



## **Université de Montpellier**

### **Master 2**

Mathématique de l'INformation et de la Décision

&

Statistiques pour l'Information et l'Aide à la Décision

Année 2018-2019

### **Mémoire de stage**

Estimation des biais de déclaration des thonidés dans les livres de bord pour la pêche à la senne en milieu pélagique tropical

Auteur :

Victor-Alexandre Aragno

Encadrants :

Antoine Duparc (IRD, antoine.duparc@ird.fr)

Mathieu Depetris (IRD, mathieu.depetris@ird.fr)

## Remerciements

Je tiens d'abord à remercier copieusement mes maîtres de stages, Antoine DUPARC et Mathieu DEPETRIS de m'avoir accueilli avec eux. Je les remercie pour leurs bienveillances, leurs patiences et leurs enseignements qui m'ont permis d'appréhender ce projet avec sérénité.

Je tiens également à remercier Julien LEBRANCHU ainsi que Laurent FLOCH de m'avoir aidé à de nombreuses reprises.

J'aimerais remercier plus généralement tous les membres de la station IFERMER de Sète, faire partie de la station fut un moment plaisant et enrichissant. Je souhaite donc remercier Pascal BACH et Laurent DAGORN de m'avoir accueilli au sein de l'Observatoire des écosystèmes pélagiques tropicaux exploités (Ob7) de l'unité mixte de recherche MARBEC.

J'aimerais exprimer ma gratitude aux nombreux professeurs que j'ai pu rencontrer lors de ma scolarité. En ce sens j'aimerais remercier Xavier BRY pour son implication auprès des étudiants du master MIND.

J'aimerais également remercier les stagiaires présents lors de mon stage, leurs présences m'ont permis d'acquérir de nombreuses connaissances ainsi que de passer des moments forts agréables.

## Table des matières

1. Introduction.....	6
2. Matériel et méthodes.....	9
2.1. Sources de données.....	9
2.2. Analyses statistiques.....	9
2.2.1. Taux d'erreur et détermination des seuils de détection.....	11
2.2.2. Étude des erreurs de détections dans les déclarations de bord.....	13
2.2.3. Biais de compositions.....	13
2.2.4. Étude du delta en fonction de proportion de l'espèce contenu dans l'échantillon.....	15
3. Résultats.....	16
3.1. Classification des calées par leurs compositions spécifiques et par leurs compositions en classe de poids.....	16
3.2. Taux d'erreurs et comparaison des seuils de détections.....	17
3.2.1. Taux d'erreurs.....	17
3.2.2. Comparaison des seuils de détection.....	17
3.3. Étude des erreurs de détections dans les déclarations de bord.....	19
3.4. Biais de compositions.....	21
3.4.1. Étude des biais de déclaration en composition spécifique.....	21
Bancs libres.....	21
Bancs objet.....	21
3.4.2. Étude des biais de déclaration en composition par catégories de poids.....	22
Bancs libres.....	22
Bancs objet.....	23
3.5. Étude du delta en fonction de proportion de l'espèce contenu dans l'échantillon.....	24
4. Discussion.....	26
4.1. Des critères d'échantillonnages aux cuves homogènes.....	26
4.2. Les seuils de détection d'espèces et les erreurs de détection des déclarations de bord....	27
4.3. Les biais de compositions.....	28
4.4. Relation entre biais de composition et proportions spécifiques.....	29
5. Conclusion.....	30
Références.....	31
Annexe.....	34
Annexe 1 : La pêche à la senne.....	34
Annexe 2 : Prédiction et IC 95% des biais de composition spécifique pour bancs libres.....	34
Annexe 3 : Prédiction et IC 95% des biais de composition spécifique pour bancs objets.....	35
Annexe 4 : Prédiction et IC 95% des biais de composition par catégories de poids pour bancs libres.....	35
Annexe 5 : Prédiction et IC 95% des biais de composition par catégories de poids pour bancs objets.....	35
Annexe 6 : Groupes par type de banc créés par les kmeans sur la composition spécifique.....	36
Annexe 7 : Groupes par type de banc créés par les kmeans sur la composition par groupe de poids.....	37
Résumé.....	38
Abstract.....	38

## Table des figures

### Index des figures

Figure 1: Dendrogramme des classifications.....	16
Figure 2: Moyenne et quantile du score de détection des déclarations de bord en fonction du pourcentage de l'espèce dans les cuves selon l'échantillon.....	18
Figure 3: Moyenne et quantile du score des détections des échantillons en fonction du pourcentage de l'espèce dans les cuves selon les déclarations de bord.....	19
Figure 4: Facteurs explicatifs du score de détection du livre de bord. Estimé à partir d'un modèle linéaire généralisé mixte.....	20
Figure 5: Prédiction et IC 95% des biais de déclaration entre échantillon et déclaration de bord par espèce sous bancs libres. Estimé à partir d'un modèle linéaire mixte.....	21
Figure 6: Prédiction et IC 95% des biais de déclaration entre échantillon et déclaration de bord par espèce et océan sous bancs objets. Estimé à partir d'un modèle linéaire mixte.....	22
Figure 7: Prédiction et IC 95% des biais de déclaration entre échantillon et déclaration de bord par classes de poids et océans sous bancs libres. Estimés à partir d'un modèle linéaire mixte.....	23
Figure 8: Prédiction et IC 95% des biais de déclaration entre échantillon et déclaration de bord par classes de poids et océans sous bancs objets. Estimés à partir d'un modèle linéaire mixte.....	24
Figure 9: Prédiction et IC 95% des biais de déclaration entre échantillon et déclaration de bord par espèces et types de bancs. Estimés à partir d'un modèle additif mixte.....	25
Figure 10: Biais de composition par type de banc pour le YFT en fonction du groupe de proportion dans la cuve.....	25
Figure 11: Biais de composition par type de banc pour le BET en fonction du groupe de proportion dans la cuve.....	25
Figure 12: Biais de composition par type de banc pour le SKJ en fonction du groupe de proportion dans la cuve.....	25

### Index des tableaux

Tableau 1: Matrice de confusion entre les groupes créée par la classification des compositions de classes de poids et ceux des compositions spécifiques.....	17
Tableau 2: Matrice de confusion entre les groupes créée par la classification des compositions de classes de poids et ceux déjà existants du type de banc.....	17
Tableau 3: Matrice de confusion entre les groupes créée par la classification des compositions spécifiques et ceux déjà existants du type de banc.....	17
Tableau 4: Erreurs quadratiques moyennes des modèles GLM, GAM et forêt aléatoire pour la prédiction de détection du livre de bord. Calculée par validation croisée.....	18
Tableau 5: Synthèse des moyennes et quantiles du score de détection des déclarations de bord en fonction du pourcentage de l'espèce dans l'échantillon.....	18
Tableau 6: Erreurs quadratiques moyennes des modèles GLM, GAM et forêt aléatoire pour la prédiction de détection de l'échantillon. Calculée par validation croisée.....	19
Tableau 7: Synthèse des moyennes et quantiles du score de détection des échantillons en fonction du pourcentage de l'espèce dans les déclarations de bord.....	19
Tableau 8: Résultat de la sélection de modèles par critère BIC pour la régression linéaire mixte des biais de compositions spécifiques sur bancs libres. Nous avons affiché uniquement le premier ainsi que les derniers résultats.....	21
Tableau 9: Résultat de la sélection de modèles par critère BIC pour la régression linéaire mixte des biais de compositions spécifiques sur bancs objet. Nous avons affiché uniquement le premier ainsi que les derniers résultats.....	22

Tableau 10: Résultat de la sélection de modèles par critère BIC pour la régression linéaire mixte des biais de compositions en classes de poids sur bancs libres. Nous avons affiché uniquement le premier ainsi que les derniers résultats.....	23
Tableau 11: Résultat de la sélection de modèles par critère BIC pour la régression linéaire mixte des biais de compositions en classes de poids sur bancs objet. Nous avons affiché uniquement le premier ainsi que les derniers résultats.....	24
Tableau 12: Proportion avec IC95% de la quantité de pêche par marée et par espèces allant dans le marché local (Lesage, 2019).....	29

# 1. Introduction

L'accroissement démographique des populations humaines s'accompagne, entre autres, d'une augmentation de l'utilisation des ressources marines (Joseph, 1997). Selon l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) les captures halieutiques mondiales sont ainsi passées de 18.7 millions de tonnes en 1950 contre 93.7 millions de tonnes en 2014. Au sein de ces captures, 79.9 millions de tonnes sont issues de la pêche marine (FAO, 2018) pour une population mondiale passant de 2.5 à 7,3 milliards (David, 2015). Parmi ces ressources, la capture de thonidés, qui est une des pêches les plus importantes à travers le monde, est passée de 0.6 à 6 millions de tonnes sur la même période. Les premiers pays exploitants les thonidés sont l'Indonésie et le Japon avec en moyenne, respectivement, 653 et 427 milliers de tonnes de poissons capturés entre les années 2010-2014. La France quant à elle, pêche 124 milliers de tonnes (Chassot et al., 2015). Parmi tous les types de pêches, la pêche à la senne coulissante (Annexe 1) est l'une des principales techniques utilisées pour la capture des thonidés en milieu pélagique (non côtière et non abyssale). Elle représente en moyenne 67% des captures entre 2013 et 2017 (FAO, 2018; Restrepo et al., 2017). En zone tropicale, les thonidés dits « majeurs » sont ceux principalement ciblés (environ 50% des captures annuelles dans l'océan Indien, Herrera and Carlos Báez, 2018). Ces derniers sont le « Albacore » (*Thunnus albacares*, cité YFT par la suite), le « Listao » (*Katsuwonus pelamis*, SKJ) et le « Patudo » (*Thunnus obesus*, BET). La pêche à la senne coulissante s'est considérablement développée depuis les années 1980. Elle intègre des outils technologiques sophistiqués tels que des radars et des sonars pour détecter les bancs de poissons (Fonteneau and Pallarés, 1991). Mais ce qui modifia profondément les pratiques de la pêche à la senne coulissante fut le développement de l'usage des Dispositifs de Concentration de Poisson (couramment appelés DCP). En effet, les poissons ont tendance à s'agréger naturellement sous des objets flottants (Fréon and Dagorn, 2000; Romagny et al., 2000) formant des bancs plus facilement localisables et capturables que les bancs dits libres. Historiquement, les pêcheurs ciblaient les objets aléatoirement rencontrés (débris végétaux ou grands mammifères marins par exemple). Mais ces débris naturels ne représentaient qu'une faible proportion des captures (environ 15% dans l'océan Atlantique ; Ariz et al., 1993). Depuis 1990 le déploiement de DCPs s'est largement développé et 65 % des captures de bateaux senneurs auraient lieu sous ces derniers (on parle alors de captures sous banc objet). Par ailleurs, les DCPs peuvent être associés une balise géolocalisable et un échosondeur. Cela leur permet de détecter la présence ou non de banc de poissons et surtout de les localiser (Fonteneau and Pallarés, 1991). De plus, la capture sous banc objet permet une meilleure capturabilité des poissons, le taux de succès d'un coup de pêche est estimé à 89 % sur banc objets contre 46 % sur bancs libres (Davies et al., 2014). Actuellement près de 91 000 DCPs seraient mis à l'eau chaque année. L'océan le plus atteint est l'océan Pacifique qui comptabilise 80 à 85 % des déploiements de DCP. (Scott and Lopez, 2014).

Cette augmentation et ce développement des pêches posent question dans le cadre d'une gestion durable des ressources marines. Actuellement, de nombreuses espèces sont, ou ont été, surexploitées ce qui menace à la fois l'écosystème et le tissu socio-économique dont ces espèces sont la clef de voûte (Pikitch et al., 2004). L'un des exemples les plus emblématiques a été l'effondrement de la population de thon rouge suite à leurs surexploitations couplé à une augmentation croissante de l'effort de pêche associé (O'Neill, 2011). Il est donc primordial d'avoir une estimation précise des

captures afin d'améliorer les estimations des stocks de thonidés pour ainsi permettre aux décisionnaires de prendre les mesures de gestion adéquate pour une exploitation pérenne des ressources marines.

Pour se faire, de nombreuses informations sur les captures, habitats, migrations et caractéristiques biologiques des espèces sont nécessaires. Cependant le manque d'informations, ou d'utilisation d'informations partiellement erronées, rendent les estimations complexes. Par exemple, les données publiées par la FAO pour ses pays membres peuvent sous-estimer les captures (Pauly and Zeller, 2016).

En effet, les problèmes de données et de biais de données sont des questions centrales en écologie, car il est souvent très difficile de faire un recensement exhaustif d'une population. Les statistiques inférentielles essaient de pallier ce problème, en supposant que les caractéristiques propres à l'échantillon peuvent s'appliquer à la population totale. Cependant les problématiques de biais issues des échantillons de données biologiques sont multiples : taille de l'échantillon, événements climatiques, sociologique, économique, etc. (Syfert et al., 2013). Par exemple, les petites tailles d'échantillons posent des problèmes pour toutes analyses statistiques, car elles entraînent une diminution du potentiel prédictif par rapport aux modèles développés avec plus d'occurrences. Par exemple, un échantillon d'une trop petite taille est difficilement représentatif d'une population. Par ailleurs, de nombreux théorèmes mathématiques reposent sur l'hypothèse que le nombre d'individus tend vers l'infini, théorème que nous utilisons pour nos outils statistiques (par exemple la loi des grands nombres ou le théorème central limite ; Mcpherson et al., 2004; Stockwell and Peterson, 2002). De plus, lors d'études basées sur des données provenant d'observations visuelles, un biais est aussi créé par la condition de présence d'une personne pour effectuer l'estimation (Bostrom, 2002; Carter, 1974). D'autres biais sont aussi présents lorsqu'on souhaite étudier les proportions d'individus appartenant à différentes classes d'une population. Les variances des estimations de proportion pour des groupes d'individus « peu présents » sont plus grandes que celles de groupes d'individus « abondamment présents ». Cela a pour effet de donner des estimations moins précises des proportions d'espèces de classes peu représentées et biaise les résultats en faveur des classes d'individus prévalents. Résultats que l'on retrouve dans le modèle de la régression logistique (Fielding and Bell, 1997).

