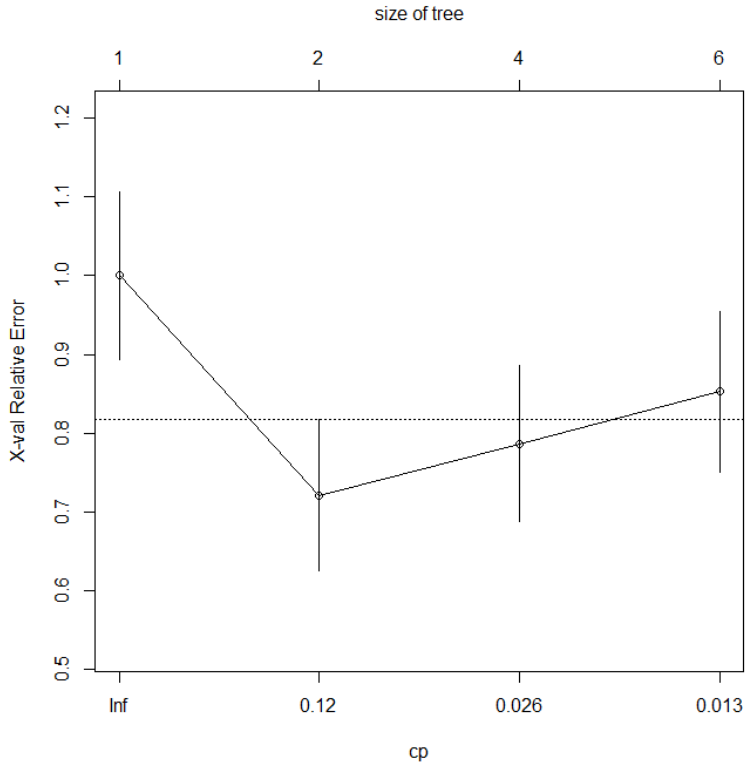


Figure annexe 26. Représentation de la moyenne et de l'écart type de l'erreur relative du modèle en fonction du nombre de nœuds.

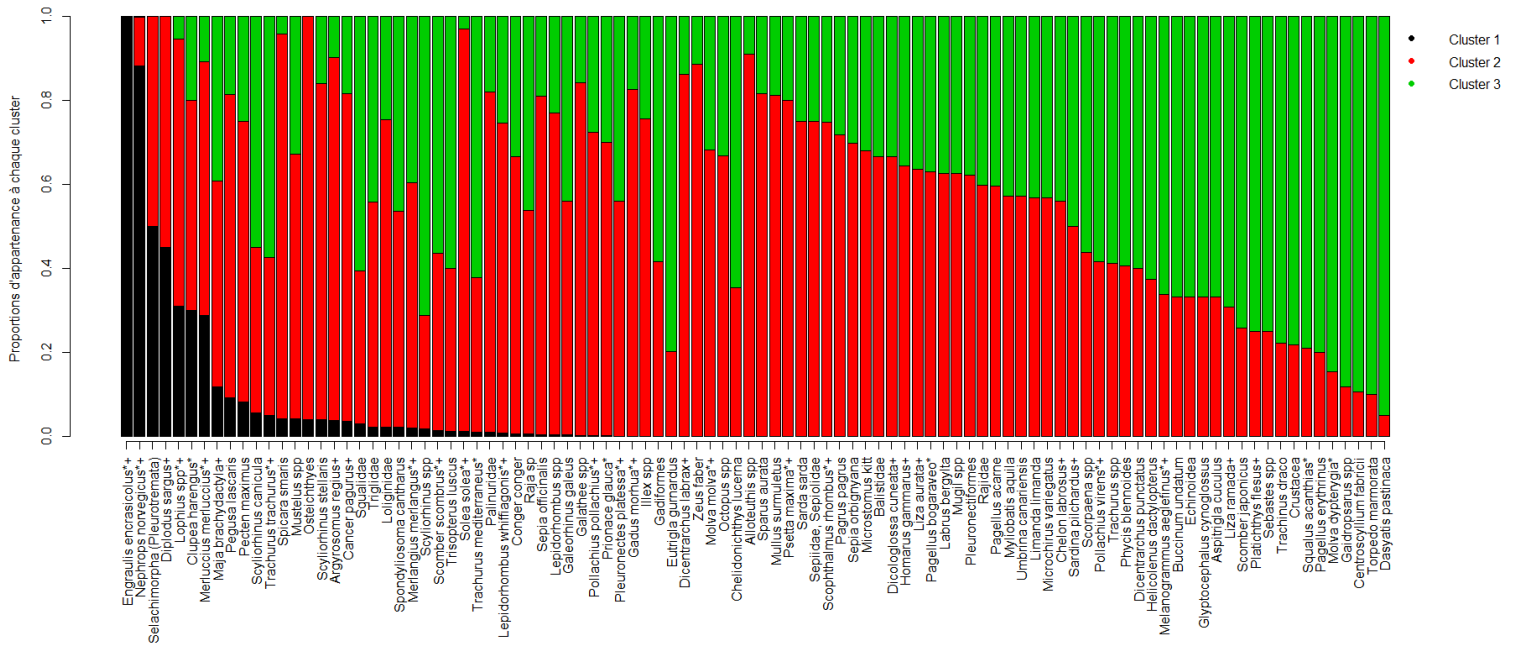


Figure annexe 27. Barplot représentant les proportions d'appartenances aux différents clusters des individus statistiques d'une même espèce.

Les espèces ayant une proportion égale à 1 ne sont pas représentées.

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD.

Correspondance des couleurs des clusters : Noir = cible, rouge = valorisé, vert = non valorisé.

- Données d'entraînement : séquences ne prenant pas en compte les rejets

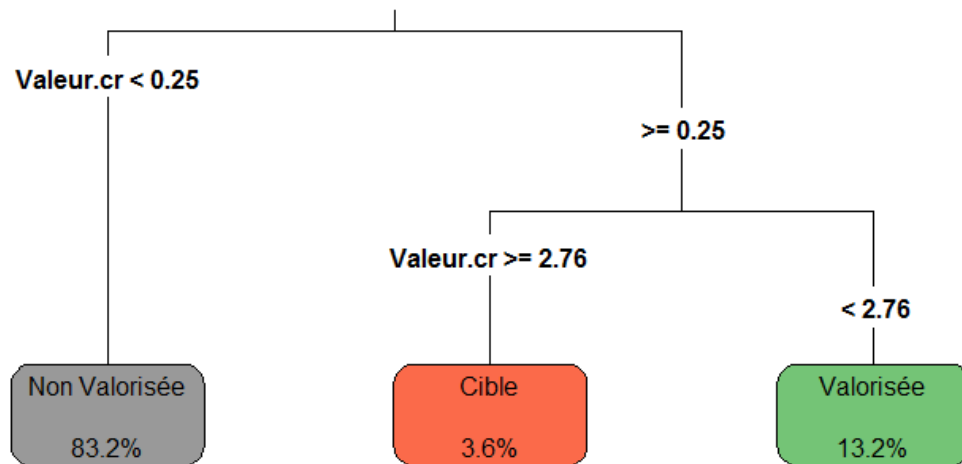


Figure annexe 28. Arbre de classification du modèle prédictif sur SACROIS dont les données d'entraînement des observations à l'échelle de la séquence de pêche sans rejets. Pour la variable Valeur.cr, les quantités représentées sont centrées réduites.

