



Estimer le nombre de captures accidentelles de dauphins communs dans les engins GNS/GTR

**Analyse préliminaire des données
OBSCAME par Régression Multi-niveau avec
Post-stratification**

2024-11-22

AUTHIER Matthieu, BREVET Mathieu, CLOÂTRE Thomas, DUBROCA Laurent

La Rochelle Université / Ifremer

2024/11/30

Rapport technique (version 2)

Citation proposée : Authier, M., Brevet, M., Cloâtre, T., & Dubroca, L. (2024) Estimer le nombre de captures accidentelles de petits cétacés dans les engins GNS/GTR - Analyse des données OBSCAMe par Régression Multiniveau avec Post-stratification. Rapport technique (version 2). Observatoire Pelagis, UAR 3462, CNRS-LRUniv / Ifremer.

Travaux effectués dans le cadre de la saisine 2024 sur les captures accidentelles de petits cétacés dans le Golfe de Gascogne. Chapitre 2a – estimation du risque et de la mortalité par captures accidentelles des petits cétacés

Acronymes

- CAR : < Conditional Auto-Regressive >
- GNS : filets maillants calés de fond à une nappe, filets droits
- GTR : filets maillants
- MrP : < Multi-level regression with Postratification > ou Régression Multi-niveau avec Poststratification
- OBSMER : programme d'observation de la mer de la Direction générale des affaires maritimes, de la pêche et de l'aquaculture (DGAMPA)
- OFB : Office Français de la Biodiversité
- RNE : Réseau National Échouages
- SACROIS : algorithme de rapprochement, vérification, contrôles de cohérence de différents flux de données afin de produire des séries de données de production et d'effort validées, consolidées et qualifiées
- SIH : Système d'Informations Halieutiques

I. Introduction

La réduction des captures accidentelles de dauphins communs (*Delphinus delphis*) est un enjeu majeur. La saisine du 19 juillet 2024 portant sur « des besoins d'analyse de risque et estimation des captures accidentelles de petits cétacés dans le golfe de Gascogne à partir des données de captures accidentelles et de l'effort de pêche dans le golfe de Gascogne, et proposition d'un plan de suréchantillonnage pour l'hiver 2024-2025 du programme ObsMer » demande d'estimer les mortalités en mer sur la base d'une collecte de données plus importante à bord des navires (observateurs ou caméras). Porté par l'Office Français de la Biodiversité (OFB), le projet OBSCAME a permis de mieux comprendre, au travers de l'observation électronique (REM, ou « Remote Electronique Monitoring »), les captures accidentelles de mammifères marins par les fileyeurs opérant dans le golfe de Gascogne (Vignard & Tachoures 2023). En 2022 et 2023, 20 fileyeurs volontaires se sont équipés de caméras afin de filmer la remontée du filet à bord du bateau. Les vidéos ont ensuite été visionnées par une équipe dédiée, pour renseigner une base de données dès lorsqu'une capture accidentelle de mammifère marin est identifiée (Vignard & Tachoures 2023). Cette base de données a été partagée avec le Système d'Informations Halieutiques (SIH) de l'Ifremer.

Dans le cadre des travaux 2a de la saisine sus-nommée, une analyse préliminaire des données OBSCAME a été réalisée, notamment pour tester l'approche de Régression Multiniveau avec Post-stratification (« Multi-level regression with Postratification », MrP; Gelman & Little 1997) afin d'estimer le nombre total de dauphins capturés dans les engins GNS (filets maillants calés de fond à une nappe, filets droits) et GTR (trémails). La méthodologie MrP a été développée et appliquée aux données OBSMER et SACROIS concernant les chaluts boeufs pélagiques (PTM; Rouby et al. 2022). Les premiers essais pour appliquer cette méthodologie aux données OBSMER collectées sur les fileyeurs n'avaient pas abouti, du fait de non-convergence de paramètres clés du modèle sous-jacent. Cette non-convergence traduisait probablement un défaut du modèle à accommoder les données des fileyeurs, caractérisées par une plus grande hétérogénéité.

Les données OBSCAME offrent une occasion unique pour poursuivre les développements et l'adaptation de l'approche MrP aux données de captures accidentelles de dauphins communs. Le présent rapport détaille ces développements et les verrous rencontrés, ainsi que les premiers résultats d'estimation issue de l'approche MrP. **Ces résultats sont préliminaires et surtout illustratifs d'un cadre analytique qui pourrait être déployé dans le futur, une fois certains verrous levés.**

II. Données

Les données sont 13943 opérations de pêche dans les sous divisions 27.7.e, 27.8.a et 27.8.b du CIEM (Fig. 1). Les analyses ont été réalisées avec le logiciel de programmation et d'analyse de données R v4.2.2 (R Core Team 2022). L'ensemble des préparations et analyses de données a été réalisé sous ce même logiciel avec les outils du *tidyverse* (Wickham et al. 2019), et pour la cartographie de la bibliothèque *sf* (Pebesma & Bivand 2023).

Estimer le nombre de captures accidentelles de dauphins communs dans les engins GNS/GTR