Dans le cadre de la pêche à la senne tropicale ciblant les thonidés, les informations de chaque coup de pêche, appelés calées, sont déclarés dans le livre de bord par le capitaine de chaque navire et sont utilisées pour estimer les quantités et les compositions des captures. Cependant, ces déclarations sont des estimations visuelles réalisées par l'équipage et donc soumises à des biais. Conscient de ces biais, les scientifiques réalisent donc, depuis de nombreuses années, des échantillonnages à bord des navires et aux ports lors des débarquements. Ainsi les premières études sur les biais de déclarations ont détecté qu'une partie des jeunes YFT et des jeunes BET (< 3 kg) sont déclarés en SKJ et qu'une partie des BET de 3 à 15 kg sont déclarés en YFT (Cayré, 1984; Fonteneau, 2008). Depuis, les procédures d'échantillonnages ont été adaptées pour tenir compte de l'évolution des activités de pêche. La conception la plus récente de l'échantillonnage a été mise au point en 1997 à la suite d'analyses de sensibilité réalisées sur différents plans d'échantillonnage lors du projet européen « Échantillonnage Thonier » (ET) (Pallarés and Hallier, 1997a; Pianet et al., 2000). Néanmoins, après 30 ans d'évolution des pratiques des pêcheries (Fonteneau et al., 2000; Parajua and Hallier, 1992), une réévaluation des biais de déclaration semble nécessaire afin de mieux

comprendre les besoins en échantillonnage et améliorer les prédictions des modèles associés d'estimation des captures.

Le but de ce travail est d'étudier les biais entre les déclarations des livres de bord grâce aux échantillonnages faits aux ports. Pour cela nous avons abordé 2 axes de travail.

Le premier axe repose sur l'hypothèse selon laquelle les détections d'espèces par les échantillons seraient plus fiables que celles issues des déclarations de bords. Cette supposition vient du fait que les déclarations de bords sont des estimations visuelles et il semble difficile de quantifier l'ensemble des poissons pris dans une calée. Afin de vérifier cette hypothèse, nous avons étudié le taux d'erreurs de détections, ainsi que les différences de seuils de détections d'espèces (d'une source de donnée, déclaration de bord ou échantillon) en fonction de la proportion de l'espèce dans la cuve (selon la seconde source de données). De plus nous avons étudié si les erreurs de détections d'espèces des déclarations de bord peuvent varier selon plusieurs paramètres :

- le type de banc (banc libre ou banc objet) sous lequel la pêche se déroule. En effet les YFT et BET sont généralement de plus petite taille (moins de 10kg) sous bancs objets alors que ceux en banc libre (Fonteneau et al., 2013). Comme les petits thonidés sont difficilement différenciables entre ces deux espèces, on peut alors supposer que pêcher sous bancs objet augmente les biais de détection et ainsi les biais de déclarations.
- l'expérience de l'équipage. La pêche à la senne coulissante tel qu'elle est exercée actuellement est une activité récente, datant de 1980. Le temps d'adaptation aux nouvelles technologies et l'expérience des membres d'équipage leur permettent d'apprécier, plus efficacement les espèces, et ainsi diminuer les erreurs de détection observées.
- L'océan. Les biais de détections ne sont pas forcément les mêmes dans l'océan Atlantique et dans l'océan Indien en raison des conditions climatiques et des différentes habitudes de pêches.
- Saisonnalité. Des conditions variables extérieures à la pêche telle que des conditions économiques, climatiques ou même l'instauration de quotas sur certaines espèces pourraient se caractériser par une variation saisonnière des biais de détections.

Le second axe de mon travail a porté sur la caractérisation des biais de composition entre les déclarations de bord et les échantillons. Les facteurs précédemment cités sont aussi à même d'influencer les biais de composition. À ces derniers s'ajoutent d'autres paramètres :

- Les catégories de poids des espèces . Les déclarations de pêches se font en classes de poids (appelées catégorie commerciale). Elles sont souvent peu précises (exemple : «plus de 10kg», «moins de 30kg», «entre 20 et 40kg», etc.) et difficilement interprétables biologiquement. Par exemple, pour 1 tonne de YFT classée en catégorie supérieure à 10Kg, nous ne savons pas s'il faut considérer des poissons proches de 10kg, de 50kg ou plus encore. De plus, le poids d'une espèce est difficilement observable, car le poids d'une espèce n'est pas directement proportionnel à sa taille (un YFT de 30kg qui mesure 1m n'est pas deux fois plus lourd qu'un YFT de 50cm). Nous nous attendons donc à ce que les biais varient selon les classes considérées.

Enfin, la déclaration reposant sur des estimations, il est possible que le biais de déclaration soit dépendant de la proportion de l'espèce. Nous nous attendons par exemple à ce qu'une espèce très abondante dans les captures soit surestimée au détriment des espèces moins présentes. Nous avons donc testé cette hypothèse pour essayer de déterminer le mécanisme sous-jacent expliquant les biais de déclaration.

## 2. Matériel et méthodes

### 2.1. Sources de données

Pour pouvoir observer et analyser les éventuelles variations de captures, deux bases d'informations ont été utilisées.

#### *Les livres de bord*

Le livre de bord établi par le capitaine est transmis au retour de la marée (période de pêche entre le départ du port et le débarquement), à l'arrivée au port du navire. Ce livre de bord contient les caractéristiques de la marée. Par exemple pour chacune calée associée on pourra notamment retrouver les informations suivantes : océan, date, heure, type de banc, la quantité capturée par espèce/catégorie de poids ainsi que la ou les cuves dans lesquelles elle a été stockée. Compléter le livre de bord est une obligation administrative.

#### *L'échantillonnage des thonidés au débarquement*

L'échantillonnage des thonidés au débarquement a été mis en place en 1997 suite au programme ET (Pallarés and Hallier, 1997). C'est une base d'information issue d'échantillonnages de cuves du navire à son arrivée au port. Les cuves échantillonnées sont sélectionnées selon plusieurs critères afin de garantir leur représentativité. Ainsi sont échantillonnées les cuves ne contenant qu'une seule calée (ou dont les calées contenues dans la cuve doivent appartenir au même type de banc), à la même zone de pêche et au même trimestre.

Pour pallier la variabilité des compositions spécifique pouvant provenir du choix de la strate de cuve échantillonné, deux échantillons sont effectués par cuve : le premier à l'ouverture de la cuve et le deuxième passé la moitié de la cuve vidée. La quantité de poisson échantillonnée varie selon trois possibilités en fonction de la taille des poissons contenus dans la cuve :

- Si la cuve contient seulement des poissons mesurant plus de 70 cm, 2 fois 100 poissons seront mesurés en LD1 (longueur mâchoire supérieure – base de la première épine dorsale) dans l'océan Atlantique et 2 fois 75 poissons dans l'océan Indien.
- Si la cuve contient seulement des poissons mesurant moins de 70 cm, 300 + 200 poissons seront comptés/mesurés en LF (longueur fourche) pour les deux océans. À noter que seulement 25 SKJ seront mesurés par échantillon, le poids total sera rapporté au nombre compté.
- Si la cuve contient un mélange de petits et gros poissons, même procédure que pour les inférieurs à 70 cm excepté le fait que les gros poissons seront mesurés en LD1.

### 2.2. Analyses statistiques

#### *Prétraitement des données*

En amont, un travail de prétraitement a été nécessaire afin d'harmoniser les données et la sélection des cuves utilisables pour tester nos hypothèses.

Notre intérêt dans cette étude porte seulement sur le YFT, le BET et le SKJ, nous avons donc considéré que ces trois espèces « remplissent » la cuve, c'est-à-dire que nous avons enlevé des déclarations toutes autres espèces.

En premier lieu, pour comparer les déclarations des livres de bord aux échantillons, nous devons convertir les mesures de longueur des échantillons en LD1/LF en poids. Pour ce faire, nous avons donc utilisé des formules de conversion officielles LD1 vers LF puis taille-LF vers poids (Caverivière, 1976; Cayré and Laloë, 1986; Champagnat and Pianet, 1974; Parks et al., 1982).

En raison du trop grand nombre et de l'impossibilité d'utiliser telle quelle les classes commerciales déclarées et leurs variabilités selon les livres de bord, nous avons dû les standardiser. Pour cela nous avons fusionné des classes commerciales de poids (par espèce) des déclarations de bord à l'aide de la densité empirique des poids contenus dans chaque classe commerciale. Nous avons donc créé trois nouvelles classes pour les YFT et BET (inférieur à 10, compris entre 10 et 30 kg et supérieur à 30 en kg) et deux classes pour le SKJ (inférieur à 1.8 et supérieur à 1,8 en kg). Les poids des déclarations d'espèces et des échantillons ont été sommé selon ces huit classes de poids afin de permettre leurs comparaisons.

Dans un second temps, les cuves ont été sélectionnées selon plusieurs critères permettant de s'assurer de la représentativité des échantillons correspondants aux déclarations. Ainsi, les cuves retenues pour les analyses devaient soit être des monos calées (1 seule calée contenue dans la cuve) soit des multicalées. Pour ces dernières les calées étaient déclarées avec des compositions en espèces similaires et répondaient à plusieurs autres critères. En particulier, les cuves étaient considérées homogènes lorsque les calées étaient géographiquement proches (inférieur à 3 degrés, pour prendre en compte l'hétérogénéité spatiale), représentaient un poids minimum de 6 tonnes (corresponds en moyenne à 30 poissons échantillonnés), étaient pêchées sous le même type de banc et de composition spécifiques ou de composition par classe de poids semblables. Pour ce dernier critère, une classification par méthode des centres mobiles (k-means) suivie d'une classification ascendante hiérarchique (CAH) puis une nouvelle classification des k-means à partir des centres sélectionnés par la CAH ont été utilisées. La CAH a pour but de trouver le nombre de groupe optimal des k-premiers groupes, la seconde classification par k-means se fait alors sur le nombre de groupe choisi à l'aide de la CAH ainsi que les centres des groupes gardés (Bry, 2016).

Notons :

$x_i \in \mathbb{R}^3$  la composition spécifique de la calée  $i$ ,  $z_i^k = 1_{\{x_i \in k\}}$  l'indicatrice d'appartenance de la calée au cluster  $k$ ,  $\mu_k$  le centre du cluster  $k$ .

La méthode du k-means consiste à minimiser sure de distorsion suivante:

$$J(\mu, z) = \sum_{i=1}^n \sum_{k=1}^n z_i^k \|x_i - \mu_k\|^2$$

Le principe de la CAH : est de partir de n (= nombre d'individus) classe puis à chaque pas, agréger les deux classes les plus « proche ».

Nous avons utilisé la méthode de « Ward.D » → maximise l'inertie interclasse

$$I_e = \frac{1}{n} \sum_{i=1}^k n_i * d(g_i, g)^2$$

$g_i$  le centre de gravité du groupe  $G_i$  .

La fonction d est la distance euclidienne.

Cette méthode de classification a été utilisée pour les compositions par classes commerciales et par espèce (toute classe commerciale confondue). Les clusters obtenus par les 2 classifications ont ensuite été décrits puis comparés à l'aide de matrices de confusion.

### 2.2.1. Taux d'erreur et détermination des seuils de détection

Nous avons voulu déterminer le nombre d'erreurs de détection, c'est-à-dire le nombre d'espèces non déclaré d'une source de donnée (déclaration de bord ou échantillon), mais déclaré dans la seconde source de donnée, considérant l'identification de l'espèce correcte. Ensuite, nous avons voulu déterminer les seuils de détection des différentes espèces, c'est-à-dire estimer la probabilité de déclarer la présence d'une espèce selon une source de donnée sachant la fréquence déclarée dans la cuve selon la seconde source.

Pour ce faire, nous avons testé entre différents modèles (le modèle linéaire généralisé, GLM, le modèle additif généralisé, GAM, et la forêt aléatoire, RandomForest) celui qui prédisait le mieux les détections d'espèce à l'aide d'une validation croisée de type K-fold à 5 partitions pour comparer les erreurs quadratiques moyennes (MSE). Enfin nous avons utilisé la méthode de prédiction minimisant l'erreur quadratique et en utilisant une méthode de bootstrap pour créer un intervalle de confiance sur les probabilités de détection.

Théoriquement nous pouvons noter :

$W_{yft}^e, W_{bet}^e$  et  $W_{skj}^e$  Les poids des espèces YFT, BET et SKJ échantillonnée dans une cuve et

$W_{yft}^l, W_{bet}^l$  et  $W_{skj}^l$  ceux des déclarations de bord.

On note également :

$P_{skj}^e, P_{yft}^e$  et  $P_{bet}^e$  ainsi que  $P_{skj}^l, P_{yft}^l$  et  $P_{bet}^l$  Les proportions des espèces dans la cuve selon

l'échantillon ou selon la déclaration de bord. Pour exemple :  $P_{skj}^e = \frac{W_{skj}^e}{W_{yft}^e + W_{bet}^e + W_{skj}^e}$

Un « individu » i sera déterminé par une espèce dans une cuve.

On note alors  $x_i^l$  la proportion d'une espèce dans une cuve selon la déclaration de bord. Ainsi que

$$Y_i^l = 1(x_i^l > 0)$$

La variable indicatrice de présence d'une espèce dans la cuve selon la déclaration de bord.