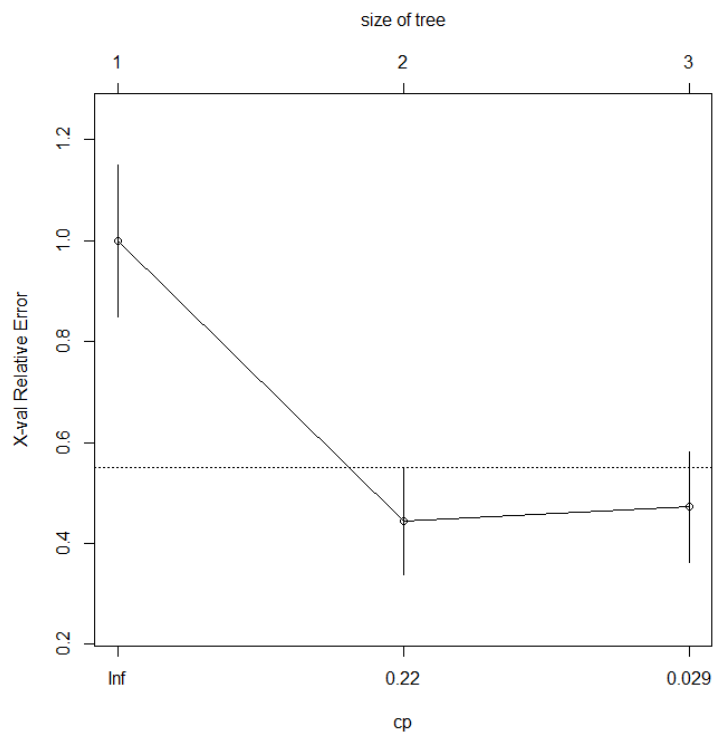


Figure annexe 29. Représentation de la moyenne et de l'écart type de l'erreur relative du modèle en fonction du nombre de nœuds.

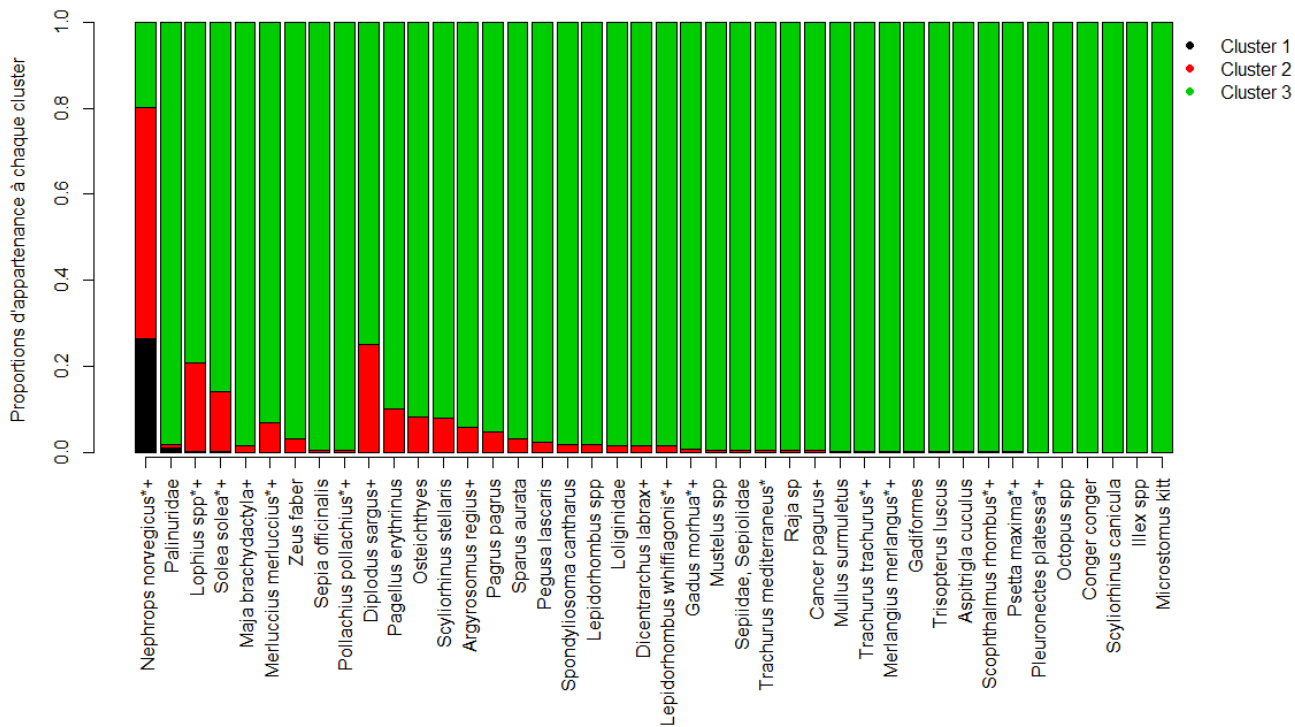


Figure annexe 30. Barplot représentant les proportions d'appartenances aux différents clusters des individus statistiques d'une même espèce

Les espèces ayant une proportion égale à 1 ne sont pas représentées.

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD.

Correspondance des couleurs des clusters : Noir = cible, rouge = valorisé, vert = non valorisé.

j. Tableau récapitulatif des classes d'espèces issues de SACROIS

Tableau annexe 1. Tableau récapitulatif de la classification des espèces à partir de SACROIS en fonction des données et de la méthode utilisée.

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD.
Noir = espèce cible, rouge = espèce valorisée, vert = espèce non valorisée, bleu = espèce mixte

Espèce	CART Séquence avec rejet (ACP+CAH)	CART Séquence sans rejet (K-means)
Alloteuthis spp	Valorisée	Non Valorisée
Alosa alosa, A. fallax	Non Valorisée	Non Valorisée
Argyrosomus regius+	Valorisée	Non Valorisée
Aspitrigla cuculus	Mixte	Non Valorisée
Balistidae	Mixte	Non Valorisée
Buccinum undatum	Mixte	Non Valorisée
Cancer pagurus+	Valorisée	Non Valorisée
Centroscyllium fabricii	Non Valorisée	Non Valorisée
Cepola macrophthalma	Non Valorisée	Non Valorisée
Chelidonichthys lucerna	Mixte	Non Valorisée
Chelon labrosus+	Mixte	Non Valorisée
Clupea harengus*	Cible	Non Valorisée
Conger conger	Mixte	Non Valorisée
Crustacea	Non Valorisée	Non Valorisée
Dasyatis pastinaca	Non Valorisée	Non Valorisée
Dicentrarchus labrax+	Valorisée	Non Valorisée
Dicentrarchus punctatus	Mixte	Non Valorisée
Dicologlossa cuneata+	Mixte	Non Valorisée
Diplodus sargus+	Cible	Non Valorisée
Echinoidea	Mixte	Non Valorisée
Engraulis encrasicolus*+	Cible	Valorisée
Eutrigla gurnardus	Non Valorisée	Non Valorisée
Gadiformes	Mixte	Non Valorisée
Gadus morhua*+	Valorisée	Non Valorisée
Gaidropsarus spp	Non Valorisée	Non Valorisée
Galathee spp	Valorisée	Non Valorisée
Galeorhinus galeus	Mixte	Non Valorisée
Glyptocephalus cynoglossus	Mixte	Non Valorisée
Helicolenus dactylopterus	Mixte	Non Valorisée
Homarus gammarus+	Mixte	Non Valorisée
Illex spp	Valorisée	Non Valorisée
Labrus bergylta	Mixte	Non Valorisée
Lepidorhombus spp	Valorisée	Non Valorisée
Lepidorhombus whiffiagonis*+	Valorisée	Non Valorisée
Limanda limanda	Mixte	Non Valorisée
Liza aurata+	Mixte	Non Valorisée
Liza ramada+	Mixte	Non Valorisée
Loliginidae	Valorisée	Non Valorisée
Lophius spp*+	Cible	Non Valorisée
Maja brachydactyla+	Mixte	Non Valorisée
Melanogrammus aeglefinus*+	Mixte	Non Valorisée
Merlangius merlangus*+	Mixte	Non Valorisée
Merluccius merluccius*+	Cible	Non Valorisée
Microchirus variegatus	Mixte	Non Valorisée
Microstomus kitt	Mixte	Non Valorisée
Molva dypterygia*	Non Valorisée	Non Valorisée
Molva molva*+	Mixte	Non Valorisée
Mugil cephalus	Non Valorisée	Non Valorisée
Mugil spp	Mixte	Non Valorisée
Mullus surmuletus	Valorisée	Non Valorisée
Mustelus spp	Mixte	Non Valorisée
Myliobatis aquila	Mixte	Non Valorisée
Nephrops norvegicus*+	Cible	Cible
Octopus spp	Mixte	Non Valorisée