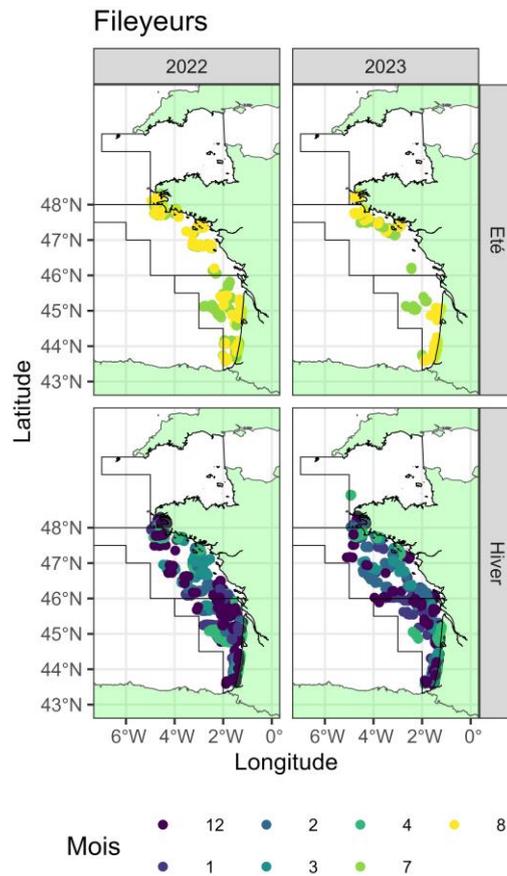


Fig. 1 — Données OBSCAMe collectées dans le golfe de Gascogne, sous-division CIEM 27.8.a et 27.8.b. Les lignes noires symbolisent les limites des sous-divisions de la division 27.8 du CIEM. Le code couleur indique le mois.

La sous division 27.7.e est moins échantillonnée que les sous divisions 27.8.a et 27.7.b.

Tableau 1 — Statistiques descriptives du jeu de données OBSCAMe

Sous division CIEM	Année	Evénements	N
27.7.e	2022	1	1012
27.7.e	2023	1	930
27.8.a	2022	27	3811
27.8.a	2023	25	2133
27.8.b	2022	29	3598
27.8.b	2023	33	2468

Les données sont collectées dans 25 rectangles statistiques (Fig. 2).

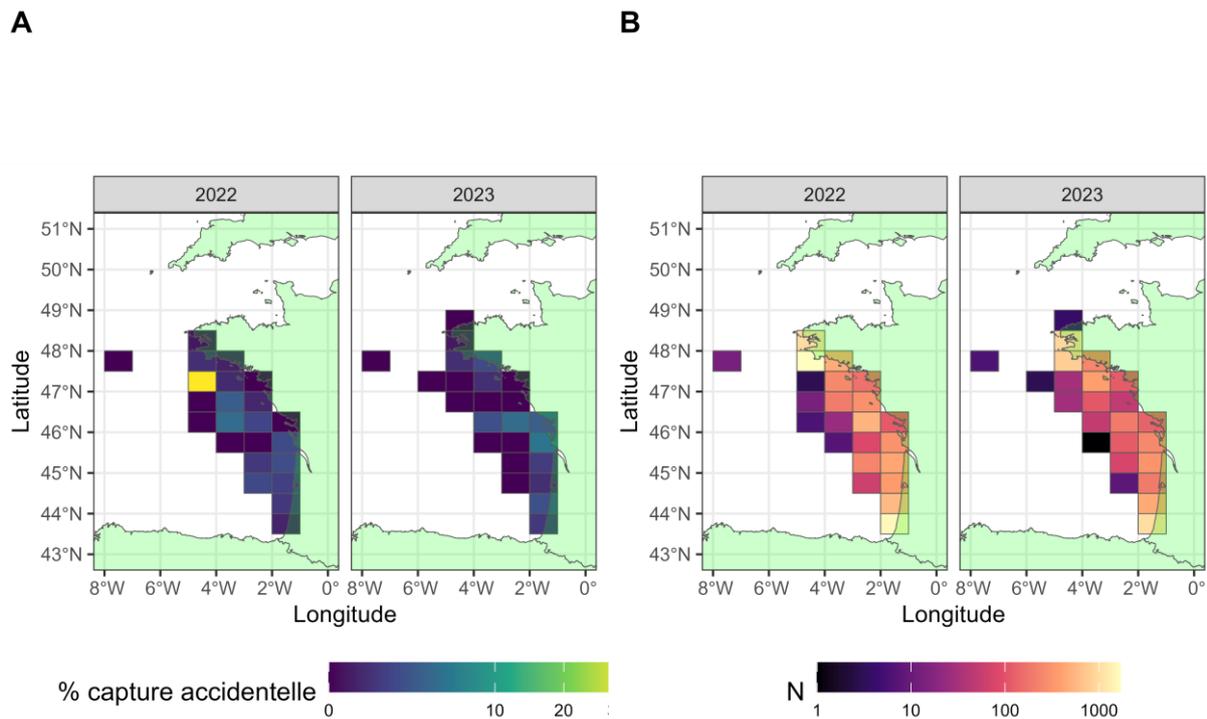


Fig. 2 — Données OBSCAME collectées dans le golfe de Gascogne. (A) Fréquence des évènements de capture accidentelle de dauphin commun (échelle racine carrée). (B) Taille d'échantillon (échelle logarithmique).

Des disparités de 2 ordres de grandeur sont visibles entre les rectangles statistiques (Fig. 2). La fréquence des évènements de captures accidentelles est plus élevée dans les rectangles statistiques les moins échantillonnés, ce qui amène le statisticien Andrew Gelman à écrire que ce type de représentation (Fig. 2 A) est nécessairement problématique (Gelman & Price 1999). Ce patron de corrélation entre observation extrême et faible taille d'échantillon est souvent observé et reflète en grande partie les aléas d'échantillonnage. Une solution est ici de recourir à la modélisation afin de lisser ces artefacts liés à l'échantillonnage.

II.1. Rareté des évènements dans le jeu de données

Un total de 116 évènements de captures accidentelles sur 13943 opérations de pêche a été observés dans les 25 rectangles statistiques retenus (répartis sur 3 sous-divisions CIEM, Fig. 3). A l'échelle d'une opération de pêche, la fréquence d'un évènement de capture accidentelle peut être estimée simplement par $\frac{116}{13943} \approx 0.0083$, soit moins de 1%. Un tel chiffre correspond à un niveau de rareté dans les données qualifié d'extrêmement rare par Shyalika et al. (2023). Shyalika et al. (2023) détaillent les défis associés à l'analyse de données avec si peu d'évènements du point de vue de l'apprentissage machine (« machine learning »).