Le GLM consiste à supposer que :

$$Y_i | X_i = x_i \sim B\left(\frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}\right) \text{ Ou les variables } Y_i \text{ sont indépendantes.}$$

En déduire que :

$$P(Y = 1 | X = x) = \frac{\exp(x^T \beta)}{1 + \exp(x^T \beta)} = E(Y | X = x)$$

La fonction de vraisemblance à maximiser en  $\beta$  est :

$$V(\beta) = \prod_{i=1}^n \left(\frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}\right)^{y_i} \left(\frac{1}{1 + \exp(x_i^T \beta)}\right)^{1-y_i}$$

La fonction de log-vraisemblance est :

$$L(\beta) = \ln(V(\beta)) = \sum_{i=1}^n y_i \ln\left(\frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}\right) + (1 - y_i) \ln\left(\frac{1}{1 + \exp(x_i^T \beta)}\right) = \sum_{i=1}^n y_i x_i^T \beta - \ln(1 + \exp(x_i^T \beta))$$

On dérive en  $\beta$  et on cherche le  $\beta$  tel que :

$$\frac{\delta L(\beta^{opt})}{\delta \beta} = 0 \Leftrightarrow \sum_{i=1}^n y_i x_i - x_i \exp\left(\frac{(x_i \beta^{opt})}{1 + \exp(x_i \beta^{opt})}\right) = 0$$

Cependant cette équation n'est pas solvable analytiquement. On utilise la méthode de Newton Raphson de descente de gradient pour trouver une solution à ce système. Cette méthode consiste à calculer une suite de  $\beta$  qui converge vers  $\beta^{opt}$ .

Cette suite est calculée par l'itération :

$$\beta^{(t)} = \beta^{(t-1)} - \left[\frac{\delta L^2}{(\delta \beta)^2}(\beta^{(t-1)})\right]^{-1} \frac{\delta L}{\delta \beta}(\beta^{(t-1)})$$

De plus,  $\frac{\delta L^2(\beta)}{\delta \beta} = -\sum_{i=1}^n \frac{\exp(x_i^T \beta)}{(1 + \exp(x_i^T \beta))^2} x_i x_i^T$  est définie négative pour tout  $\beta \in \mathbb{R}^p$ , donc  $\beta$  maximise bien la log-vraisemblance. (Marin)

Le GAM est une forme de généralisation du GLM, c'est le même fonctionnement que le GLM excepté que  $g(E(Y)) = \beta_0 + \sum_{i=1}^m f_i(x_i)$  ou  $f_i(x) = \sum_k \beta_{ik} b_{ik}(x)$  et  $g$  est la fonction de lien.

l'algorithme de Backfitting et l'algorithme général local de scoring sont généralement utilisés afin de déterminer les paramètres. (Friedman et al., 2000; Stéphane, 2012)

Le principe de la classification par forêt aléatoire est la construction de 500 arbres aléatoires avec des sous-échantillons (de taille 0.632 fois le nombre d'individus ; Breiman, 2001) ainsi que sur des variables explicatives tirées de façon aléatoire (si leurs nombres est supérieur 1, ce qui n'est pas notre cas). Puis vote par moyenne des 500 arbres.

### 2.2.2. Étude des erreurs de détections dans les déclarations de bord

Afin de pouvoir étudier si le nombre d'activités ainsi que les poids des activités pour les cuves multicalées a une importance pour l'erreur de détection ainsi que pour les biais de composition, nous avons utilisé un critère de «structure de distribution des calées ». Nous avons utilisé l'indice d'équitabilité de Pielou species evenness, qui est un indice de diversité dérivé de l'indice de Shannon, et détermine le degré d'équilibre d'abondance entre les espèces d'une population. Il permet par exemple de quantifier la similarité de structure des populations. Elle varie entre 0 et 1, tend vers 0 quand la quasi-totalité des effectifs est concentrée sur une espèce ; elle est de 1 lorsque

toutes les espèces ont même abondance. Il est défini comme  $J' = \frac{H'}{H'_{max}}$  ou  $H'$  est l'indice de diversité de Shannon et  $H'_{max} = \ln(n_i)$  est le maximum de l'indice de Shannon pour une population  $i$  de taille  $n_i$ . On rappelle que l'indice de Shannon est défini comme :

$$H' = - \sum_{i=1}^R p_i * \ln(p_i)$$
 ou  $R$  est la richesse spécifique (par exemple les espèces d'un environnement) et  $p_i$  est la proportion de l'espèce  $i$  dans la population totale  $R$ . Dans notre cas d'étude, nous avons appliqué cet indice à chaque cuve en regardant l'équitabilité des calées la composant. Les cuves ayant une équitabilité proche sont similaires en composition de calées.

Pour étudier la fiabilité des déclarations de bord, nous avons utilisé uniquement les données où l'espèce a été détectée dans la cuve par l'échantillonnage. Pour cette étude, nous avons utilisé un modèle linéaire généralisé mixte avec comme variable réponse la variable indicatrice de présence d'une espèce dans la cuve selon la déclaration de bord et comme variables explicatives fixes le type de banc en interaction avec l'espèce, l'océan en interaction avec l'espèce, le pourcentage dans l'échantillon par espèces, l'année, le nombre d'activités et le critère de structure de distribution des calées. Le facteur bateau était la variable aléatoire afin de prendre en compte la dépendance des données.

Nous avons choisi le modèle linéaire généralisé mixte, car nous supposons que  $Y_i^l = 1(x_i^l > 0)$  suit une loi binomiale, mais que les variables ne sont pas indépendantes.

### 2.2.3. Biais de compositions

Pour l'étude des biais de déclaration en composition spécifique, nous avons choisi de faire une première analyse sur la composition spécifique et une seconde analyse sur la composition de poids. De plus, pour des raisons de dynamique écologique différente nous avons séparé chaque analyse par leurs types de bancs. Nous avons donc pour chaque composition (spécifique et de poids) une analyse pour les biais de composition pour les cuves de bancs objet et une analyse pour les cuves de bancs libres.

Pour ce faire, nous avons utilisé un modèle linéaire mixte avec comme variable réponse la différence de composition spécifique (entre échantillon et déclaration de bord) pour les cuves de bancs libres dans un premier cas, pour les cuves de banc objet dans un deuxième cas. Puis la différence de composition en poids pour les bancs libres dans un troisième cas et pour les bancs objets dans le dernier cas. Les variables explicatives fixes étaient l'espèce, l'océan, l'année, le nombre d'activités et le critère de spécificité des calées.

Pour définir les différences de compositions, on note :  $P_{SKJ}^e, P_{YFT}^e$  et  $P_{BET}^e$  les proportions des espèces dans une cuve mesurées par un échantillonnage et  $P_{SKJ}^l, P_{YFT}^l$  et  $P_{BET}^l$  les proportions des espèces dans une cuve déclarée par le personnel de bords (cf. partie taux d'erreur et détermination des seuils de détection). On note l'erreur de compositions de la déclaration de la cuve pour une espèce par :  $\Delta_{SKJ} = P_{SKJ}^l - P_{SKJ}^e, \Delta_{YFT} = P_{YFT}^l - P_{YFT}^e, \Delta_{BET} = P_{BET}^l - P_{BET}^e$

$$\Delta_{SKJ}, \Delta_{YFT}, \Delta_{BET} \in [-1, 1]$$

On définit un individu comme précédemment, c'est-à-dire une espèce d'une cuve. On note donc  $Y_i = \Delta_i, i \in n$  l'erreur de déclaration d'une espèce d'une cuve.

Contrairement au modèle linéaire classique où les variables sont indépendantes :  $Y \sim N(X\beta, \sigma^2)$  le modèle linéaire mixte ne considère pas cette indépendance,  $Y \sim N(X\beta, \Sigma)$  ou,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix} = \sigma_a^2 + \sigma^2$$

Le modèle s'écrit :

$$Y_{ij} = \mu + A_i + E_{ij}; A_i \sim N(0, \sigma_a^2), E_{ij} \sim N(0, \sigma^2)$$

Les paramètres du modèle sont trouvés par maximum de vraisemblance restreint (REML)

(Ballesteros, 2008; Bates et al., 2014; Pinheiro and Bates, 2006)

Nous utiliserons le package lme4

Pour les sélections de variables du/es modèles, nous avons choisis les modèles qui minimise le critère d'information bayésien.

$$BIC = -2 \ln(L) + k * \ln(n)$$

L: la vraisemblance du modèle

n : le nombre d'observations

k : le nombre de paramètres libres

Le modèle BIC pénalise davantage le surparamétrage que le critère d'information d'Akaike ( $AIC = -2 \ln(L) + 2 * k$ ). N'ayant pas trouvé d'algorithme de sélection de modèles pour les

analyses linéaires généralisées mixtes, nous testerons un par un les modèles qui nous semblent potentiellement explicatifs.

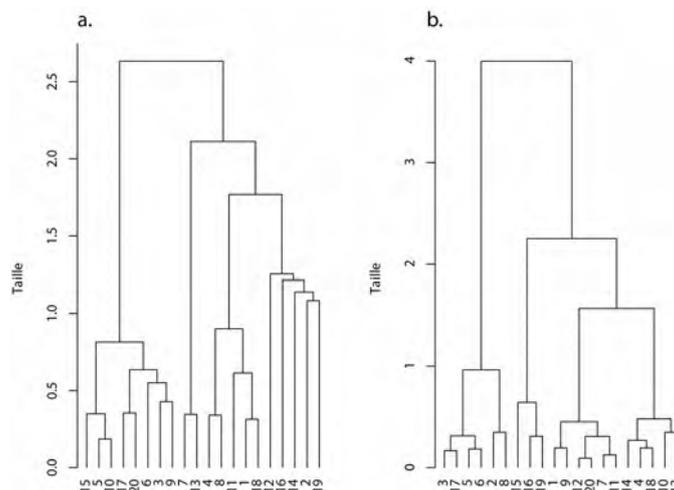
#### **2.2.4. Étude du delta en fonction de proportion de l'espèce contenu dans l'échantillon**

Pour étudier la dépendance des erreurs de composition spécifique avec la proportion de l'espèce dans la cuve, nous avons d'abord fait une analyse succincte à l'aide de boîtes à moustaches en séparant les proportions d'espèces dans la cuve en 10 groupes (de 0 à 1 par 0,1) puis nous avons utilisé un modèle additif généralisé mixte afin d'observer l'influence de la proportion d'une espèce par type de banc sur les biais de composition. Nous avons fait le choix d'utiliser ce modèle, car nos observations ne sont pas indépendantes pour les mêmes raisons que pour l'étude des biais de compositions précédents. Nous utiliserons le package gamm4 (Wood, 2006).

### 3. Résultats

#### 3.1. Classification des calées par leurs compositions spécifiques et par leurs compositions en classe de poids

Les résultats de la CAH sur les centroïdes des 20 groupes déterminés par k-means suggèrent un découpage en 2 groupes pour les 2 niveaux de description des captures (composition spécifique et classe de poids, Figure 1).



**Figure 1:** Dendrogramme des classifications.

a: Dendrogramme des groupes de compositions spécifiques.

b: Dendrogramme des groupes de compositions par catégories de poids.

Les 2 groupes créés par la classification sur la composition spécifique et la composition par classe de poids se confondent à près de 95 % démontrant une discrimination stable des groupes (Tableau 1). En effet, les premiers groupes des 2 CAH sont caractérisés par une dominance de SKJ (Pourcentages supérieurs à 80 %) et le second groupe par une dominance de YFT (Pourcentages supérieurs à 70 %; annexe 6 et 7). Le BET n'est que peu discriminant (proche de 5 % dans les groupes 1 et 2). Ces 2 groupes reflètent donc majoritairement les bancs sous objets, pour le groupe 1, les bancs libres de YFT, pour le groupe 2. La comparaison avec les déclarations des types de bancs dans les livres de bord montre que les classifications ont une précision de près de 80 % pour tous les types de bancs.

En revanche la sensibilité est meilleure pour les bancs libres que les bancs objet (environ 82 % et 76 % respectivement).

**Tableau 1:** Matrice de confusion entre les groupes créée par la classification des compositions de classes de poids et ceux des compositions spécifiques.

**Tableau 2:** Matrice de confusion entre les groupes créée par la classification des compositions de classes de poids et ceux déjà existants du type de banc.

**Tableau 3:** Matrice de confusion entre les groupes créée par la classification des compositions spécifiques et ceux déjà existants du type de banc.

	Espèce 1	Espèce 2			Banc objet	Banc libre			Banc objet	Banc libre	
Catégorie 1	38851	1591	0,961	Catégorie 1	32282,000	8160,000	0,798	Espèce 1	32159,000	8959,000	0,782
Catégorie 2	2267	32768	0,935	Catégorie 2	7196,000	27839,000	0,795	Espèce 2	7319,000	27040,000	0,787
	0,945	0,954	0,949		0,818	0,773	0,797		0,815	0,751	0,784

Sur la sélection des cuves « homogènes » : 13 177 cuves ont été échantillonnées. Avec les conditions d'homogénéité (selon la composition par catégorie d'espèces) des calées dans la cuve, le nombre de cuves restant est de 7979, soit près de 53 % des cuves échantillonnées. Le nombre de cuves homogènes est de 7909 si l'on utilise la classification par la composition spécifique.

## 3.2. Taux d'erreurs et comparaison des seuils de détections

### 3.2.1. Taux d'erreurs

Sur les 22 659 déclarations d'espèces, nous avons dénombré 5162 erreurs :

- 4741 sont de fausses absences dans les déclarations de bord selon l'échantillon, dont 68,8 % sont des déclarations de captures sous banc objet.
- 421 sont de fausses absences dans l'échantillon selon les déclarations de bord, dont 34 % sont des déclarations de capture sous banc objet.