Osteichthyes	Valorisée	Non Valorisée
Pagellus acarne	Mixte	Non Valorisée
Pagellus bogaraveo*	Mixte	Non Valorisée
Pagellus erythrinus	Non Valorisée	Non Valorisée
Pagrus pagrus	Valorisée	Non Valorisée
Palinuridae	Valorisée	Non Valorisée
Pecten maximus	Mixte	Non Valorisée
Pegusa lascaris	Valorisée	Non Valorisée
Phycis blennoides	Mixte	Non Valorisée
Platichthys flesus+	Non Valorisée	Non Valorisée
Pleuronectes platessa*+	Mixte	Non Valorisée
Pleuronectiformes	Mixte	Non Valorisée
Pollachius pollachius*+	Valorisée	Non Valorisée
Pollachius virens*+	Mixte	Non Valorisée
Prionace glauca*	Mixte	Non Valorisée
Psetta maxima*+	Valorisée	Non Valorisée
Raja sp	Mixte	Non Valorisée
Rajidae	Mixte	Non Valorisée
Sarda sarda	Valorisée	Non Valorisée
Sardina pilchardus+	Mixte	Non Valorisée
Scomber japonicus	Non Valorisée	Non Valorisée
Scomber scombrus*+	Mixte	Non Valorisée
Scophthalmus rhombus*+	Valorisée	Non Valorisée
Scorpaena spp	Mixte	Non Valorisée
Scyliorhinus canicula	Mixte	Non Valorisée
Scyliorhinus spp	Non Valorisée	Non Valorisée
Scyliorhinus stellaris	Valorisée	Non Valorisée
Sebastes spp	Non Valorisée	Non Valorisée
Selachimorpha (Pleurotremata)	Cible	Non Valorisée
Sepia officinalis	Valorisée	Non Valorisée
Sepia orbignyana	Mixte	Non Valorisée
Sepiidae, Sepiolidae	Valorisée	Non Valorisée
Solea solea*+	Valorisée	Non Valorisée
Soleidae	Non Valorisée	Non Valorisée
Sparus aurata	Valorisée	Non Valorisée
Spicara smaris	Valorisée	Non Valorisée
Spondyliosoma cantharus	Mixte	Non Valorisée
Sprattus sprattus*	Non Valorisée	Non Valorisée
Squalidae	Mixte	Non Valorisée
Squalus acanthias*	Non Valorisée	Non Valorisée
Torpedo marmorata	Non Valorisée	Non Valorisée
Trachinus draco	Non Valorisée	Non Valorisée
Trachurus mediterraneus*	Mixte	Non Valorisée
Trachurus spp	Mixte	Non Valorisée
Trachurus trachurus*+	Mixte	Non Valorisée
Triglidae	Mixte	Non Valorisée
Trisopterus luscus	Mixte	Non Valorisée
Umbrina canariensis	Mixte	Non Valorisée
Zeus faber	Valorisée	Non Valorisée

ANNEXE IV. Profil des variables pour les espèces sous gestion

Dans les nominations d'espèces : * = espèce soumise à un TAC, + = espèce soumise à une TMD. Correspondance des couleurs des clusters : Noir = cible, rouge = valorisé, vert = non valorisé, bleu = mixte.

Les graphiques de l'annexe IV, représentent les espèces sous gestion dont la proportion dans les captures (ou les débarquements) est supérieure à 1%

1. Echelle de l'opération de pêche prenant en compte les rejets

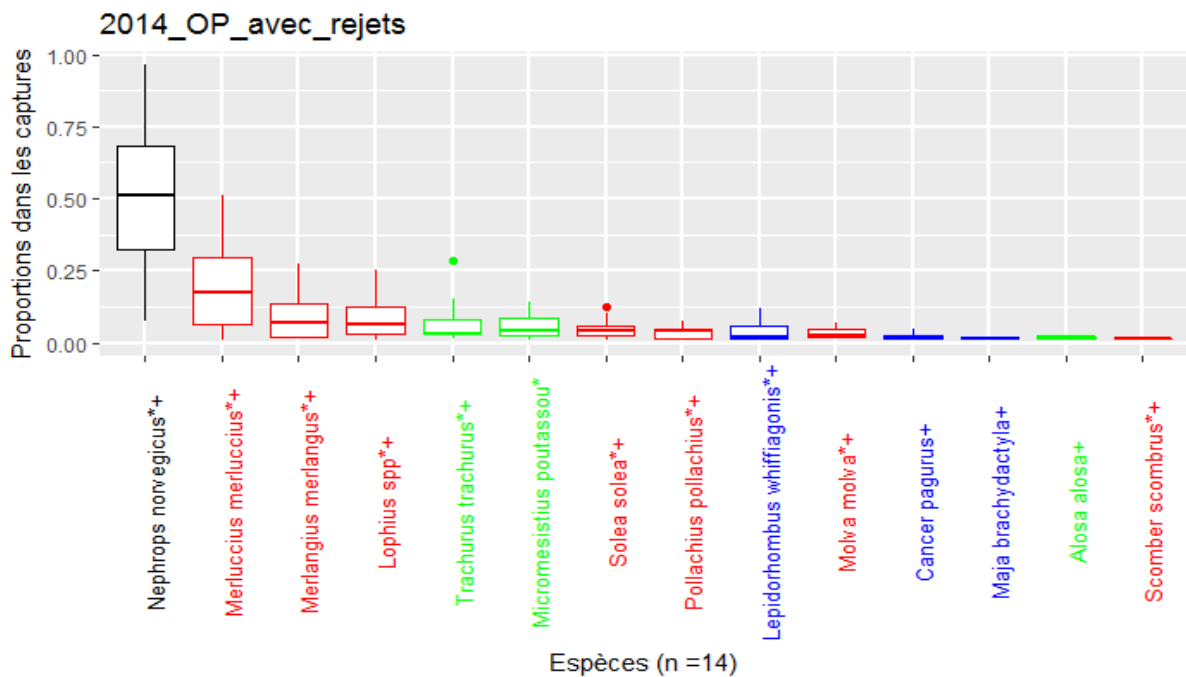


Figure annexe 31. Proportions moyennes dans les captures des espèces sous gestion à l'échelle de l'opération de pêche avec les rejets.