Estimer le nombre de captures accidentelles de dauphins communs dans les engins GNS/GTR

Van der Paal (2014) offre des recommandations et une discussion de point de vue statistique pour accommoder un tel niveau de rareté des événements dans les données lorsqu' une régression logistique est utilisée pour l'analyse.

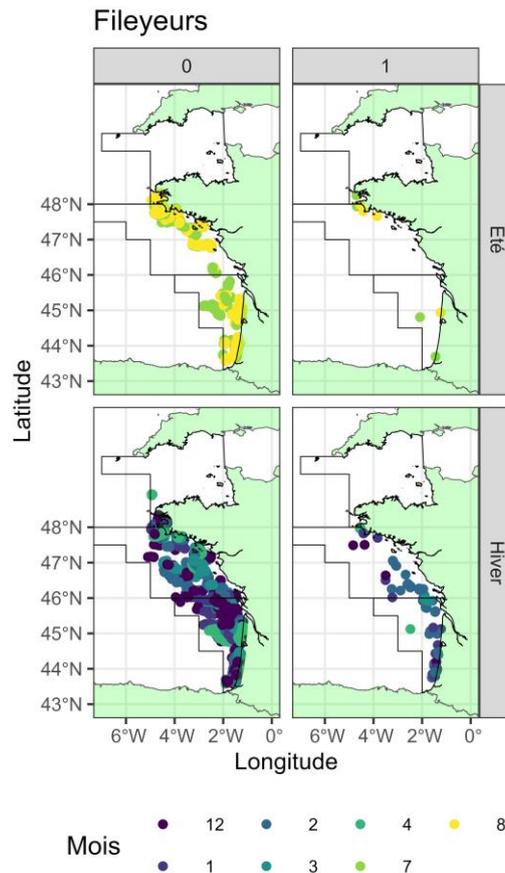


Fig. 3 — Données OBSCAME collectées dans le golfe de Gascogne, sous-division CIEM 27.8.a et 27.8.b. Les lignes noires symbolisent les limites des sous-divisions de la division 27.8 du CIEM. Les données relatives aux évènements de captures accidentelles (1) ou non (0) sont représentées par saison (2022 et 2023 agrégées). Le code couleur indique le mois.

III. Méthodologie

La structure globale du modèle utilisé pour l'analyse est décrite par Rouby et al. (2022). Le modèle analyse conjointement :

1. l'occurrence d'un évènement de capture à chaque opération de pêche (variable réponse: < by-catch >, vraisemblance Bernoulli notée B);
2. la durée d'une opération de pêche (variable réponse: < duration > en heures, vraisemblance Gamma notée Γ); et
3. le nombre d'opération de pêche par jour de mer (variable réponse: < Fishing operation >, vraisemblance Poisson tronquée à 0 notée P^+).

Prédire l'occurrence d'un événement de capture à l'échelle de pêche repose sur l'utilisation d'un processus Bernoulli de paramètre Π où Π est une probabilité qui peut être interprétée comme le risque de capture accidentelle, *i.e.* le produit de la probabilité de capture dans un engin de pêche et la probabilité de présence des animaux dans la zone où se déroule l'opération de pêche. Les données disponibles permettent d'estimer ce risque Π à l'échelle de l'opération de pêche (*e.g.* Rouby et al. 2022).

III.1. Choix de la vraisemblance pour la durée

La vraisemblance Γ choisie pour modéliser les durées par Rouby et al. (2022) est restrictive car elle suppose un coefficient de variation constant. D'autres choix ont été testés : une vraisemblance Weibull (notée W) et une vraisemblance \langle log-skew-normal \rangle (notée logSN; Ben Hacine & Bouallegue 2015).

III.2. Fonction de lien asymétrique

Van der Paal (2014) discute et compare différentes fonctions de lien (« link function ») pour régression logistique. Des fonctions asymétriques en particulier permettent de donner plus de poids aux événements rares lors de l'estimation des paramètres et offrent une solution pour accommoder le niveau de rareté sans modifier le jeu de données (*e.g.* sur-échantillonner les événements ou sous-échantillonner les non-événements). La fonction de lien correspondant à la fonction de densité de probabilité cumulée d'une distribution logistique standard asymétrique est illustrée pour différent choix du paramètre d'asymétrie (paramètre contrôlé par le choix d'un quantile ; Fig. 4).