### 3.2.2. Comparaison des seuils de détection

#### 1. Seuil de détection d'une espèce dans les déclarations de bord en fonction du pourcentage de cette espèce dans l'échantillon

Selon les résultats issus des validations croisées sur les différents modèles, le modèle minimisant l'erreur quadratique moyenne est le GAM (Tableau 4).

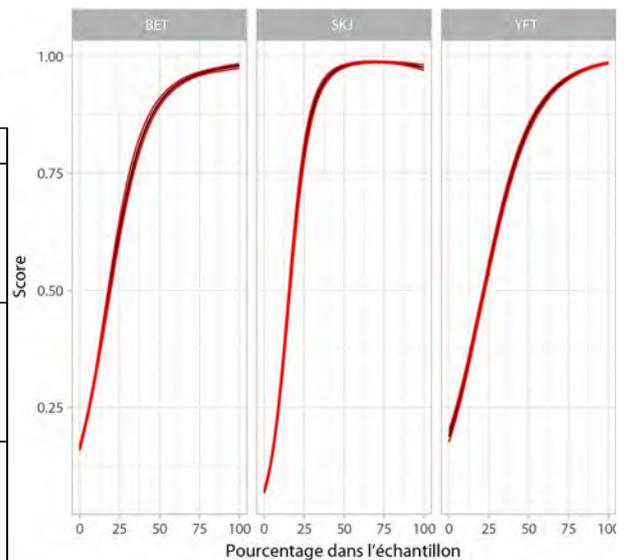
**Tableau 4:** Erreurs quadratiques moyennes des modèles GLM, GAM et forêt aléatoire pour la prédiction de détection du livre de bord. Calculée par validation croisée

	EQM
GLM	0,092
GAM	<b>0,086</b>
RF	0,103

Pour une présence à 25 % dans la cuve selon l'échantillon, le score de détection moyen par les déclarations de bord est compris entre 0,536 et 0,559 pour le YFT, 0,609 et 0,650 pour le BET et entre 0,775 et 0,812 pour le SKJ (Tableau 5). Le SKJ est l'espèce la plus facilement détectable selon la modélisation. Les scores de détection augmentent avec le pourcentage présumé de l'espèce dans la cuve selon l'échantillon (Figure 2).

**Tableau 5:** Synthèse des moyennes et quantiles du score de détection des déclarations de bord en fonction du pourcentage de l'espèce dans l'échantillon.

Pourcentage	Espèce	Score moyen	Quantile 0,025	Quantile 0,975
25,000	YFT	0,549	0,536	0,559
50,000	YFT	0,847	0,837	0,855
75,000	YFT	0,954	0,950	0,957
90,000	YFT	0,977	0,975	0,979
25,000	BET	0,627	0,609	0,650
50,000	BET	0,900	0,893	0,909
75,000	BET	0,962	0,958	0,966
90,000	BET	0,974	0,969	0,978
25,000	SKJ	0,793	0,775	0,812
50,000	SKJ	0,979	0,975	0,982
75,000	SKJ	0,987	0,986	0,989
90,000	SKJ	0,982	0,979	0,985



**Figure 2:** Moyenne et quantile du score de détection des déclarations de bord en fonction du pourcentage de l'espèce dans les cuves selon l'échantillon.

## 2. Seuil de détection d'une espèce dans l'échantillon en fonction du pourcentage de cette espèce dans les déclarations de bord

Selon les résultats issus des validations croisées sur les différents modèles, le modèle minimisant l'erreur quadratique moyenne est le GAM (Tableau 6).

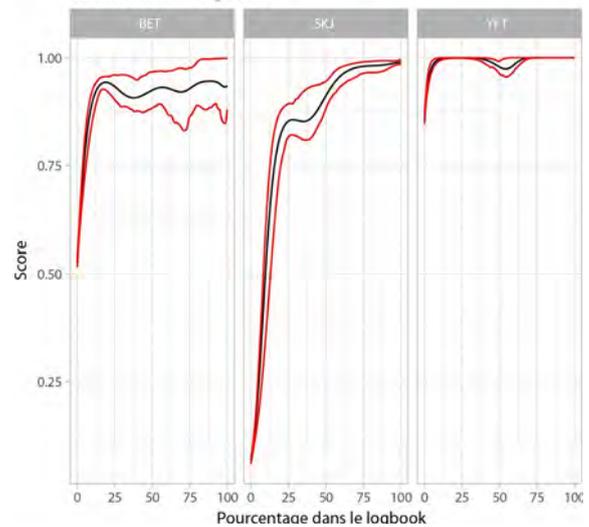
**Tableau 6:** Erreurs quadratiques moyennes des modèles GLM, GAM et forêt aléatoire pour la prédiction de détection de l'échantillon. Calculée par validation croisée.

	EQM
GLM	0,058
GAM	<b>0,053</b>
RF	0,060

Pour une présence à 25 % dans la cuve selon les déclarations de bord, le score de détection moyen de l'échantillon est de 1 pour le YFT, varie entre 0,90 et 0,96 pour le BET et varie entre 0.816 et 0,893 pour le SKJ (Tableau 7). Les scores de détection de l'échantillon sont plus forts que ceux des déclarations de bord pour des pourcentages dans les cuves faible. Les scores de détection augmentent avec le pourcentage présumé de l'espèce dans la cuve selon les déclarations de bord (Figure 3)

**Tableau 7:** Synthèse des moyennes et quantiles du score de détection des échantillons en fonction du pourcentage de l'espèce dans les déclarations de bord.

Pourcentage	Espèce	score	Quantile 0,025	Quantile 0,975
25,000	YFT	1,000	1,000	1,000
50,000	YFT	0,979	0,963	0,992
75,000	YFT	1,000	1,000	1,000
90,000	YFT	1,000	1,000	1,000
25,000	BET	0,933	0,902	0,960
50,000	BET	0,925	0,883	0,967
75,000	BET	0,928	0,864	0,977
90,000	BET	0,945	0,887	0,997
25,000	SKJ	0,853	0,816	0,893
50,000	SKJ	0,912	0,882	0,945
75,000	SKJ	0,980	0,964	0,990
90,000	SKJ	0,984	0,972	0,993



**Figure 3:** Moyenne et quantile du score des détections des échantillons en fonction du pourcentage de l'espèce dans les cuves selon les déclarations de bord.

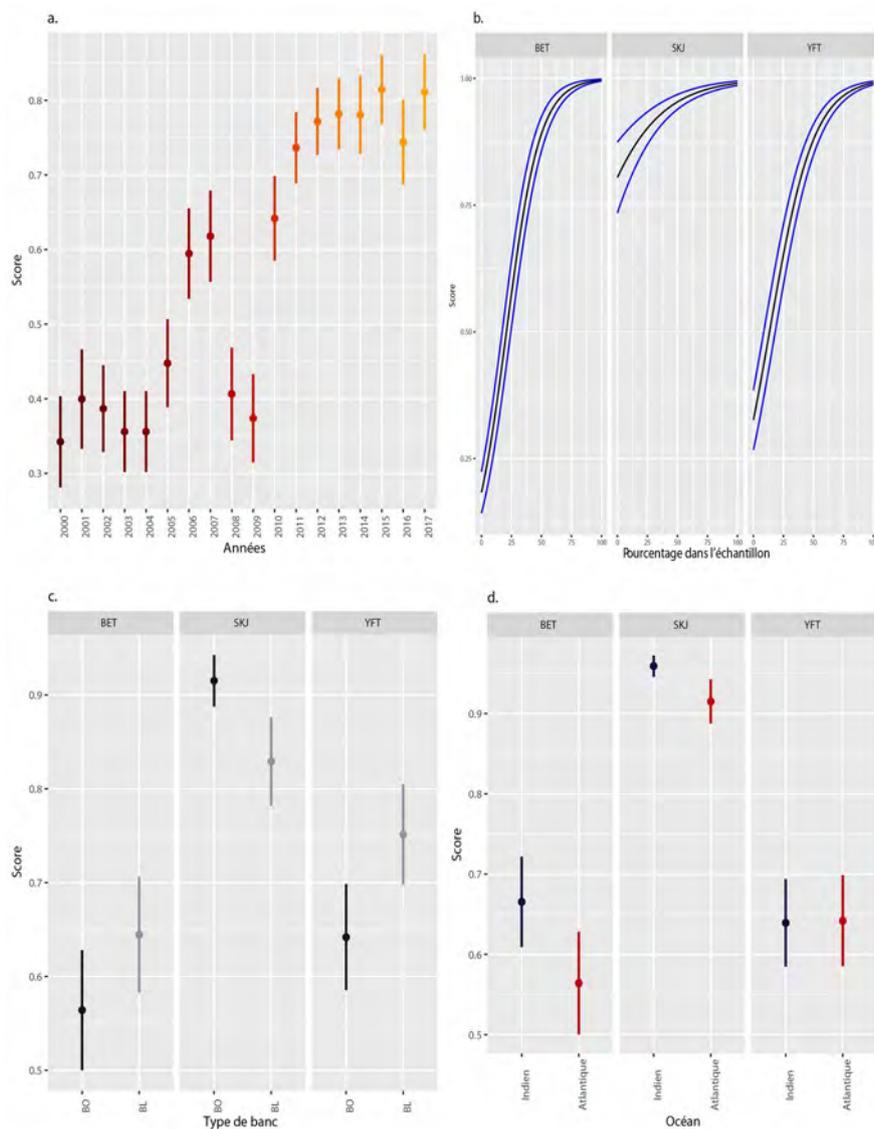
### 3.3. Étude des erreurs de détections dans les déclarations de bord

La restriction de variables par le critère BIC nous donne le modèle avec comme variables explicatives :

- L'interaction entre l'océan et les espèces.
- L'interaction entre le pourcentage présent dans la cuve (selon l'échantillon) et les espèces.

- L'année de pêche.
- L'interaction entre le type de banc sous lesquels sont pêchés les calées de la cuve et les espèces.

Ainsi, le score de détection s'est amélioré avec les années excepté pour les années 2008-2009 (Figure 4.a). Il augmente aussi avec le pourcentage de l'espèce dans la cuve (Figure 4.b). On remarque que le SKJ est l'espèce la plus facilement détectable (Figures 4.b, 4.c et 4.d). Cependant les erreurs de détection du BET et du SKJ sont plus faibles dans l'océan indien (Figure 4d). Et enfin, on remarque que le SKJ est mieux détecté sous bancs objets plutôt que sous bancs libres, à l'inverse le YFT et le BET dans une moindre mesure sont mieux détectés sous banc libre que sous bancs objets (Figure 4c).



**Figure 4:** Facteurs explicatifs du score de détection du livre de bord. Estimé à partir d'un modèle linéaire généralisé mixte.

a: Effet des années sur le score de détection.

b: Effet du pourcentage par espèces dans l'échantillon sur le score de détection.

c: Effet du type de banc par espèces sur le score de détection.

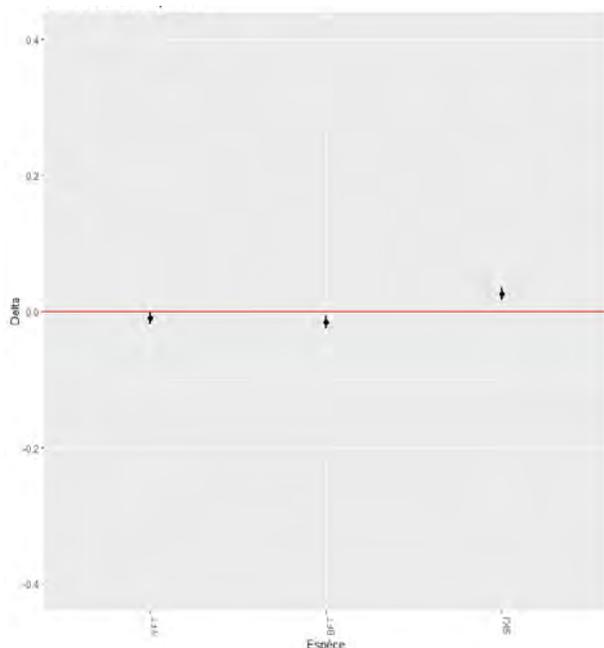
d: Effet de l'océan par espèces sur le score de détection.

### 3.4. Biais de compositions

#### 3.4.1. Étude des biais de déclaration en composition spécifique

##### Bancs libres

La sélection de modèles par le critère BIC nous donne comme meilleur modèle explicatif le modèle nulle, c'est à dire sans les variables explicatives l'année, le mois, l'océan, l'espèce et l'hétérogénéité de la cuve (Tableau 8). Le modèle de régression avec la variable explicative « Espèce » a été fait à titre indicatif (Figure 5 ; Annexe 2).



**Figure 5:** Prédiction et IC 95% des biais de déclaration entre échantillon et déclaration de bord par espèce sous bancs libres. Estimé à partir d'un modèle linéaire mixte.

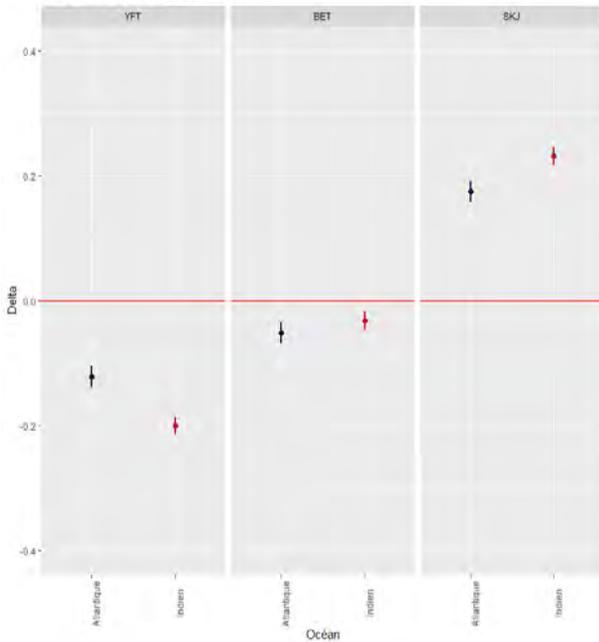
**Tableau 8:** Résultat de la sélection de modèles par critère BIC pour la régression linéaire mixte des biais de compositions spécifiques sur bancs libres. Nous avons affiché uniquement le premier ainsi que les derniers résultats.