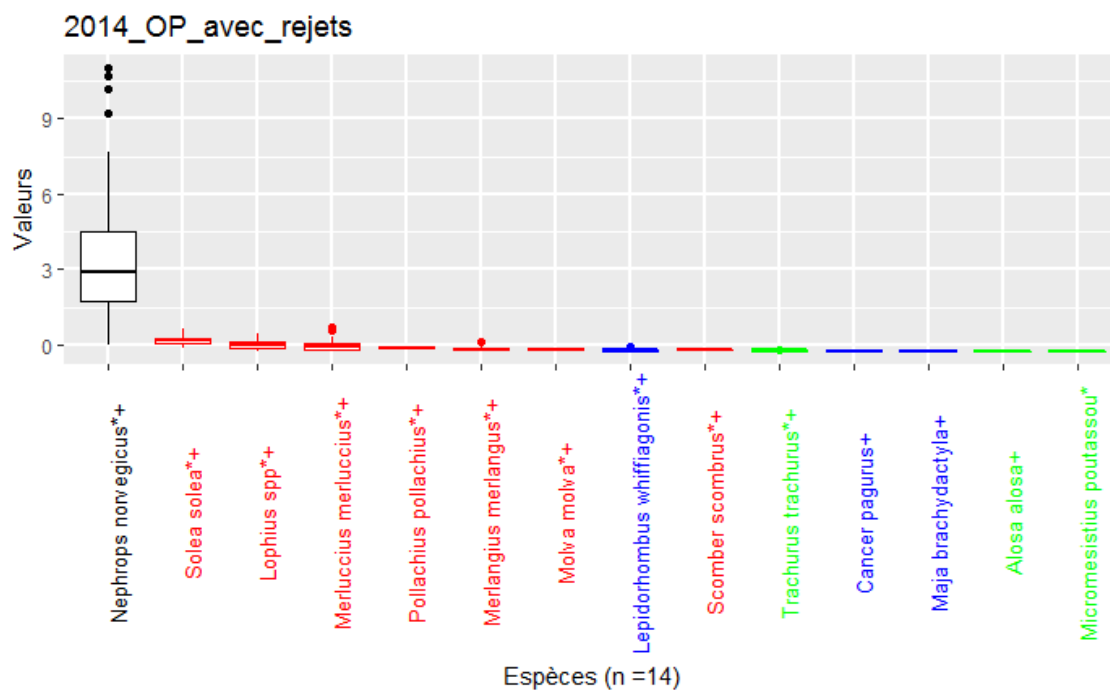


Figure annexe 32. Profil de valeur des espèces sous gestion à l'échelle de l'opération de pêche avec les rejets.

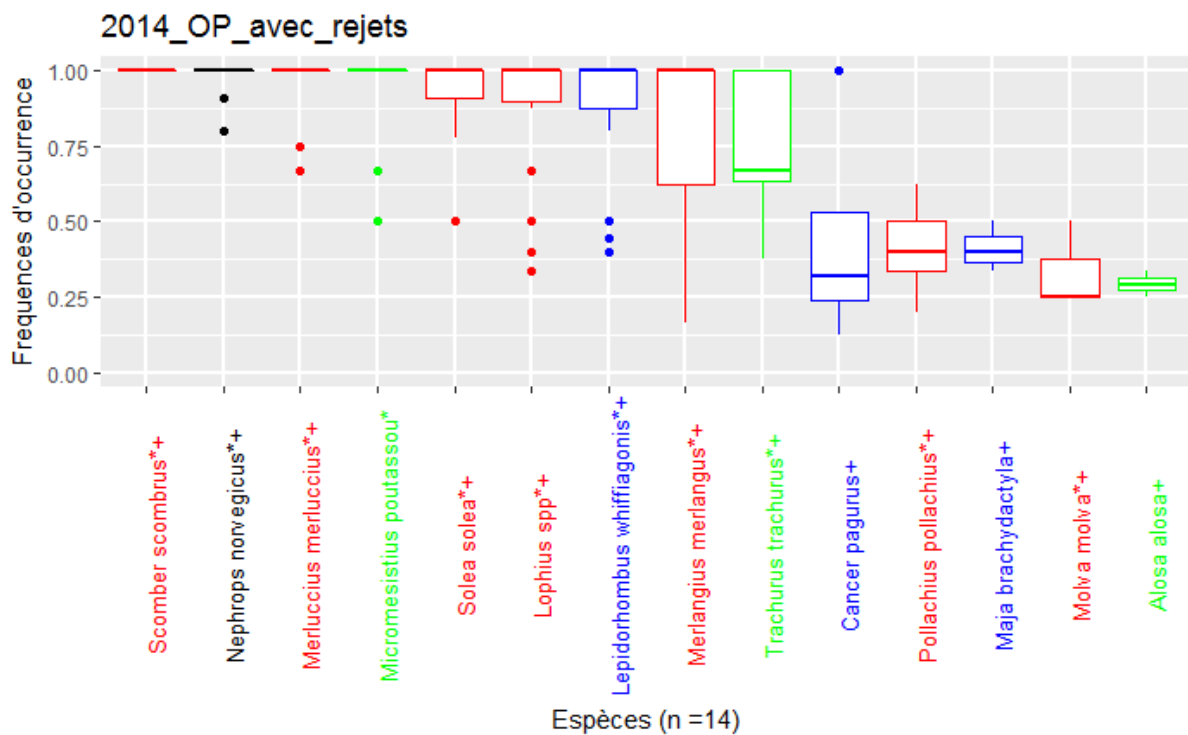


Figure annexe 34. Fréquences d'occurrence des espèces sous gestion à l'échelle de l'opération de pêche avec les rejets.

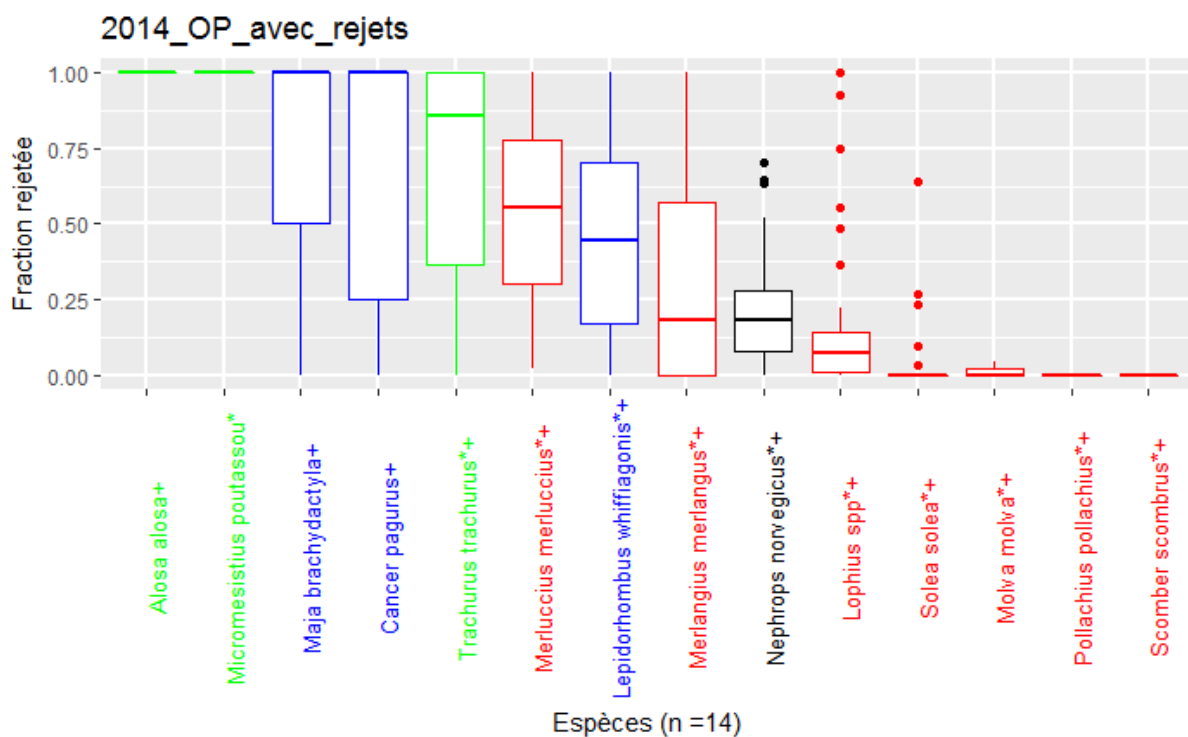


Figure annexe 33. Fractions de rejets des espèces sous gestion à l'échelle de l'opération de pêche avec les rejets.