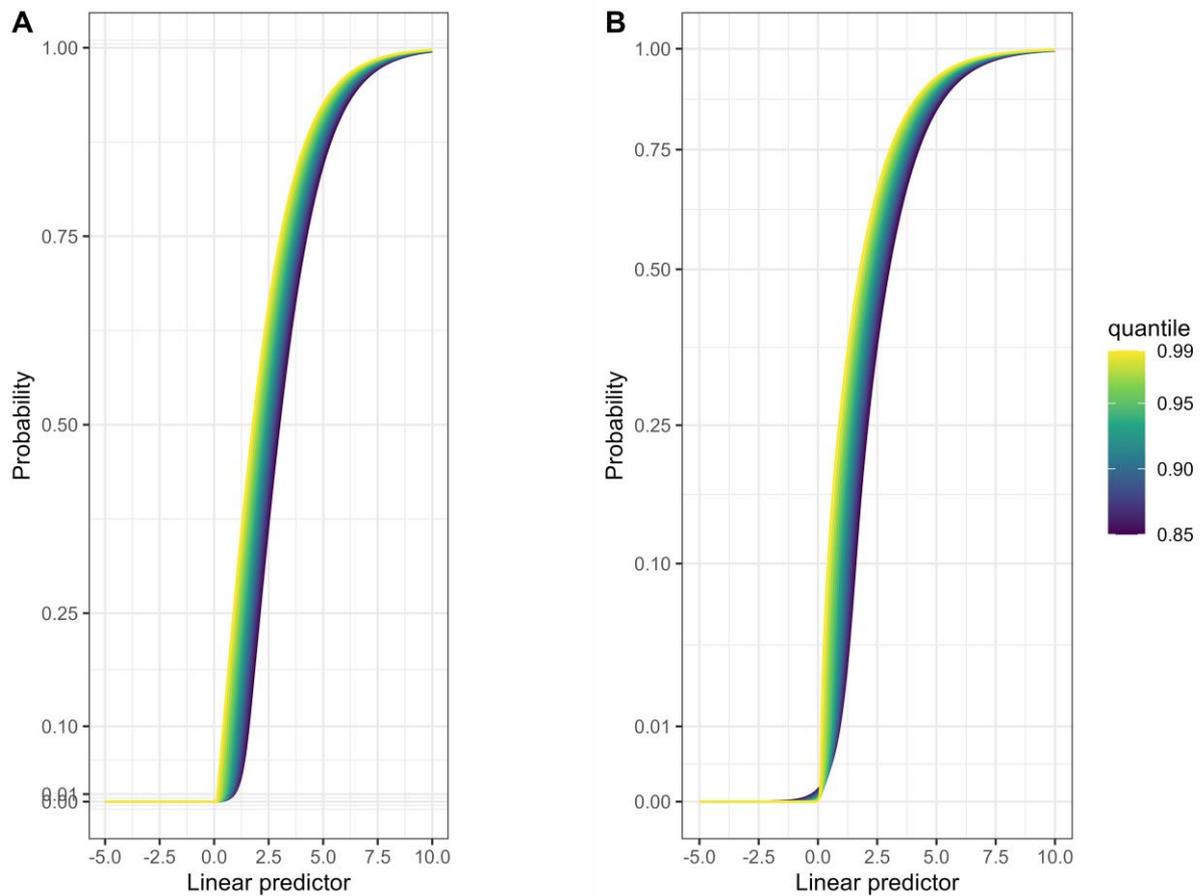


Fig. 4 — Fonction de lien asymétrique pour régression logistique. (A) échelle naturelle sur l'axe des ordonnées; (B) échelle racine carrée sur l'axe des ordonnées permettant de mieux visualiser les faibles probabilités.

La vitesse avec laquelle la probabilité approche la valeur 0 est différente de celle avec laquelle la valeur 1 est approchée : cette forme asymétrique permet de donner moins de poids aux non-événements. Par exemple, une valeur du prédicteur linéaire de -5 ou de -1 ne fait que peu de différence en termes de probabilité. En revanche, une valeur du prédicteur linéaire de 5 correspond à une probabilité très différente de celle associée à une valeur du prédicteur linéaire de 1.

L'utilisation d'une fonction de lien asymétrique (pour un quantile de 0.99 correspondant à la fréquence empirique < 0.1%) a été testée. La log-vraisemblance associée à ce choix est représentée sur le graphique Fig. 5.