Le modèle « complet » contient l'interaction entre l'espèce et l'océan, l'année et le mois. Il contient aussi l'hétérogénéité de la cuve.

	BIC
Modèle « complet »	-14089
Espèce*Océan+Années	-15085
Espèce*Océan	-15407
Espèce + Océan	-15438
Espèce	-15457
Nulle	-15466

##### Bancs objet

Les biais de déclarations concernant les cuves pêchées sous banc objet varient en fonction de l'espèce et de l'océan (Tableau 9). Le SKJ a été fortement surestimé (une prévision de  $0,17 \pm 0,02$  pour l'océan Atlantique et  $0,23 \pm 0,01$  de différence entre la proportion de SKJ indiqué dans l'échantillon et celle dans la déclaration de bord, Annexe 3) tandis que le YFT est fortement sous-estimé ( $-0,12 \pm 0,02$  en Atlantique et  $-0,2 \pm 0,01$ , Annexe 3). Le BET a été aussi surestimé, mais dans une moindre mesure. Les biais de composition spécifique étaient plus importants pour les déclarations dans l'océan Indien pour les YFT et SKJ mais pas pour le BET (Figure 6).



**Figure 6:** Prédiction et IC 95% des biais de déclaration entre échantillon et déclaration de bord par espèce et océan sous bancs objets. Estimé à partir d'un modèle linéaire mixte.

**Tableau 9:** Résultat de la sélection de modèles par critère BIC pour la régression linéaire mixte des biais de compositions spécifiques sur bancs objet. Nous avons affiché uniquement le premier ainsi que les derniers résultats.

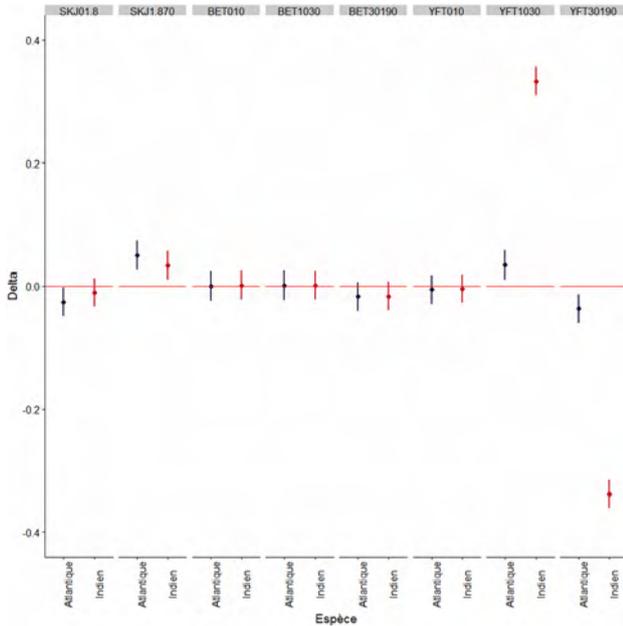
Le modèle « complet » contient l'interaction entre l'espèce et l'océan, l'année et le mois. Il contient aussi l'hétérogénéité de la cuve.

	<b>BIC</b>
<b>Modèle « complet »</b>	-1861
<b>Espèce*Océan+Années</b>	-2565
<b>Espèce*Océan</b>	-2863
<b>Espèce + Océan</b>	-2820
<b>Espèce</b>	-2837
<b>Nulle</b>	-2595

### 3.4.2. Étude des biais de déclaration en composition par catégories de poids.

#### Bancs libres

Les biais de compositions pour les catégories de poids concernant les cuves pêchées sous banc libre ont varié en fonction de l'espèce et de l'océan (Tableau 10). Dans l'océan Indien, le YFT de classe 10-30kg a été fortement surestimé ( $0,33 \pm 0,03$  de différences entre la proportion de YFT indiqué dans l'échantillon et celle dans la déclaration de bord) tandis que la classe de YFT de poids supérieur à 30kg semble sous-estimé dans les mêmes proportions ( $-0,34 \pm 0,02$  ; annexe 4). Dans l'océan Atlantique, on retrouve le même schéma de sous-déclaration de YFT de classe intermédiaire et sous-déclaration de gros YFT mais dans une proportion beaucoup plus faible ( $-0,03 \pm 0,03$ ). Les gros SKJ (supérieur à 1.8kg) ont aussi été légèrement surdéclarés par rapport aux échantillons sans toutefois de différence entre les deux océans (Figure 7).



**Figure 7:** Prédiction et IC 95% des biais de déclaration entre échantillon et déclaration de bord par classes de poids et océans sous bancs libres. Estimés à partir d'un modèle linéaire mixte.

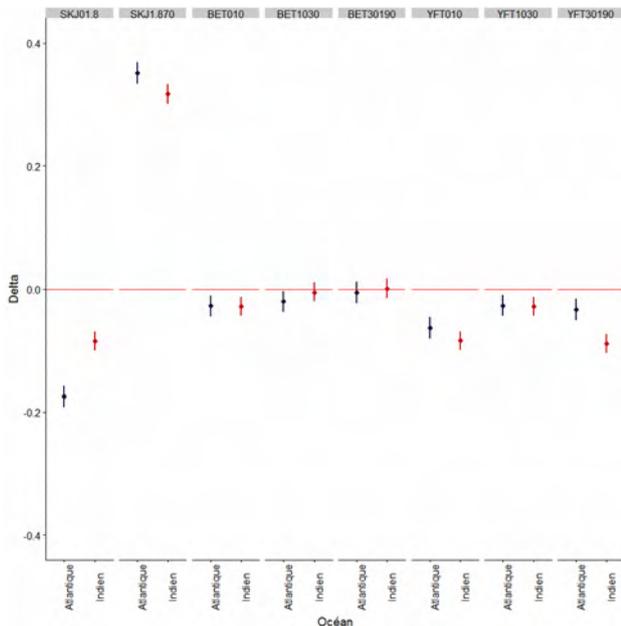
**Tableau 10:** Résultat de la sélection de modèles par critère BIC pour la régression linéaire mixte des biais de compositions en classes de poids sur bancs libres. Nous avons affiché uniquement le premier ainsi que les derniers résultats.

Le modèle « complet » contient l'interaction entre l'espèce et l'océan, l'année et le mois. Il contient aussi l'hétérogénéité de la cuve.

	BIC
Modèle « complet »	-12700
Espèce*Océan+Années	-13146
Espèce*Océan	-14024
Espèce + Océan	-13867
Espèce	-13887
Nulle	-13739

## Bancs objet

Les biais de compositions pour les catégories de poids concernant les cuves pêchées sous banc objet varient également en fonction de l'espèce et de l'océan (Tableau 11). Les gros SKJ (supérieur à 1.8kg) ont été largement surévalués (Figure 8), d'autant plus dans l'océan Atlantique ( $0,35 \pm 0,02$  ; annexe 5). À l'inverse, toutes les classes de YFT ainsi que les petits SKJ et dans une moindre mesure les petits BET ont été surdéclarés. L'océan dans lequel la capture a eu lieu influence de manière variable le biais de déclaration : tantôt supérieur dans l'océan atlantique (petits et gros YFT), tantôt inférieur (petit SKJ).



**Figure 8:** Prédiction et IC 95% des biais de déclaration entre échantillon et déclaration de bord par classes de poids et océans sous bancs objets. Estimés à partir d'un modèle linéaire mixte.

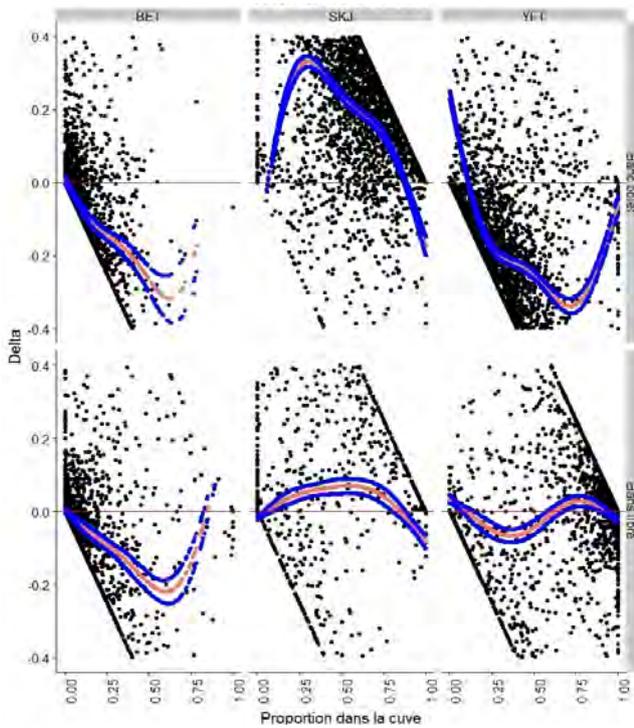
**Tableau 11:** Résultat de la sélection de modèles par critère BIC pour la régression linéaire mixte des biais de compositions en classes de poids sur bancs objet. Nous avons affiché uniquement le premier ainsi que les derniers résultats.

Le modèle « complet » contient l'interaction entre l'espèce et l'océan, l'année et le mois. Il contient aussi l'hétérogénéité de la cuve.

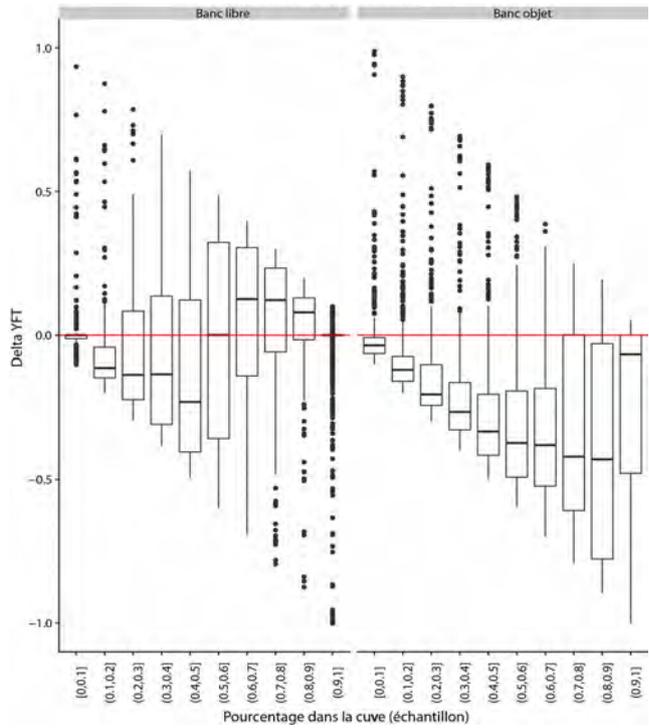
	BIC
Modèle « complet »	-18833
Espèce*Océan+Années	-20900
Espèce*Océan	-21242
Espèce + Océan	-21221
Espèce	-21230
Nulle	-20804

### 3.5. Étude du delta en fonction de proportion de l'espèce contenu dans l'échantillon.

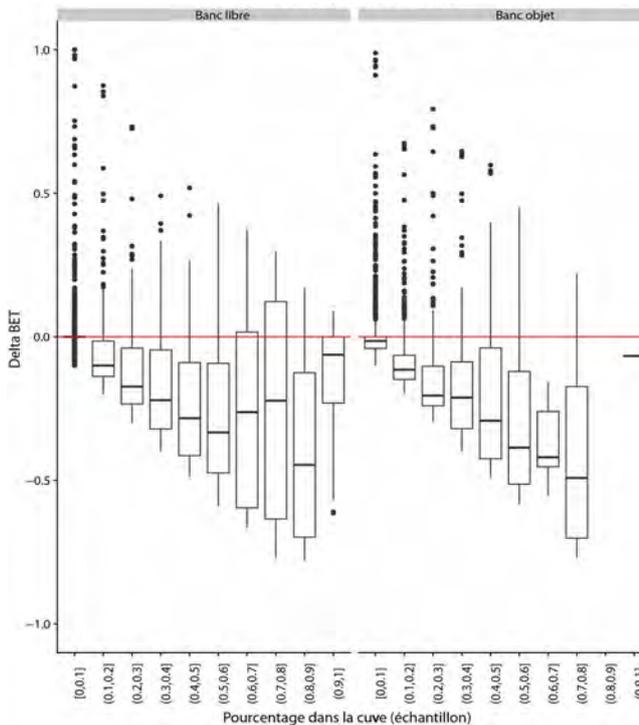
Nous avons observé (Figure 9, 10, 11 et 12) que les erreurs de déclarations des espèces sont en moyenne plus forte lorsque leurs fréquences dans les cuves sont dans la gamme intermédiaire (entre 0.2 et 0.8). Dans cette gamme le SKJ a été surdéclaré et le BET et YFT sous-déclarés quelques soit leur fréquence dans la cuve sous bancs objets, à l'exception des fréquences que le BET n'atteint pas ou trop peu. Lorsque les espèces sont peu ou très abondantes, les biais de déclaration diminuent ainsi que leur variance. Nous pouvons noter que les BET ont été fortement sous-déclarés même dans les gammes fortes de fréquence, contrairement au YFT sous banc libre surdéclaré en moyenne dans les gammes 0.6-0.9.