2. Echelle de l'opération de pêche ne prenant pas en compte les rejets

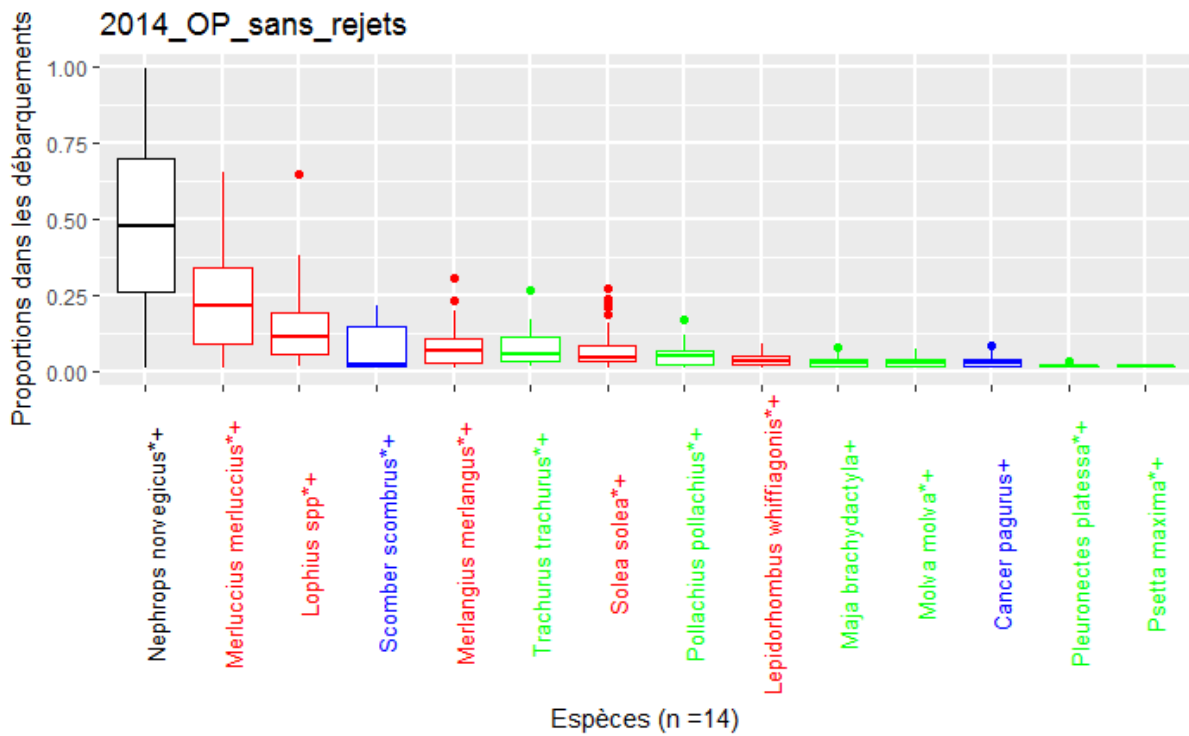


Figure annexe 35. Proportions moyennes dans les débarquements des espèces sous gestion à l'échelle de l'opération de pêche sans les rejets.

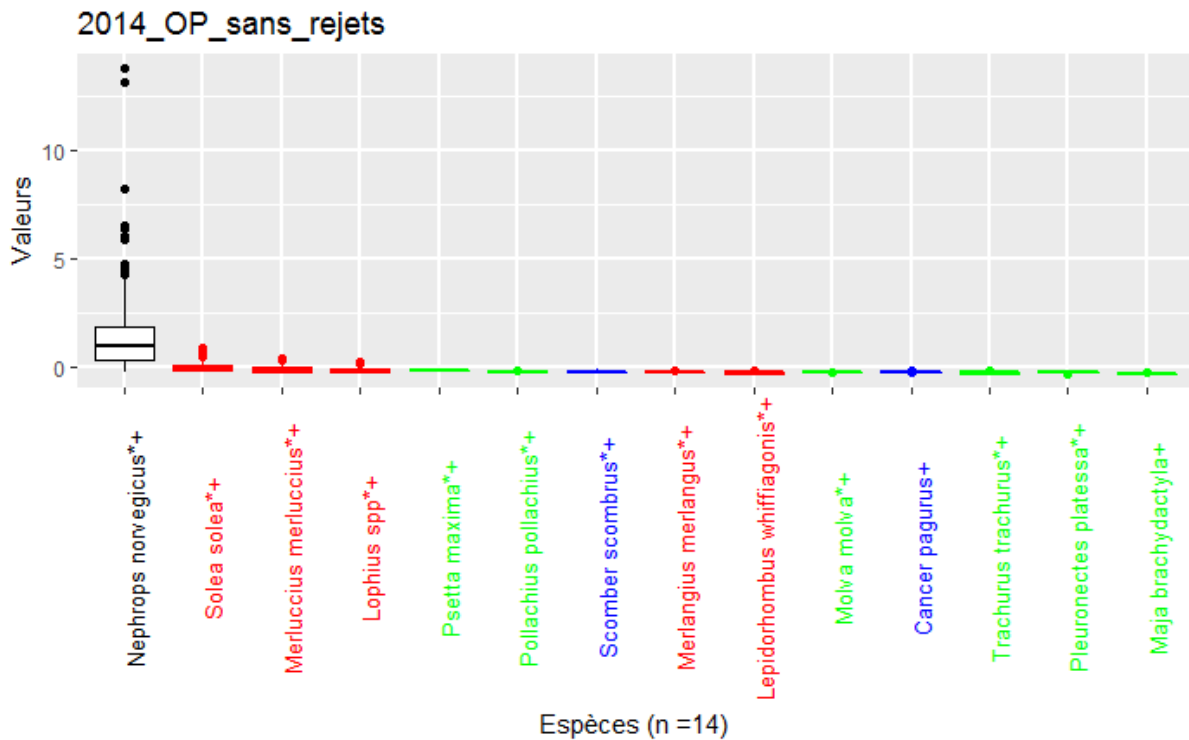


Figure annexe 36. Profil de valeur des espèces sous gestion à l'échelle de l'opération de pêche sans les rejets.