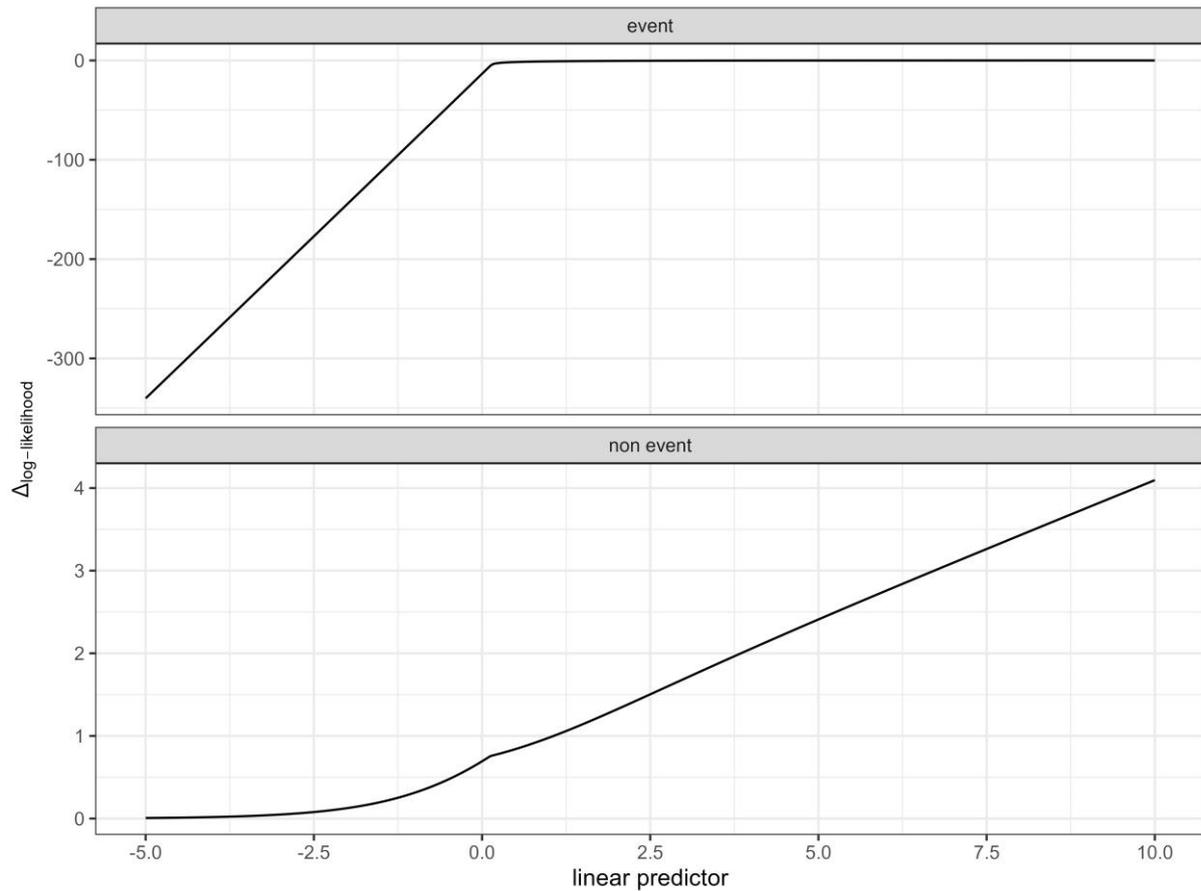


Fig. 5 — Différence entre les log-vraisemblance Bernoulli associées à un évènement (1) et un non-évènement (0) avec un lien logit asymétrique et un lien logit symétrique en fonction du prédicteur linéaire. La fonction de lien asymétrique pénalise beaucoup plus fortement une mauvaise classification d'évènements (valeurs de log-vraisemblance très négatives pour des valeurs du prédicteur linéaire également négatives) qu'une mauvaise classification d'un non-évènement. Elle permet ainsi de donner plus de poids aux évènements rares lors de l'estimation des paramètres du modèle.

III.3. Structure spatiale

Un autre développement en termes de modélisation est de faire une analyse à l'échelle du rectangle statistique et de complexifier le modèle de Rouby et al. (2022) pour tenir compte de l'agencement spatiale de ceux-ci (Fig. 6) de sorte à prendre en considération une autocorrélation spatiale. Le modèle utilisé ici est modèle CAR (« Conditional Auto-Regressive »; Morris et al. 2019) qui permet compte de tenir compte du voisinage spatial d'un rectangle.

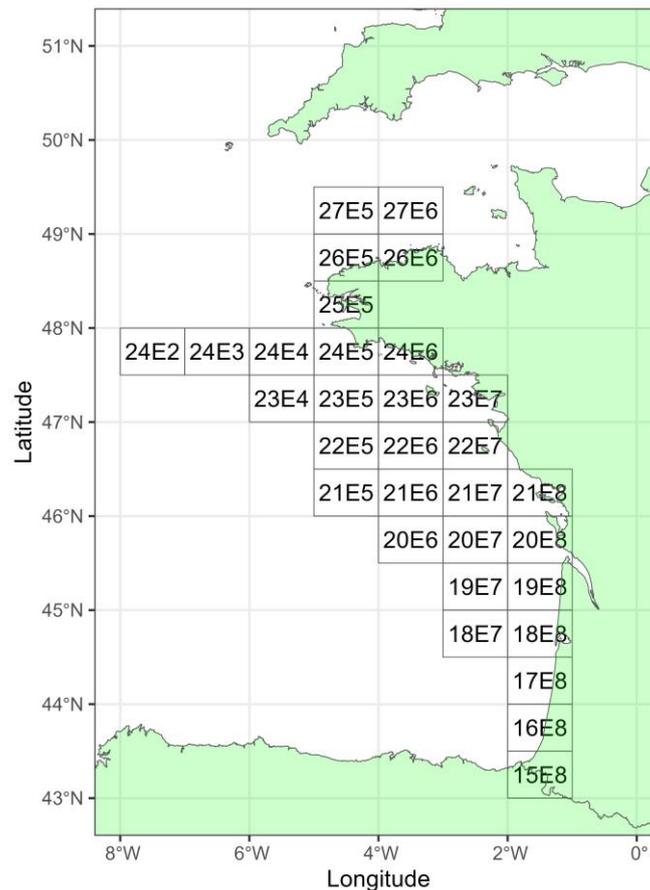


Fig. 6 — Agencement spatial des rectangles statistiques CIEM dans le golfe de Gascogne.

III.4. Ajustement des modèles

Un total de 12 modèles ont été encodés dans le langage *Stan* (Carpenter et al. 2017) et ajusté sur le cluster de calcul *Curta* du Mésocentre de Calcul Intensif Aquitain et hébergé à l'Université de Bordeaux, Nouvelle Aquitaine. Ces modèles permettent d'évaluer les différents choix de modélisation détaillé plus haut afin de réaliser une exploration des différents choix de modélisations possible, et notamment le choix de la vraisemblance pour les données de durées (Gamma $M_{\{1-3\}}$ vs Weibull $M_{\{3-6\}}$ vs log-skew-normal $M_{\{6-12\}}$), le choix d'une résolution spatiale (sous-division $M_{\{1,4,7,10\}}$ ou rectangle statistique $M_{\{2-3, 5-6, 8-9, 11-12\}}$), et le choix de la fonction de lien (logit symétrique $M_{\{1-9\}}$ ou asymétrique $M_{\{10-12\}}$).

Estimer le nombre de captures accidentelles de dauphins communs dans les engins GNS/GTR

Huit chaînes de Markov sont initialisées sur des valeurs tirées aléatoirement pour les paramètres puis un total de 3000 valeurs sont échantillonnées avec un algorithme de Monte Carlo Hamiltonien. Les 2000 premières valeurs sont utilisées pour une phase adaptative d'échauffement de l'algorithme et les 1000 dernières valeurs sont conservées. La convergence des paramètres d'un modèle est évaluée à l'aide du diagnostic de Gelman-Rubin ($\hat{R} < 1.025$).

Tableau 2 — Spécification et sélection du modèle.