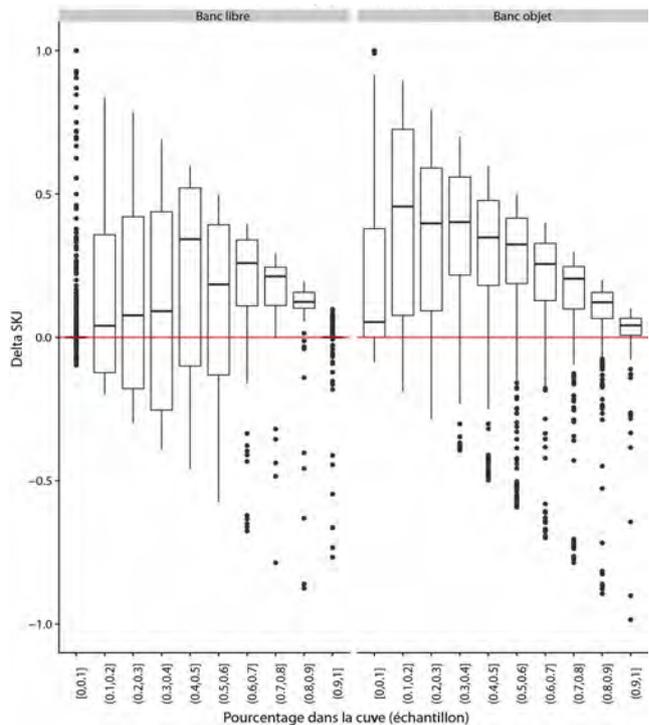
**Figure 9:** Prédictions et IC 95% des biais de déclaration entre échantillon et déclaration de bord par espèces et types de bancs. Estimés à partir d'un modèle additif mixte.



**Figure 10:** Biais de composition par type de banc pour le YFT en fonction du groupe de proportion dans la cuve.



**Figure 11:** Biais de composition par type de banc pour le BET en fonction du groupe de proportion dans la cuve.



**Figure 12:** Biais de composition par type de banc pour le SKJ en fonction du groupe de proportion dans la cuve.

## 4. Discussion

Suite à une sélection statistique des cuves basée sur de nombreux critères d'homogénéité des calées, nous avons mis en évidence que les groupes issus des compositions (spécifique) coïncidaient à 80 % avec les types de bancs. Cependant nous avons perdu 47 % de notre jeu de données dans ce processus de sélection. Nous avons ensuite démontré que le nombre d'erreurs de déclaration ainsi que le seuil de détection des espèces sont plus importants pour les déclarations des livres de bord que pour les échantillons. Puis, nous avons montré que les erreurs de détections des déclarations des livres de bord sont influencées par les facteurs suivant : le type de banc, l'année, l'océan, la présence dans la cuve de l'espèce ainsi que l'espèce elle-même. Nous avons retrouvé une partie de ces facteurs : type de banc, océan et espèces ainsi que les catégories de poids dans l'étude des biais de compositions entre déclarations des livres de bord et les échantillons. Enfin nous avons remarqué que sur ce jeu de données, une proportion intermédiaire des espèces dans les cuves augmente les biais de composition.

### 4.1. Des critères d'échantillonnages aux cuves homogènes

Nous avons procédé à une sélection des cuves déclarées basée sur leurs types de bancs, leurs compositions et leurs dates de pêche, facteurs que nous pensons importants pour s'assurer de l'homogénéité des cuves et ainsi obtenir une bonne représentativité des échantillons de celles-ci. Cependant après sélections des cuves (contenant des calées proches en aire géographique, pêchés sur le même type de banc et ayant une composition similaire) près de la moitié des cuves échantillonnées se sont révélées trop hétérogènes pour être conservées. Ce résultat s'explique par le que nos critères d'homogénéité étaient plus contraignants que ceux utilisés lors de l'échantillonnage. Ce travail démontre la nécessité de remettre en question certains points de l'échantillonnage, notamment vis-à-vis de la sélection des échantillons. Ce type de réflexions a déjà été initié, notamment à travers le projet européen MARE/2016/22 (Strengthening regional cooperation in the area of fisheries data collection ; Ruiz & al., 2019). Par exemple, il serait judicieux de mettre en place des analyses antérieures à l'échantillonnage, tel que celles appliquées dans ce travail, afin d'améliorer la récolte de données.

On peut également s'interroger sur les déclarations du type de banc sous lequel est effectuée la pêche. En effet, nous avons remarqué que les groupes créés par notre classification en fonction de la composition coïncidaient à près de 80 % avec les types de bancs déclarés. Donc environ 20 % ne correspondent pas. Par exemple, il est possible qu'une partie des calées dans le groupe à majorité de bancs objets soit pêchés sous des bancs libres de SKJ . Les groupes sont classifiés par la présence d'une majorité de SKJ pour le groupe à majorité banc objet (annexe 6). Une autre hypothèse pouvant expliquer en partie le décalage entre les compositions et le type de banc viendrait de la difficulté pour l'équipage de déterminer le type de banc. Un banc objet est défini par la présence d'un objet flottant. Or il peut s'avérer qu'un banc libre passe à proximité d'un objet flottant ou bien qu'un banc objet s'est éloigné de tout objet flottant. Dans ces deux cas de capture, cela peut amener l'équipage à faire une erreur de déclaration. En outre, il n'est pas établi de norme de distance du banc aux DCPs pour lequel la pêche est déclarée comme pêche sous banc objet. Une étude récente sur l'impacte des DCP et plus généralement des bancs objets dans l'océan Indien a montré que la distance moyenne entre bancs objets avoisinent les 20 km (Dagorn et al., 2013). L'interprétation du

type de banc devient alors difficile, car certaines espèces, dont les thons sont attirés par ces objets. On estime ainsi l'attraction de l'objet sur les YFT jusqu'à une distance de 10 km (Girard et al., 2004). Néanmoins, le type de banc reste un facteur important rendant compte de conditions écologiques bien distinctes (Fréon and Dagorn, 2000). Dans le cadre de modélisation à but prédictif, la connaissance précise de ce critère peut donc être importante. Ce travail soulève donc la question d'utiliser les classes de type de banc corrigées par classification supervisée basée sur la composition spécifique à la place des déclarations soumises aux erreurs de jugement.

## **4.2. Les seuils de détection d'espèces et les erreurs de détection des déclarations de bord**

En supposant les déclarations de bords justes, on dénombre 278 erreurs de détection pour les 22659 cuves échantillonnées, soit en moyenne 24 par années. Cela peut être dû à l'échantillon lui-même ou bien à des erreurs liées à l'échantillonnage (erreurs de saisie par exemple). À l'inverse, en supposant les échantillons justes, on trouve 4741 erreurs dans les déclarations de bord. Ces déclarations se basant sur une interprétation furtive et visuelle des espèces prises, il n'est pas anormal que le nombre d'erreurs soit plus élevé que pour les échantillons. Il n'est pas étonnant non plus de voir que 68,8% des erreurs de détections dans les déclarations de bord sont sur des cuves contenant des calées pêchées sous banc objets, la composition y étant généralement plus variée (Fonteneau, 2008) et donc plus complexe à estimer.

La détection d'espèce par l'échantillon est plus efficace que la détection par la déclaration du livre de bord surtout pour de petits pourcentages d'espèces présents dans la cuve (partie 3.2.2). Cette différence s'estompe avec l'augmentation de la proportion de l'espèce. Cela est dû au fait que l'équipage a plus de difficulté à détecter les espèces en faible abondance. Cela corrobore l'hypothèse selon laquelle les déclarations de bords sont moins fiables que les déclarations issues des échantillons.

Cependant les raisons des erreurs de détection dans les déclarations de bord sont diverses. Elles varient en fonctions de plusieurs variables : l'espèce, le type de banc, l'océan et l'année. Concernant les SKJ, sa morphologie est sensiblement différente du YFT et du BET de petite taille présent en banc objet et l'on peut remarquer facilement ses 3 à 5 rayures sur le ventre que les autres espèces n'ont pas (Damiano, 2011). Cela pourrait expliquer en partie la surestimation observée du fait de sa facilité de détection. On peut également remarquer que le score de détection est meilleur sous banc libre que sous banc objet pour les BET et les YFT, cela peut être dû aux faits que ce sont les espèces "attendues" dans un banc libre (Fonteneau, 2008). À l'inverse le SKJ qui n'est pas une espèce souvent présente (ils ne forment que rarement des bancs libres) souffre d'une détection plus mauvaise pour les calées de banc libre que pour les calées de banc objet. De plus, nous avons observé que la détection par les déclarations de bord est meilleure d'année en année excepté en 2008-2009. Cette amélioration peut en partie être due au temps d'adaptation des membres d'équipage aux améliorations technologiques (Dueri et al., 2018; Torres-Irineo et al., 2013). Nous avons remarqué qu'il existe des différences entre les scores de détections dans les deux océans (dans l'océan Indien, les scores de détection des BET et SKJ sont meilleurs que ceux dans l'Océan Atlantique). Ces effets entre océans sont complexes à expliquer, car multifactoriels, par exemple les conditions océanographiques, climatiques ainsi que les pratiques de pêche (exemple : les anomalies

de réchauffement des températures de surface ; Cane, 1983). Des analyses complémentaires sont nécessaires pour comprendre cette problématique qui dépasse le cadre de cette étude.

Un autre facteur que l'on doit prendre en compte, mais que l'on ne peut pas estimer est le fait que l'absence d'erreur de déclaration d'une cuve n'implique pas l'absence d'erreur de déclarations pour la déclaration de bord. En effet, les déclarations de bords sont faites par calées alors que notre étude porte sur les erreurs de déclarations dans les cuves. Or toutes les cuves ne sont pas monocalées. Concernant les multicalées, la présence d'une espèce dans une seule calée, selon la déclaration de bord, suffit pour que l'espèce soit présente dans la cuve. Ainsi, cela ne compte pas comme une erreur de détection même si l'espèce était en fait absente dans les autres calées. De ce point de vue, notre étude est donc plutôt conservatrice, car les erreurs de détection sont potentiellement sous-estimées.

### **4.3. Les biais de compositions**

Il faut souligner que même si les erreurs de détections impliquent des biais de compositions, cela ne s'est pas totalement ressenti dans nos résultats sur les biais de composition. Le fait qu'une espèce ne soit pas détectée dans une cuve malgré sa présence n'est pas entièrement déterminant dans la présence de biais de détection, car, comme nous l'avons vu, les erreurs de détections surviennent plus facilement lorsque la proportion de l'espèce est faible dans la cuve. Donc cette non-détection à un impact moindre sur la composition.

Ces études nous ont permis de montrer que l'on retrouve sur toute la période étudiée une partie des biais détectés dans les déclarations de bord il y a plus de 30 ans. Les premières études avaient montré que les jeunes YFT et BET étaient déclarés en SKJ (Cayré, 1984). Notre étude confirme donc que pour la pêche sous banc objet, les petits BET (inférieurs à 10kg) ainsi que toutes les classes de YFT sont sous-déclarés au profit des gros SKJ (supérieur à 1,8kg). Cela nous a également confortés dans l'idée que la pêche sous bancs objet augmente les biais de composition spécifique. Idée que nous avons, car les bancs de poissons sous DCP sont de compositions plus variées que les bancs libres (Chavance et al., 2011; Kothias Amon et al., 1996). De plus les YFT et BET y sont de plus petites tailles et donc plus difficilement différenciables (Fonteneau et al., 2013; Fréon and Dagorn, 2000). En outre, ces études nous ont aussi permis d'identifier des biais de composition au sein des classes de poids des espèces (SKJ sur bancs objets, YFT sur bancs libres et notamment dans l'océan Indien), biais que l'on pourrait aussi expliquer en partie par le fait que les déclarations de bord soient une estimation visuelle et qu'estimer le poids d'un poisson n'est pas chose aisée. Mais ces biais pourraient aussi venir en partie du fait que les déclarations de bords sont en premier lieu faites en classe commerciales qui sont nombreuses et qui s'entrecoupent. Ces catégories sont larges (exemple : supérieur à 10kg) et ne déterminent pas suffisamment les poids des espèces, elles peuvent donc expliquer les biais de composition interspèces. Si les prix de vente sont les mêmes pour une catégorie quelque soit l'espèce (exemple : les poissons inférieurs à 1.8kg), les pêcheurs n'ont aucun intérêt à déterminer avec précision la composition spécifique et déclare l'ensemble sous un même libellé. Enfin, de la même manière que pour les biais de détection, le facteur océan a eu un effet différent sur les biais de composition. Des analyses complémentaires sont nécessaires afin de quantifier les possibles variations.

Cependant, il existe des biais dans cette analyse des biais de compositions. En effet, les conversions taille-poids sont anciennes et ne tiennent pas compte des variations géographiques et temporelles

(Duparc et al., 2018). Cela peut créer un décalage entre l'échantillon et les déclarations de bords. Cette étude nous interroge sur l'importance de mettre à jour les relations tailles-poids régulièrement ainsi que dans des zones géographiques éparses. De plus, une partie du YFT, BET et SKJ déclaré dans le livre de bord ne part pas aux conserveries, et est débarqué pour être vendu au marché local. Or l'échantillonnage ne concerne que les thonidés entrant aux conserveries. Ce décalage entre déclaration et échantillon pourrait expliquer les différences observées. Or Les estimations du flux de thonidés du marché local représente de 0,3 à 0,9 % du BET pêché lors de la marée, de 4 à 5,2 % pour le SKJ et de 1,3 à 1,8 % pour le YFT (Tableau 12). Elles ne peuvent donc pas expliquer l'ensemble des différences de compositions observées entre les déclarations de bord et les échantillons.