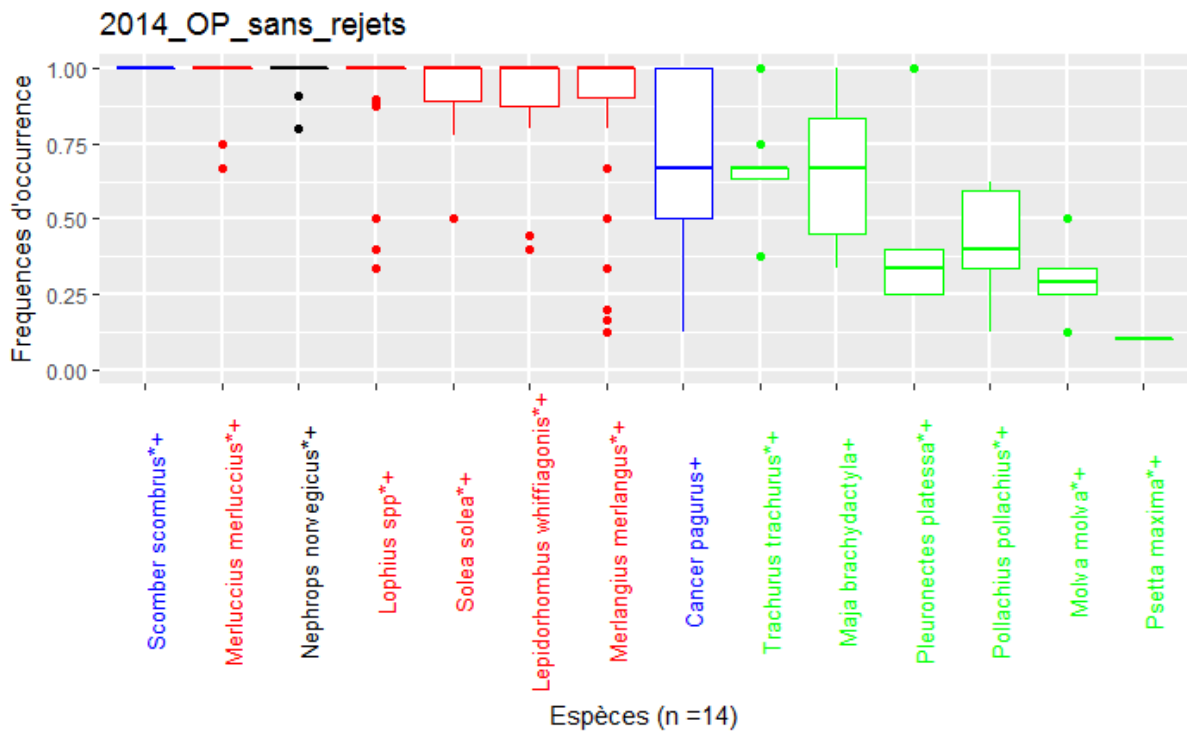


Figure annexe 37. Fréquences d'occurrence des espèces sous gestion à l'échelle de l'opération de pêche sans les rejets.

3. Echelle de la séquence de pêche prenant en compte les rejets

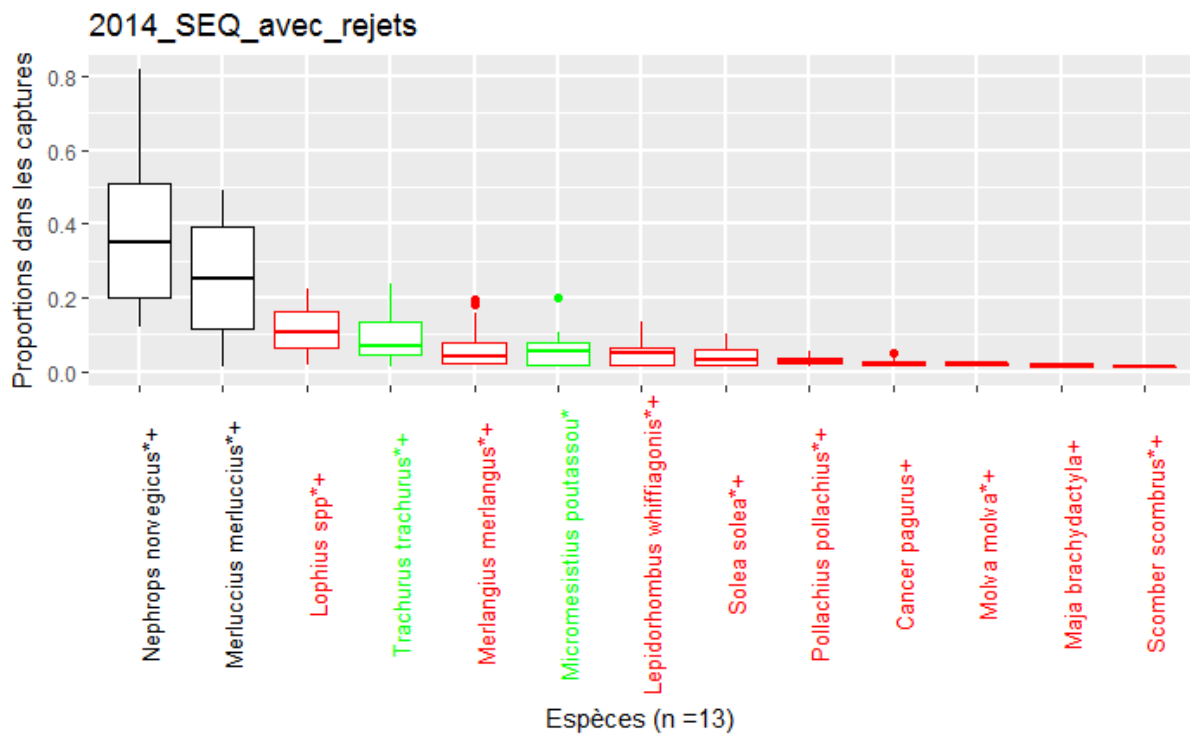


Figure annexe 38. Proportions moyennes dans les captures des espèces sous gestion à l'échelle de la séquence de pêche avec les rejets.



Figure annexe 39. Profil de valeur des espèces sous gestion à l'échelle de la séquence de pêche avec les rejets.

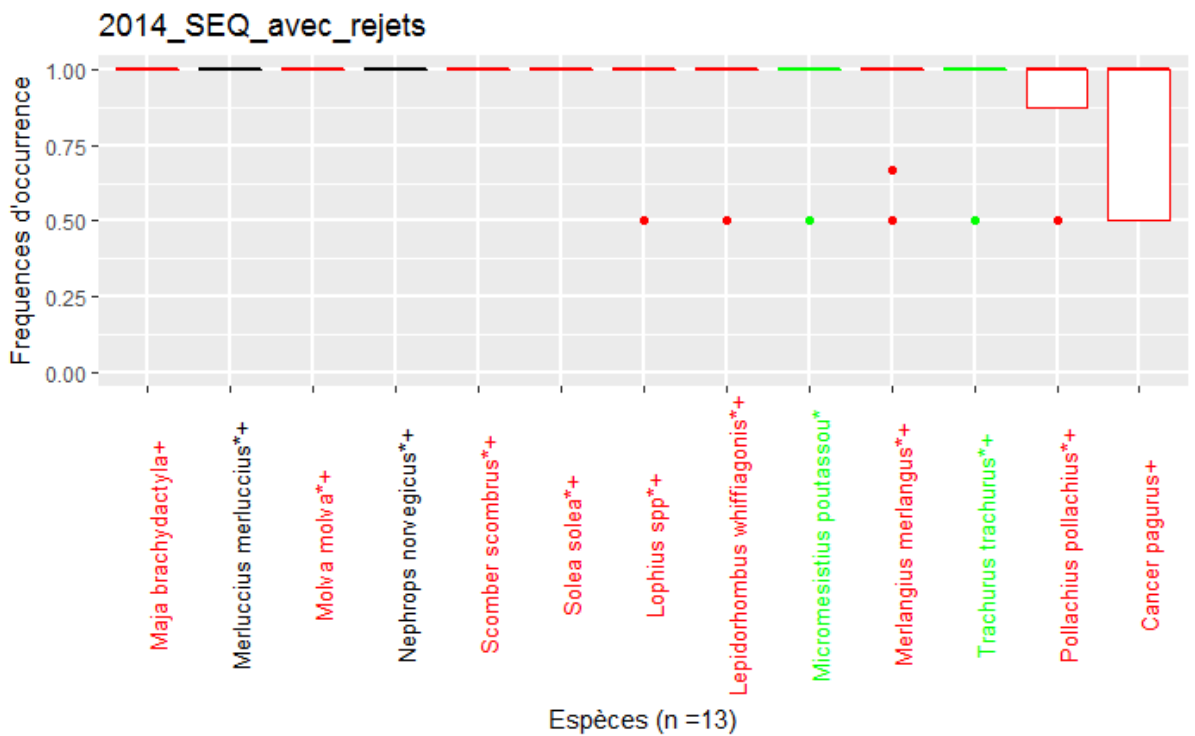


Figure annexe 40. Fréquences d'occurrence des espèces sous gestion à l'échelle de la séquence de pêche avec les rejets.