Modèle	Echelle spatiale	Variable réponse et vraisemblance Statistique				Leave-One-Out Information Criterion (looiC)
		Evènement de capture	Duration	Jour de Mer	Intensité des évènements	
M_1	subdivision	B (lien logit)	Γ	P^+	NB^+	116407 ₂₈₅
M_2	rectangle	B (lien logit)	Γ	P^+	NB^+	116177 ₃₁₁
M_3	CAR	B (lien logit)	Γ	P^+	NB^+	116200 ₃₁₁
M_4	subdivision	B (lien logit)	W	P^+	NB^+	41056 ₂₄₄
M_5	rectangle	B (lien logit)	W	P^+	NB^+	40128 ₂₈₀
M_6	CAR	B (lien logit)	W	P^+	NB^+	40139 ₂₇₉
M_7	subdivision	B (lien logit)	logSN	P^+	NB^+	37570 ₂₂₇
M_8	rectangle	B (lien logit)	logSN	P^+	NB^+	37406 ₂₆₀
M_9	CAR	B (lien logit)	logSN	P^+	NB^+	37422 ₂₅₉
M_{10}	subdivision	B (lien logit asymétrique)	logSN	P^+	NB^+	37566 ₂₂₇
M_{11}	rectangle	B (lien logit asymétrique)	logSN	P^+	NB^+	37381 ₂₅₈
M_{12}	CAR	B (lien logit asymétrique)	logSN	P^+	NB^+	37403 ₂₅₉

Sélection des modèles

L'ajustement du modèle aux données est évalué avec le critère *looiC* (Gelman et al. 2014) : plus ce critère est faible, meilleure est l'ajustement du modèle. La valeur (X), ainsi que son erreur d'estimation, de ce critère prédictif est reportée dans le Tableau 2 sous la forme X_e . Tenir compte du rectangle statistiquement et de leur agencement spatial améliore de manière marginale l'ajustement (comparer le *looiC* des modèles M_{2-3} , M_{5-6} , M_{8-9} & M_{11-12}). Le meilleur modèle est le modèle M_{11} mais un modèle plus complexe (M_{12}) est compétitif au vu de l'erreur associée au critère *looiC*. Ainsi la différence de *looiC* entre les modèles M_{11} et M_{12} est de l'ordre de 20 points à mettre en regard d'une erreur d'estimation du critère de l'ordre de 260 points : cette différence n'est donc pas significative et l'ajustement des deux modèles aux données est donc similaires. Ces modèles M_{11} et M_{12} sont très similaires : M_{12} permet toutefois une

cartographie à l'échelle du rectangle statistique alors que M_{11} reste à l'échelle de la sous-division CIEM. Les modèles M_{8-9} sont également compétitifs mais l'estimation du risque moyen de capture est problématique semble trop élevée (≈ 0.0191 ou 2% ; soit plus du double de la valeur de la moyenne empirique) : cette possible sur-estimation semble ensuite compensée par le paramètre de variabilité temporelle du risque. L'estimation du risque moyen avec le lien logit asymétrique est plus proche de la moyenne empirique (≈ 0.0128 ou 1.3%).

Validation des modèles

La validation des paramètres estimés d'un modèle a été graphique. Par exemple, le graphique Fig. 7 montre un ajustement acceptable du modèle M_{11} aux données de captures accidentelles. Un autre paramètre du modèle, lié aux variations spatiales (entre sous-division), permet néanmoins d'estimer des risques non-nuls (mais dans des gammes raisonnables, *i.e.* < 0.1) au cours du temps dans les différentes sous-division.

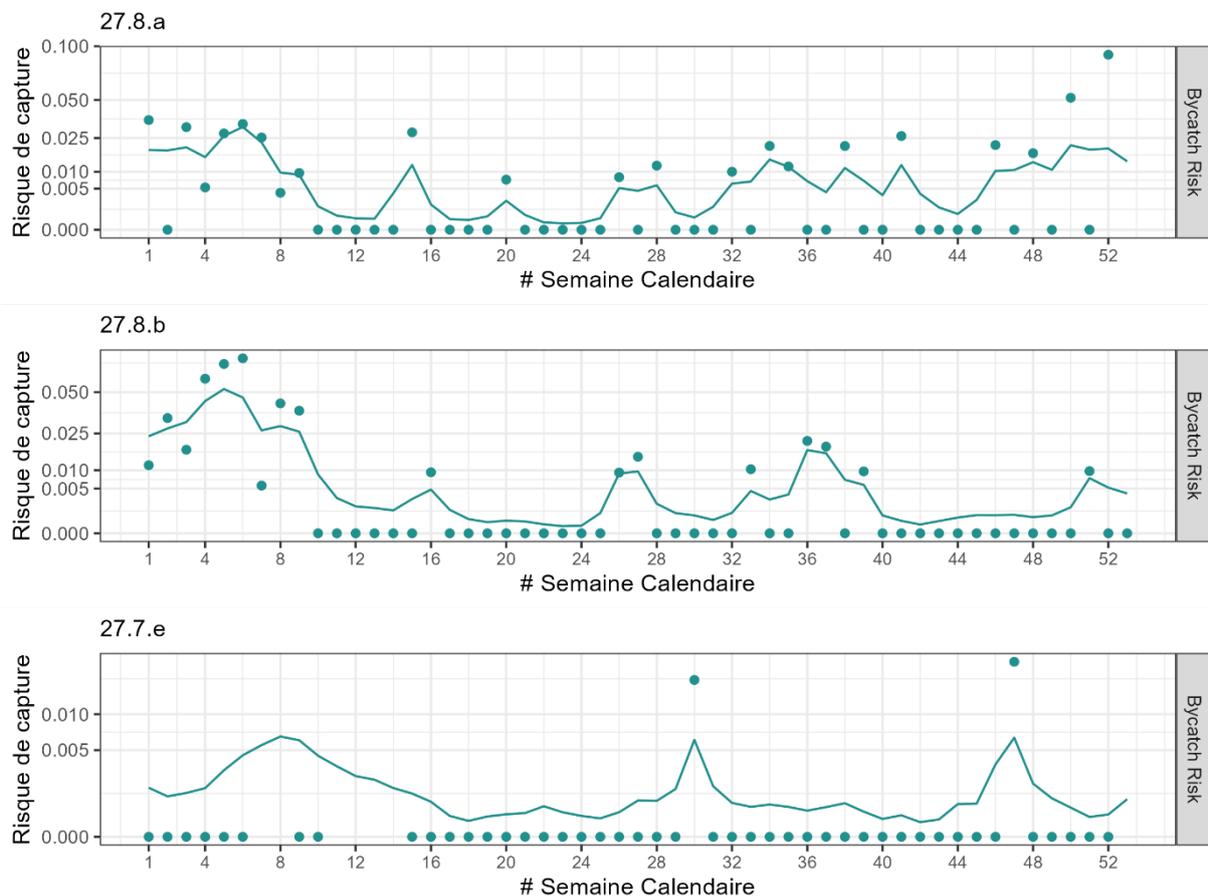


Fig. 7 — Risque de capture accidentelle pendant une opération de pêche par sous-division CIEM (estimations issues du modèle M_{11}). Les points représentent les données et la ligne l'estimation moyenne issue du modèle. Les incertitudes ne sont pas représentées par souci de lisibilité.