**Tableau 12:** Proportion avec IC95% de la quantité de pêche par marée et par espèces allant dans le marché local (Lesage, 2019).

	Proportion	IC 95 % min	IC 95 % sup
BET	0,004	0,003	0,009
SKJ	0,044	0,040	0,052
YFT	0,014	0,013	0,018

#### 4.4. Relation entre biais de composition et proportions spécifiques.

L'étude de l'erreur de composition en fonction de la proportion de l'espèce dans la cuve nous a montré que les erreurs de composition des espèces sont en moyenne plus fortes lorsque leurs fréquences dans les cuves sont dans la gamme intermédiaire. Ce qui confirme en partie l'hypothèse selon lequel il est difficile de distinguer une espèce particulière dans un mélange d'espèce et que cela entraîne des biais de compositions. Cela conforte aussi les résultats obtenus lors de l'étude de la composition spécifique : sous bancs objets, les SKJ sont surdéclarés et les BET et YFT sous-déclarés quelque soit leur fréquence dans la cuve. Nous retrouvons pour les YFT sous bancs libres l'idée qu'une espèce minoritaire est sous-déclarée tandis qu'une espèce majoritaire est surdéclarée, modèle que l'on ne retrouve pas pour les autres catégories. Pour étudier plus précisément les erreurs de composition en fonction des proportions d'espèces dans les cuves, nous aurions pu aussi effectuer une analyse par classes de poids, car comme nous l'a montré l'analyse en composition par classe de poids, il existe des différences de tendance entre les catégories des YFT sous banc libre et les SKJ sous banc objet.

## 5. Conclusion

Les premières études sur les biais de compositions dans les déclarations de bord ont montré que les BET et les YFT de petite taille étaient souvent déclarés comme des SKJ et que de plus, les petits BET et YFT sont difficilement différenciables (Bard, 1986; Cayré, 1984). Depuis les procédures d'échantillonnages ont été modifiées afin de tenir compte de l'évolution de la pêche à la senne due aux nouvelles technologies et en particulier à l'apparition des dispositifs de concentration de poissons (Fonteneau et al., 2000; Pallarés and Hallier, 1997a; Parajua and Hallier, 1992). Dans un premier temps nous avons montré que les déclarations de bord commettent des erreurs de détections d'espèce. Nous avons par la même occasion montré que plusieurs facteurs pouvaient influencer sur ces erreurs : le type de banc, l'océan, l'année et enfin en fonction de chaque espèce. Dans un second temps, nous avons étudié les biais de compositions entre les déclarations de bord et les échantillons faits à l'arrivée aux ports des navires. Lors de cette étude, nous avons retrouvé une partie des facteurs explicatifs des erreurs de détections : le type de banc, l'espèce et l'océan. Pour la composition spécifique nous avons remarqué que la pêche sous banc objet était un facteur amplificateur, voir déclencheur de biais de compositions, biais composé d'une surdéclaration de SKJ et donc de sous-déclarations de BET et YFT. Pour la composition par classes de poids, les biais de compositions sont présents pour les deux types de bancs. Le biais de composition pour les cuves de calées pêchés sur bancs libres est caractérisée par une surdéclaration des YFT de classe intermédiaire et une sous-déclaration des gros YFT, pour les cuves de calées pêchés sur bancs objets une surdéclaration des gros SKJ et une sous-déclaration des trois classes de YFT ainsi que de petites classes de BET et SKJ. Ce dernier résultat arrive donc dans la continuité des premières études sur les biais de compositions. Une étude sur les variations des biais entre les habitudes de pêche de pays différents aurait un intérêt certain. En ce sens, une demande de données concernant les déclarations de bord espagnoles a été transmise à l'Institut Océanographique espagnol.

## Références

- Ariz, J., Delgado, A., Fonteneau, A., Gonzales Costas, F., Pallarés, P., 1993. Logs and tunas in the eastern Tropical Atlantic. A review of present knowledges and uncertainties. ColVolSciPap ICCAT, SCRS/1992/039 40, 421–446.
- Ballesteros, S., 2008. Introduction au modèle linéaire mixte 57.
- Bard, F.X., 1986. Note sur l'échantillonnage pluridisciplinaire des thons tropicaux au port d'Abidjan -SCRS/1985/047. ColVolSciPap ICCAT, SCRS/1985/047 25, 375–45.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting Linear Mixed-Effects Models using lme4. ArXiv14065823 Stat.
- Bostrom, N., 2002. Anthropoc Bias: Observation Selection Effects in Science and Philosophy. Routledge. <https://doi.org/10.4324/9780203953464>
- Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bry, X., 2016. Classification automatique.
- Cane, M.A., 1983. Oceanographic Events During El Niño. Science 222, 1189–1195. <https://doi.org/10.1126/science.222.4629.1189>
- Carter, B., 1974. Large Number Coincidences and the Anthropic Principle in Cosmology, in: Longair, M.S. (Ed.), Confrontation of Cosmological Theories with Observational Data, International Astronomical Union / Union Astronomique Internationale. Springer Netherlands, Dordrecht, pp. 291–298. [https://doi.org/10.1007/978-94-010-2220-0\\_25](https://doi.org/10.1007/978-94-010-2220-0_25)
- Caverivière, A., 1976. Longueur prédorsale, longueur à la fourche et poids des albacores (Thunnus albacares) de l'Atlantique. Cah. ORSTOM Série Océan. 14, 201–208.
- Cayré, P., 1984. Procédure suivie pour la révision de la composition spécifique des statistiques thonières FISM (France, Côte d'Ivoire, Sénégal et Maroc), in: <title/>, Collective Volume of Scientific Papers - ICCAT. ICCAT, Madrid, pp. 102–107.
- Cayré, P., Laloë, F., 1986. Review of the Gonad Index (GI) and an introduction to the concept of its “critical value”: application to the skipjack tuna Katsuwonus pelamis in the Atlantic Ocean | SpringerLink [WWW Document]. URL <https://link.springer.com/article/10.1007/BF00428558> (accessed 8.16.19).
- Champagnat, C., Pianet, R., 1974. Croissance du patudo (Thunnus obsesus) dans les régions de Dakar et de Pointe Noire.
- Chassot, E., Assan, C., Soto, M., Damiano, A., Delgado de Molina, A., Joachim, L.D., Cauquil, P., Lesperance, F., Curpen, M., Lucas, J., Floch, L., 2015. Statistics of the European Union and associated flags purse seine fishing fleet targeting tropical tunas in the Indian Ocean 1981-2014, in: 17ème Groupe de Travail Sur Les Thons Tropicaux. p. 31 p.
- Chavance, P., Amon Kothias, J.B., Dewals, P., Pianet, R., Amandè, M.J., Delgado de Molina, A., Djoh, A., 2011. Statistics on tuna surface fishery's bycatch landed in Abidjan, Côte d'Ivoire, for the 1982-2009 period, in: Thunnus Alalunga (ALB), Collective Volume of Scientific Papers. pp. 2104–2112.
- Dagorn, L., Bez, N., Fauvel, T., Walker, E., 2013. How much do fish aggregating devices (FADs) modify the floating object environment in the ocean? Fish. Oceanogr. 22, 147–153. <https://doi.org/10.1111/fog.12014>
- Damiano, A., 2011. Guide d'identification des principales espèces marines destinés aux observateurs scientifiques.
- David, O., 2015. La population mondiale - 3e édition: Répartition, dynamique et mobilité. Armand Colin.
- Davies, T.K., Mees, C.C., Milner-Gulland, E.J., 2014. The past, present and future use of drifting fish aggregating devices (FADs) in the Indian Ocean. Mar. Policy 45, 163–170. <https://doi.org/10.1016/j.marpol.2013.12.014>

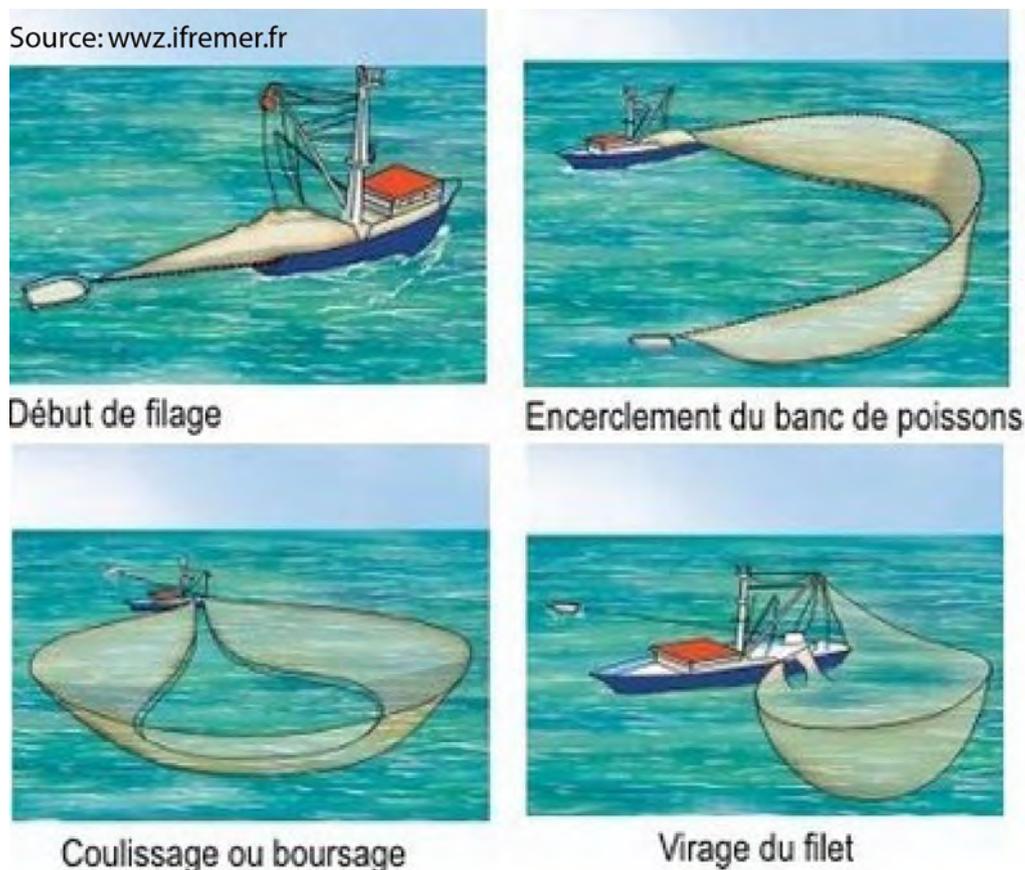
- Dueri, S., Bopp, L., Maury, O., 2018. Projecting the impacts of climate change on skipjack tuna abundance and spatial distribution. *Glob. Change Biol.* 20, 742–753. <https://doi.org/10.1111/gcb.12460>
- Duparc, A., Cauquil, P., Depetris, M., 2018. Assessment of accuracy in processing purse seine tropical tuna catches with the T3 methodology using French fleet data.
- FAO, 2018. La situation mondiale des pêches et de l'aquaculture.
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38–49. <https://doi.org/10.1017/S0376892997000088>
- Fonteneau, A., 2008. Species composition of tuna catches taken by purse seiners 14.
- Fonteneau, A., Ariz, J., Gaertner, D., Nordstrom, V., Pallares, P., 2000. Observed changes in the species composition of tuna schools in the Gulf of Guinea between 1981 and 1999, in relation with the Fish Aggregating Device fishery. *Aquat. Living Resour.* 13, 253–257. [https://doi.org/10.1016/S0990-7440\(00\)01054-8](https://doi.org/10.1016/S0990-7440(00)01054-8)
- Fonteneau, A., Chassot, E., Bodin, N., 2013. Global spatio-temporal patterns in tropical tuna purse seine fisheries on drifting fish aggregating devices (DFADs): Taking a historical perspective to inform current challenges★. *Aquat. Living Resour.* 26, 37–48. <https://doi.org/10.1051/alr/2013046>
- Fonteneau, A., Pallarés, P., 1991. Evolution récente des pêcheries thonières à la senne dans le Golfe de Guinée. *Rec Doc Sci ICCAT* 36, 427–478.
- Fréon, P., Dagorn, L., 2000. Review of fish associative behaviour: Toward a generalisation of the meeting point hypothesis. *Rev. Fish Biol. Fish.* 10, 183–207. <https://doi.org/10.1023/A:1016666108540>
- Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *Ann. Stat.* 28, 337–407. <https://doi.org/10.1214/aos/1016218223>
- Girard, C., Benhamou, S., Dagorn, L., 2004. FAD: Fish Aggregating Device or Fish Attracting Device? A new analysis of yellowfin tuna movements around floating objects. *Anim. Behav.* 67, 319–326. <https://doi.org/10.1016/j.anbehav.2003.07.007>
- Herrera, M., Carlos Báez, J., 2018. On the potential biases of scientific estimates of catches of tropical tunas of purse seiners monitored by european scientists and catches reported to the ICCAT and IOTC.
- Joseph, J., 1997. Atlas des pêcheries thonières tropicales-préface.
- Kothias Amon, J.B., Bard, F.X., Hervé, A., 1996. Mise à jour des quantités de Faux poissons débarquées par les senneurs à Abidjan. *Collect Vol Sci Pap ICCAT* 45, 227–228.
- Lesage, M., 2019. Etude de la filière du marché local dans le cadre de la pêche thonière tropicale par les senneurs dans l'océan Atlantique.
- Marin, J.-M., n.d. La régression logistique.
- Mcperson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *J. Appl. Ecol.* 41, 811–823. <https://doi.org/10.1111/j.0021-8901.2004.00943.x>
- O'Neill, R.N., 2011. Le thon rouge en miettes ! 24.
- Pallarés, P., Hallier, J.-P., 1997a. Analyse du schéma d'échantillonnage multispécifique des thonidés tropicaux 1 - rapport administratif (Rapport administratif No. Programme 95/37). IEO/ORSTOM, Madrid, Spain.
- Pallarés, P., Hallier, J.-P., 1997b. Analyse du schéma d'échantillonnage multispécifique des thonidés tropicaux 5- analyses statistiques (Rapport administratif No. Programme 95/37). IEO/ORSTOM, Madrid, Spain.
- Parajua, J.I., Hallier, J.-P., 1992. Western Indian Ocean purse seine fishery on yellowfin tuna : its specifications regarding log and free swimming school catches.