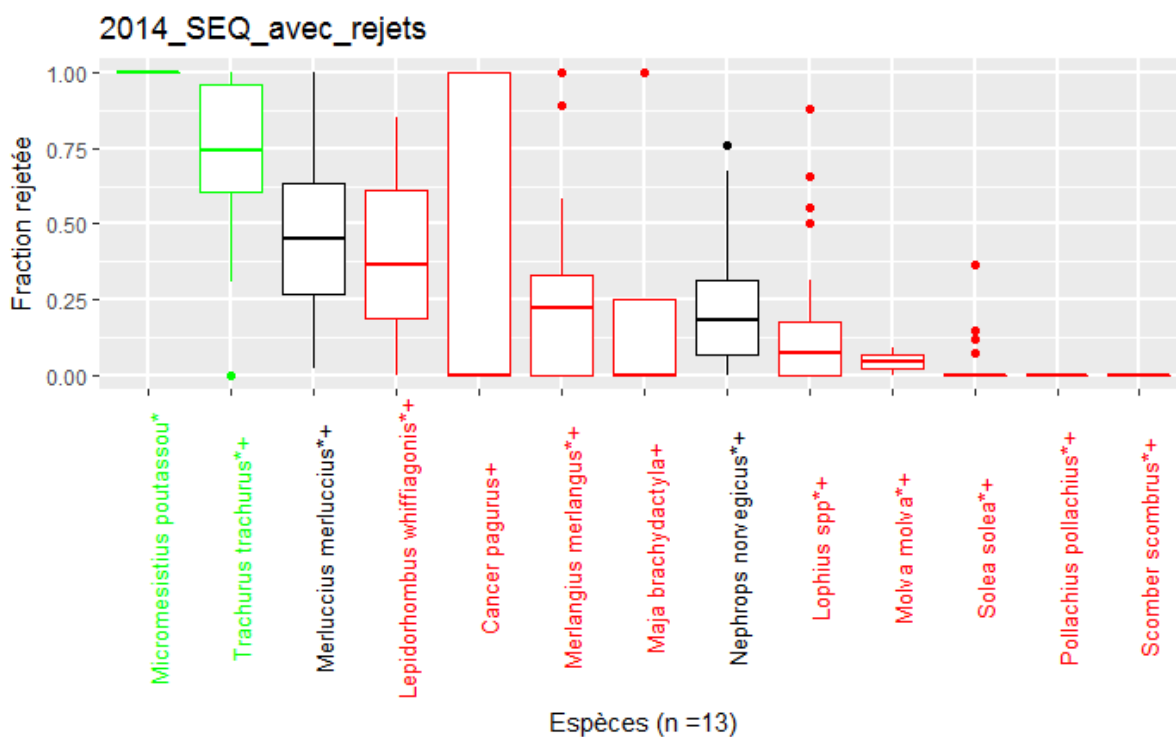


Figure annexe 41. Fractions de rejets des espèces sous gestion à l'échelle de la séquence de pêche avec les rejets.

4. Echelle de la séquence de pêche ne prenant pas en compte les rejets

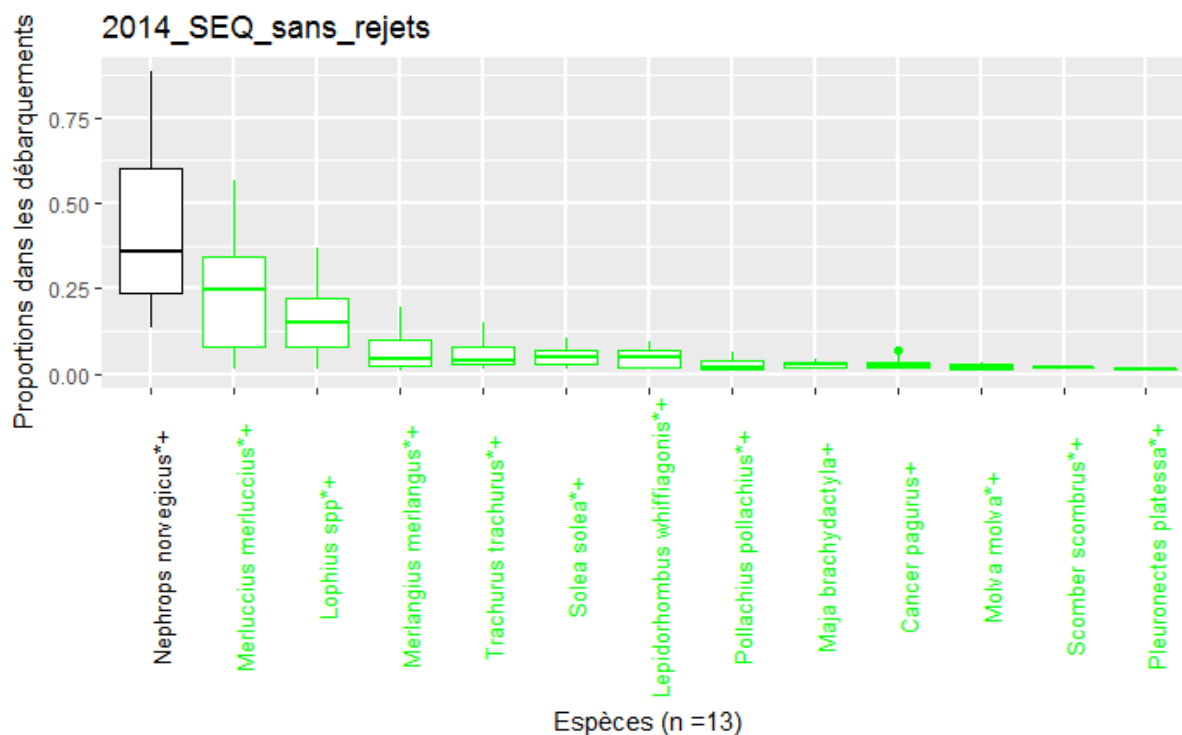


Figure annexe 42. Proportions moyennes dans les débarquements des espèces sous gestion à l'échelle de la séquence de pêche sans les rejets.

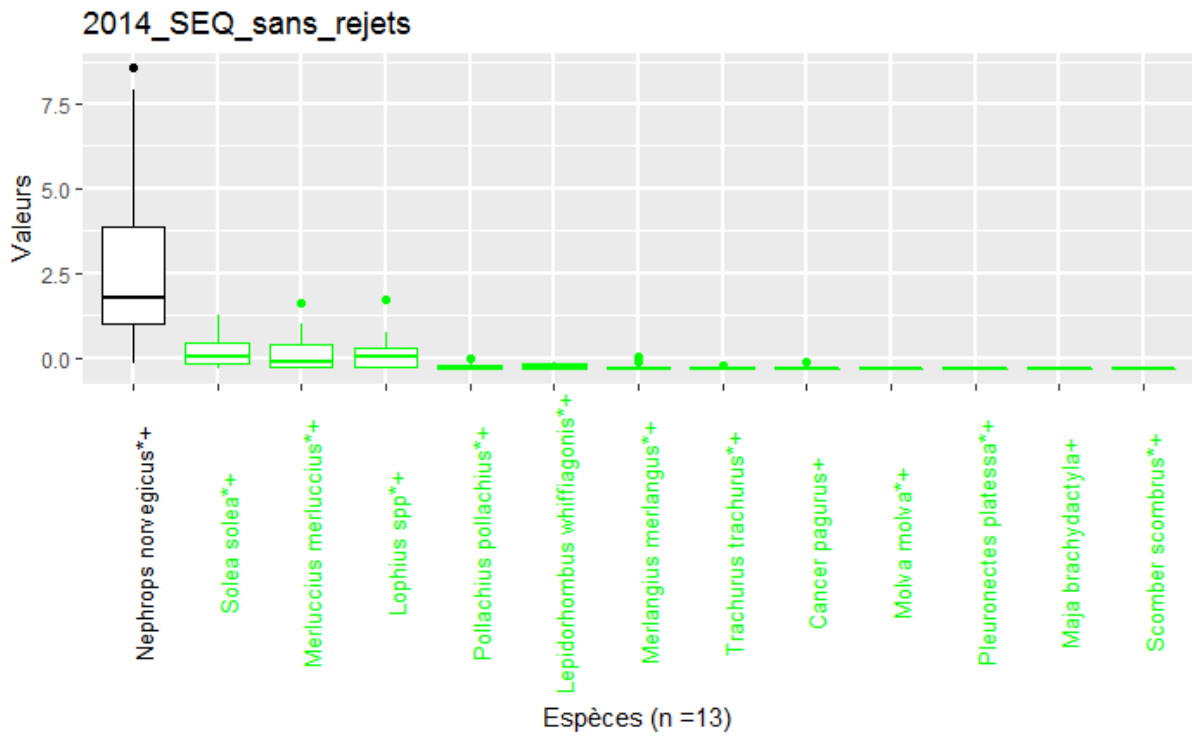


Figure annexe 44. Profil de valeur des espèces sous gestion à l'échelle de la séquence de pêche sans les rejets.