L'ajustement du modèle aux données de captures accidentelles est correct pour chaque sous-division (Fig. 8).

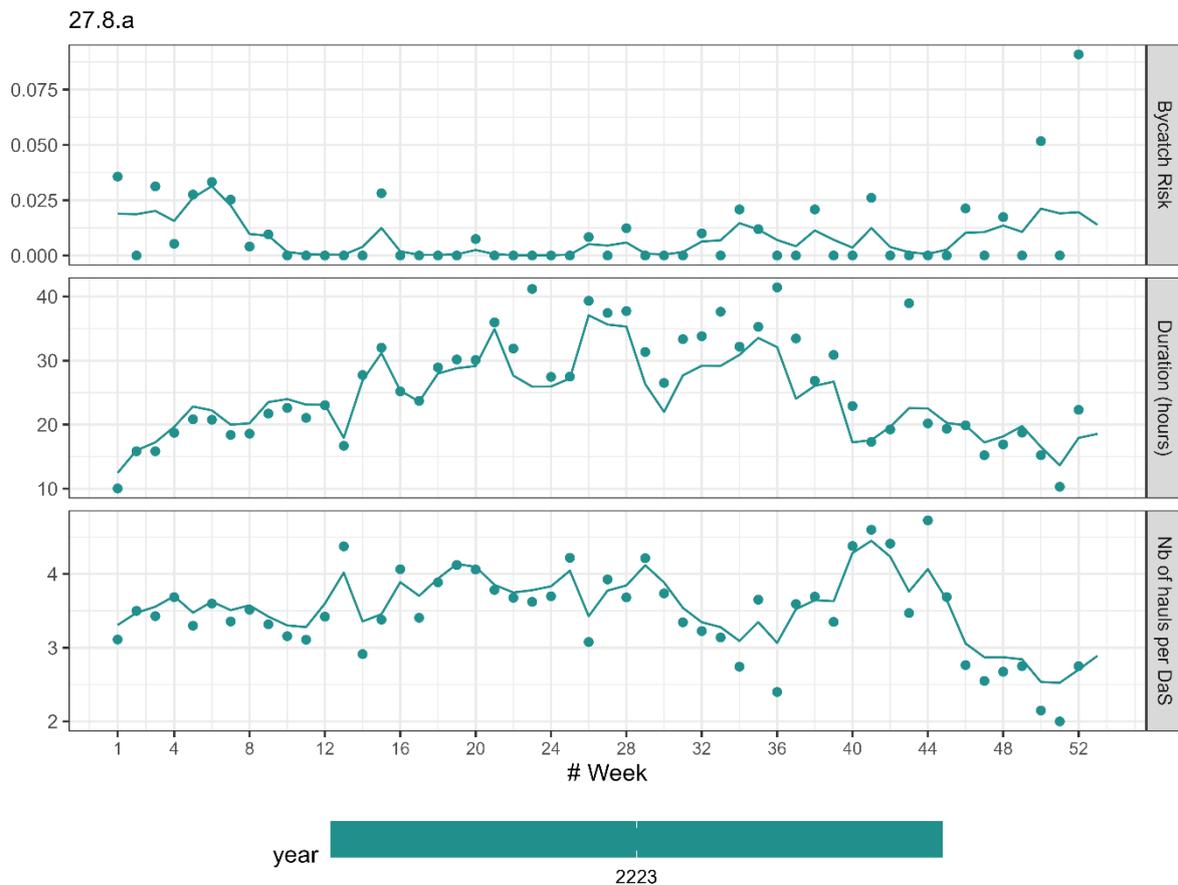


Fig. 8 — Comparaison données-estimations moyennes pour la sous-division CIEM 27.8.a (estimations issues du modèle M_{11}).

Les estimations des durées moyennes issues du modèle M_{11} ne sont pas corrélées aux variations du risque de capture. Rouby et al. (2022) incorporaient ces données de durées car, dans le cas des chaluts pélagiques en bœuf, cette durée est informative sur le risque de capture accidentelle (données OBSMER). Avec les données OBSCAME, la corrélation estimée entre le risque de capture et la durée d'immersion d'un filet est faible. Une simplification du modèle de Rouby et al. (2022) pour les données OBSCAME serait donc justifiée pour écarter les données de durée de trempé.

Enfin, l'ajustement du modèle aux données concernant le nombre d'opération de pêche par jour de mer est satisfaisant (Fig. 8). Les estimations du risque (probabilité de capture lors d'une opération de pêche) et du nombre d'opérations par jour de mer permettent de réaliser une estimation du nombre d'évènements de captures accidentelles à partir du nombre de jour de mer.