- Parks, W., Bard, X., Cayré, P., Kume, S., 1982. Length-weight relations for Bigeye Tuna captured in the eastern Atlantic.
- Pauly, D., Zeller, D., 2016. Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. *Nat. Commun.* 7, 10244. <https://doi.org/10.1038/ncomms10244>
- Pianet, R., Pallarés, P., Petit, C., 2000. New sampling and data processing strategy for estimating the composition of catches by species and sizes in the European purse seine tropical tuna fisheries 36.
- Pikitch, E.K., Santora, C., Babcock, E.A., Bakun, A., Bonfil, R., Conover, D.O., Dayton, P., Doukakis, P., Fluharty, D., Heneman, B., Houde, E.D., Link, J., Livingston, P.A., Mangel, M., McAllister, M.K., Pope, J., Sainsbury, K.J., 2004. Ecosystem-Based Fishery Management. *Science* 305, 346–347. <https://doi.org/10.1126/science.1098222>
- Pinheiro, J., Bates, D., 2006. *Mixed-Effects Models in S and S-PLUS*. Springer Science & Business Media.
- Restrepo, V., Dagorn, L., Ana, J.-R., 2017. Technical and Meeting Reports. *Int. Seaf. Sustain. Found.* URL <https://iss-foundation.org/knowledge-tools/technical-and-meeting-reports/> (accessed 7.4.19).
- Romagny, B., Ménard, F., Dewals, P., Gaertner, D., N’Goran, N., 2000. Le “faux-poisson” d’Abidjan et la pêche sous DCP dérivants dans l’Atlantique tropical Est : circuit de commercialisation et rôle socio-économique.
- Scott, G.P., Lopez, J., 2014. The use of FADS in tuna fisheries 32.
- Stéphane, T., 2012. *Data Mining et statistique décisionnelle: L’intelligence des données*. Editions TECHNIP.
- Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecol. Model.* 148, 1–13. [https://doi.org/10.1016/S0304-3800\(01\)00388-X](https://doi.org/10.1016/S0304-3800(01)00388-X)
- Syfert, M.M., Smith, M.J., Coomes, D.A., 2013. The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models. *PLOS ONE* 8, e55158. <https://doi.org/10.1371/journal.pone.0055158>
- Torres-Irineo, E., Gaertner, D., Chassot, E., Dreyfus-Leon, M., 2013. Changes in fishing power and fishing strategies driven by new technologies: The case of tropical tuna purse seiners in the eastern Atlantic Ocean - ScienceDirect.
- Wood, S.N., 2006. Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models. *Biometrics* 62, 1025–1036. <https://doi.org/10.1111/j.1541-0420.2006.00574.x>
- Ruiz, J., Depetris, M., Grande, M., Tserpes, G., Carbonara, P., Bach, P., Krug, I., Spedicato, M.T., Capello, M., Gaertner, D., Mugerza, E., Thasitis, I., Garibaldi, F., Mariani, A., Mariani, J., Murua, H., Pascual, P., Baez, J.C., Abascal, F., Uranga, J., Baidai, Y., 2019. Final report of the MARE/2016/22 projet, Strengthening regional cooperation in the area of large pelagic fishery data collection (acronym: RECOLAPE).

# Annexe

## Annexe 1 : La pêche à la senne

La pêche à la senne consiste à entourer le banc de poissons d'une senne à l'aide d'un petit navire nommé « skiff ». Une fois les deux bouts rattachés, il faut fermer la senne par le fond puis resserrer la senne près du navire. On peut ensuite remplir les cuves à l'aide d'une salabarde. On appelle l'ensemble de cette action « coup de pêche » ou « calée ».



**Illustration 1** : Pêche à la senne

## Annexe 2 : Prédiction et IC 95% des biais de composition spécifique pour bancs libres

**Tableau 1**: Prédiction et IC 95% des biais de composition spécifique pour bancs libres. Résultats du modèle linéaire mixte.

species	prediction	int_low	int_up
YFT	-0.00926442697430443	-0.0188915525551591	0.000362698606550279
BET	-0.0160200180805586	-0.0256471436614132	-0.00639289249970388
SKJ	0.0252844450548628	0.0156573194740081	0.0349115706357175

## Annexe 3 : Prédiction et IC 95% des biais de composition spécifique pour bancs objets

**Tableau 1:** Prédiction et IC 95% des biais de composition spécifique pour bancs objets. Résultats du modèle linéaire mixte.

species	ocean	prediction	int_low	int_up
YFT	Atlantique	-0.122587832302926	-0.139927640235158	-0.105248024370694
BET	Atlantique	-0.0521408285388417	-0.0694806364710737	-0.0348010206066097
SKJ	Atlantique	0.174728660841766	0.157388852909534	0.192068468773998
YFT	Indien	-0.199637604556294	-0.214195887786493	-0.185079321326095
BET	Indien	-0.0321393805475744	-0.0466976637777734	-0.0175810973173753
SKJ	Indien	0.23177698510387	0.217218701873671	0.246335268334069

## Annexe 4 : Prédiction et IC 95% des biais de composition par catégories de poids pour bancs libres

**Tableau 1:** Prédiction et IC 95% des biais de composition par catégories de poids pour bancs libres. Résultats du modèle linéaire mixte.

species	ocean	prediction	int_low	int_up
SKJ01.8	Atlantique	-0.0259229483198012	-0.0495977670516207	-0.00224812958798163
SKJ1.870	Atlantique	0.0502130284382217	0.0265382097064022	0.0738878471700412
BET010	Atlantique	2.53037076649878e-05	-0.0236495150241529	0.0237001224394829
BET1030	Atlantique	0.00100906275911599	-0.0226657559727036	0.0246838814909355
BET30190	Atlantique	-0.0170686611893215	-0.0407434799211411	0.00660615754249799
YFT010	Atlantique	-0.00558979893603868	-0.0292646176678582	0.0180850197957809
YFT1030	Atlantique	0.0341934231300577	0.0105186043982382	0.0578682418618773
YFT30190	Atlantique	-0.0368594095898777	-0.0605342283216972	-0.0131845908580582
SKJ01.8	Indien	-0.0104826465257484	-0.0336653300557116	0.0127000370042149
SKJ1.870	Indien	0.0338204839200077	0.0106378003900445	0.0570031674499709
BET010	Indien	0.00137498749582944	-0.0218076960341336	0.0245576710257925
BET1030	Indien	0.001166604369189	-0.0220160791607742	0.0243492878991522
BET30190	Indien	-0.0163858668548005	-0.0395685503847637	0.00679681667516275
YFT010	Indien	-0.00426538253519407	-0.0274480660651573	0.0189173009947691
YFT1030	Indien	0.333165307311882	0.309982623781919	0.356347990841845
YFT30190	Indien	-0.338393487181145	-0.361576170711108	-0.315210803651182

## Annexe 5 : Prédiction et IC 95% des biais de composition par catégories de poids pour bancs objets

**Tableau 1:** Prédiction et IC 95% des biais de composition par catégories de poids pour bancs objets. Résultats du modèle linéaire mixte.

species	ocean	prediction	int_low
SKJ01.8	Atlantique	-0.174762778567122	-0.192143566891535
SKJ1.870	Atlantique	0.351325699191623	0.33394491086721
BET010	Atlantique	-0.0274698268292624	-0.044850615153677
BET1030	Atlantique	-0.0201394831289193	-0.0375202714533324
BET30190	Atlantique	-0.00558338626108312	-0.0229641745854962
YFT010	Atlantique	-0.0631737518342927	-0.0805545401587058
YFT1030	Atlantique	-0.0267121484183792	-0.0440929367427923
YFT30190	Atlantique	-0.0334843241525614	-0.0508651124769745
SKJ01.8	Indien	-0.0846826073660641	-0.100283173402693
SKJ1.870	Indien	0.316728265306221	0.301127699269593
BET010	Indien	-0.0279489832574432	-0.0435495492940722
BET1030	Indien	-0.00480877535749765	-0.0204093413941262
BET30190	Indien	0.0014182599750288	-0.0141823060615997
YFT010	Indien	-0.0840283692728453	-0.0996289353094738
YFT1030	Indien	-0.0282917663794095	-0.0438923324160381

## Annexe 6 : Groupes par type de banc créés par les kmeans sur la composition spécifique

Pour chaque groupe issu de la classification par k-means sur la composition spécifique, nous avons calculé la composition moyenne des calées pour chaque type de banc ainsi que le nombre de fois ou l'espèce était présente.

**Tableau 1:** Groupes par type de banc créés par les kmeans sur la composition spécifique.  
Première partie : Composition moyenne  
Deuxième parti : nombre d'occurrences de l'espèce

Groupes	Type de banc	SKJ	YFT	BET	
1	BL	0,822	0,058	0,121	
1	BO	0,843	0,114	0,043	
2	BL	0,010	0,965	0,025	
2	BO	0,144	0,812	0,044	
Groupes	Type de banc	SKJ	YFT	BET	Maximum
1	BL	8377,000	1913,000	1932,000	8959,000
1	BO	31926,000	15722,000	6916,000	32159,000
2	BL	1188,000	27040,000	3061,000	27040,000
2	BO	3938,000	7319,000	1418,000	7319,000

## Annexe 7 : Groupes par type de banc créés par les kmeans sur la composition par groupe de poids

Pour chaque groupe issu de la classification par k-means sur la composition par classes de poids, nous avons calculé la composition moyenne des calées pour chaque type de banc ainsi que le nombre de fois ou l'espèce était présente.

**Tableau 1:** Groupes par type de banc créés par les kmeans sur la composition par groupe de poids.

Première partie : Composition moyenne

Deuxième parti : nombre d'occurrences de l'espèce

Groupes	type_banc	SKJ01.8	SKJ1.870	YFT010	YFT1030	YFT30190	BET010	BET1030	BET30190	Présence maximum
1	BL	183,000	8160,000	894,000	917,000	269,000	643,000	543,000	49,000	8160
1	BO	382,000	32282,000	11284,000	7771,000	1821,000	5218,000	972,000	625,000	32282
2	BL	255,000	1185,000	981,000	8332,000	18261,000	544,000	767,000	2593,000	27839
2	BO	870,000	2764,000	3377,000	3585,000	1206,000	1164,000	275,000	197,000	7196

Groupes	type_banc	SKJ01.8	SKJ1.870	YFT010	YFT1030	YFT30190	BET010	BET1030	BET30190
1	BL	0,006	0,876	0,032	0,031	0,007	0,025	0,021	0,001
1	BO	0,003	0,829	0,076	0,050	0,009	0,026	0,005	0,003
2	BL	0,007	0,009	0,020	0,279	0,637	0,012	0,014	0,024
2	BO	0,098	0,086	0,268	0,353	0,110	0,064	0,014	0,008

## Résumé

La pêche à la senne est l'une des pêches thonnières les plus importantes au monde, c'est pourquoi il est important d'avoir une gestion durable des stocks de thonidés. Au milieu des années 80, des études ont montré des erreurs dans la composition spécifique déclarée par les membres d'équipage. En particulier, le thon Albacore (*Thunnus albacares*) et le Patudo (*Thunnus obesus*) de petites tailles étaient souvent déclarés comme des thons Listao (*Katsuwonus pelamis* ; Bard, 1986; Cayré, 1984). Depuis les procédures échantillonnages ont été modifiées afin de tenir compte de l'évolution de la pêcherie due aux nouvelles technologies (Fonteneau et al., 2000; Pallarés and Hallier, 1997b; Parajua and Hallier, 1992). Cette étude a pour but de quantifier les biais de composition spécifique entre les déclarations des livres de bord et les échantillons faits aux ports entre les années 2000 et 2017 puis de proposer un début d'explication sur les mécanismes induisant ces biais. Il traitera également des effets de la procédure d'échantillonnage ainsi que du prétraitement nécessaire à cette analyse sur la finesse des résultats que l'on peut obtenir.

**Mots-clés :** pêche, thon, échantillon, composition spécifique, patudo, albacore, listao, *Thunnus albacares*, *Thunnus obesus*, *Katsuwonus pelamis*

## Abstract

Purse seine fishing is one of the most important tuna fisheries, therefore it is important to have sustainable tunas' stock assessment. In the mid-1980s, data collection was conducted to investigate errors in the species composition reported by crews. Studies have shown that small Bigeye tuna and Yellowfin tuna are often reported as Skipjack, and that smaller Bigeye tuna and Yellowfin tuna are difficult to differentiate (Bard, 1986; Cayré, 1984). The sampling procedures have been thereafter modified to take into account the evolution of purse seine fisherie due to technological development (Fonteneau et al., 2000; Pallarés and Hallier, 1997b; Parajua and Hallier, 1992). The purpose of this report is to quantify compositional biases between logbook and samples performed at ports between the years 2000 and 2017, and to propose an explanation of the process leading to these biases. It will also speak about the effects of the sampling procedure and the pre-treatment required for this analysis of the fineness of the results that can be obtained.

**Keywords :** fishing, tuna, logbook, sample, species composition, bias, yellowfin, begeye, skipjack