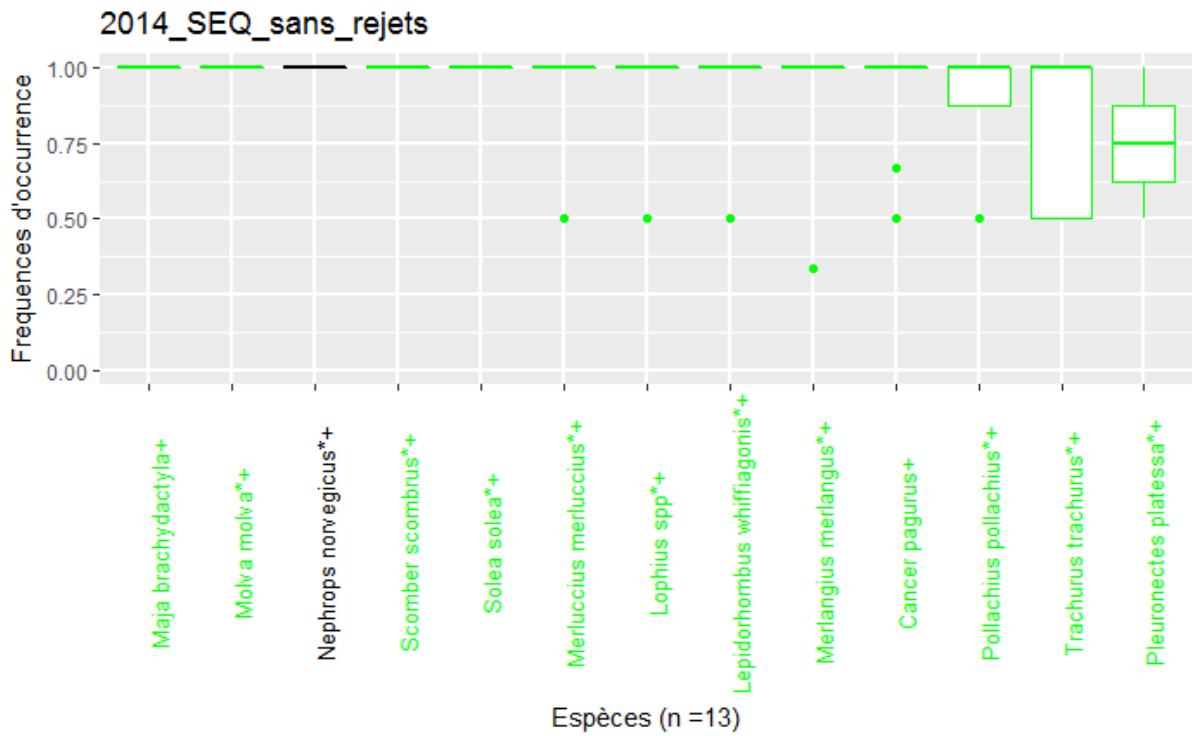



Figure annexe 43. Fréquences d'occurrence des espèces sous gestion à l'échelle de la séquence de pêche sans les rejets.

	Diplôme : Master Spécialité : SML-Biologie Parcours (option) : Sciences Halieutiques et Aquacoles (REA) Enseignants référents : Marie-Pierre ETIENNE et Jérôme GUITTON
Auteur(s) : JULIEN Timothée Date de naissance* : 08/06/1995	Organisme d'accueil : Ifremer Nantes Adresse : Rue de l'Île d'Yeu, 44980 Nantes
Nb pages : 31 Annexe(s) : 5	Maîtres de stage : Anne–Sophie CORNOU,
Année de soutenance : 2018	Stéphanie MAHEVAS et Youen VERMARD
Titre français : Identification des espèces cibles et accessoires dans la pêche mixte démersale du golfe de Gascogne.	
Titre anglais : Identification of target and accessory species in the mixed demersal fishery of the Bay of Biscay.	
Résumé (1600 caractères maximum) : La gestion avec des TAC (Total Autorisé de Capture) mono-spécifiques et les plans de gestion couplés à la mise en place de l'obligation de débarquement peut conduire à l'apparition d'espèces bloquantes dans les pêcheries mixtes. En effet, une espèce sous quotas même si capturée en faible quantité peut bloquer une pêche si celle-ci n'a pas de TAC alloué. Les plans de gestions pluriannuels pourraient permettre de limiter l'impact des espèces bloquantes dans le cas où celles-ci ne sont pas les espèces cibles des flottilles. Il est donc nécessaire de définir le statut des espèces. La définition <i>a priori</i> repose sur différentes variables : la contribution de l'espèce dans les captures, la valeur, la fraction rejetée et la fréquence d'occurrence. L'étude s'intéresse à la pêche mixte démersale du golfe de Gascogne et plus particulièrement aux métiers chaluts de fond et jumeaux à panneaux ciblant la langoustine pour l'année 2014. Dans un premier temps, la démarche consiste à catégoriser les espèces à partir d'un jeu de données exhaustif pour les captures mais limité en termes d'observation. Différentes méthodes d'analyse multivariées et de classification seront utilisées. Ensuite, les données de rejets seront supprimées puis une élévation à une échelle supérieure sera réalisée afin de comprendre les biais de classification sur le jeu de données final, ayant la même forme d'agrégation que les données de déclaration des journaux de pêche réglementaires. La suppression des rejets ainsi que le changement d'échelle induisent d'importants biais entraînant de fortes modifications dans les résultats de catégorisation. Enfin, une méthode de classification supervisée sur les données finales est également réalisée afin de prédire les statuts des espèces de cette pêche.	
Abstract (1600 caractères maximum) : Managing with monospecific TACs (Total Allowable Catches) and the introduction of the Multi-Annual Plans and the implementation of the Landing Obligation may lead to the emergence of choke species in mixed fisheries. Indeed, a species under quota even if caught in small quantities can lead to the closure of a fishery if it does not have an allocated TAC. The Multi-Annual Plan might help reducing the impact of the Landing Obligation if a species is not targeted. It is therefore necessary to define the status of species. The <i>a priori</i> definition of target/non target species is based on several variables: the contribution of the species in the catches, the value, the discard fraction and the frequency of occurrence. The study deals with the mixed demersal fishery of the Bay of Biscay and more precisely with bottom trawl and twin trawlers targeting Norway lobster for the year 2014. Initially, the approach consists in categorizing the species from a comprehensive dataset for catches but with a limited number of observations. Different methods of multivariate analysis and classification will be used. Then, the discard data will be deleted and a scaling up will be done to understand the classification bias on the final dataset, built with the same level of aggregation that the logbooks. The elimination of discards and the change of scale induce important biases leading to significant changes in the categorization results. Finally, a supervised classification method on the final data is also carried out to predict the status of the species in this fishery.	
Mots-clés : classification, analyses multivariées, espèces accessoires, pêcheries mixtes, golfe de Gascogne	
Key Words : clustering, multivariate analysis, accessory species, mixed fisheries, Bay of Biscay	