III.5. Post-stratification

Avec le modèle M_{11} , le nombre de combinaisons pour la post-stratification est $3 \times 53 = 159$: trois sous-divisions CIEM et jusqu'à 53 semaines dans une année. Le nombre de jours de mer

pour les engins GNS/GTR est fourni par SACROIS pour chacune de ces combinaisons. Il est dès lors possible d'obtenir une estimation du nombre de captures accidentelles et de dauphins capturés à partir des estimations des paramètres d'un modèle (e.g. Fig. 8, sous-panneaux haut et bas) et du nombre de jours de mer. Pour les jours de mers sans caméras (i.e. on retranche de l'estimation SACROIS le nombre de jours de mer correspondant au projet OBSCAMe), le nombre d'opérations de pêche pour une combinaison sous-division/ semaine est obtenu, puis multiplié par le risque de capture accidentelle pour une opération de pêche et par le nombre moyen de dauphins capturés lors d'un événement (estimé lui aussi à partir des données OBSCAMe uniquement). Cette étape de post-stratification est relativement simple mais peut-être intensive en calcul pour tenir compte de manière exacte des incertitudes d'estimations (e.g. pour obtenir la distribution *a posteriori* du nombre de captures sur un an). Deux approches sont ici possibles :

- un calcul exact de la distribution prédictive *a posteriori*; ou
- une approximation de cette même distribution.

La seconde approche est immédiate en calcul alors que la première va nécessiter plusieurs heures de calculs. Pour être plus précis, la première approche implique pour chaque valeur issue de l'échantillon de la distribution postérieure des paramètres du modèle sélectionné :

- prédire le nombre d'opérations de pêche à partir du nombre de jour de mers par tirage dans une loi Poisson tronquée à 0 notée P^+ ;
- prédire pour chaque opération de pêche si il y a eu capture accidentelle ou non par tirage dans une loi Bernoulli B ;
- prédire pour chaque événement le nombre de dauphins concernés par tirage dans une loi Binomiale Négative tronquée à 0 notée NB^+ ; et
- sommer le nombre de dauphins.

Cette approche implique donc d'obtenir la distribution prédictive pour chaque combinaison (sous-division CIEM, année, semaine). L'approche approximative quant-à-elle repose sur un facteur multiplicatif dérivé à partir de la distribution *a posteriori* des paramètres du modèle sélectionné. Ce facteur multiplicatif correspond au nombre moyen de dauphins capturés par jour de mer qu'il suffit dès lors de multiplier à la somme des jours de mer pour obtenir une estimation. Toutefois, cela représente une approximation qui ignore une partie de la variabilité puisque le processus de capture accidentelle à l'échelle de l'opération de pêche n'est pas explicité. En résumé il est attendu de la première approche, coûteuse en calcul une meilleure prise en compte de la variabilité, et donc des estimations (i.e. somme des dauphins capturés par événements) plus élevées avec un intervalle de confiance plus large.

Ces deux approches ont été mises en œuvre dans le but d'évaluer ici la qualité de l'approximation en utilisant les données SACROIS à partir de 2017. Le choix de cette année reflète le début des échouages massifs de dauphins communs pendant l'hiver et permet donc à l'hypothèse d'un profil hebdomadaire du risque comparable entre années sur la période 2017-2023.

Approche exacte

Les résultats de l'approche exacte mise en œuvre à partir des estimations du modèle M_{11} sont rapportées dans le Tableau 3.

Estimer le nombre de captures accidentelles de dauphins communs dans les engins GNS/GTR

Tableau 3 — Estimation du nombre d'événements et de dauphins capturés par l'approche exacte. La moyenne *a posteriori* est encadrée par un intervalle de confiance à 80% (Louis & Zeger 2009).

Année	Jour de Mer	Nb d'événements	Nb de dauphins
2017	56859	.297 1117 .2182	.359 1367 .2735
2018	57270	.297 1114 .2156	.374 1355 .2716
2019	57516	.328 1162 ..2235	.391 1426 .2802
2020	54222	.320 1115 .2131	.369 1358 .2661
2021	56160	.333 1137 .2197	.397 1391 .2759
2022	53244	.278 990 .1952	.315 1211 .2428
2023	52006	.295 1050 .2006	.337 1281 .2504

Sous l'hypothèse que les données OBSCAME soient représentatives de l'ensemble de la flotille et d'un risque constant depuis 2017, les estimations médianes du nombre de dauphins capturés accidentellement dans les filets sont de l'ordre de 1300 – 1500 animaux par an (pouvant monter à 2800 individus en tenant compte des incertitudes) pour l'ensemble des sous-divisions 27.7.e, 27.8.a et 27.8.b.

Approximation

Les résultats de l'approximation mise en œuvre à partir des estimations du modèle M_{11} sont rapportées dans le Tableau 4.

Tableau 4 — Estimation du nombre d'événements et de dauphins capturés par l'approche approximative. La moyenne *a posteriori* est encadrée par un intervalle de confiance à 80% (Louis & Zeger 2009).

Année	Jour de Mer	Nb d'événements	Nb de dauphins
2017	56859	.318 1173 .2406	.318 1272 .2708
2018	57270	.324 1156 .2366	.324 1252 .2667
2019	57516	.337 1213 .2440	.337 1313 .2752
2020	54222	.335 1126 .2268	.335 1219 .2555
2021	56160	.342 1177 .2389	.342 1274 .2692
2022	53244	.351 1094 .2181	.351 1185 .2456
2023	52006	.390 1155 .2258	.390 1264 .2554

Sous l'hypothèse que les données OBSCAME soient représentatives de l'ensemble de la flotille et d'un risque constant depuis 2017, les estimations du nombre de dauphins capturés accidentellement dans les filets sont de l'ordre de 1000 – 1200 animaux par an pour l'ensemble des sous-divisions 27.7.e, 27.8.a et 27.8.b.

Qualité de l'approximation

La qualité de l'approximation est acceptable : elle est très correcte pour le nombre d'événements de captures accidentelles (bien qu'elle semble induire une légère sur-estimation du nombre d'évènements) mais aboutit à une légère sous-estimation de la moyenne et de la largeur de l'intervalle de confiance pour le nombre de dauphins. Cela vient du fait que l'approche exacte prend bien en compte la possibilité de quelques événements d'un grand nombre d'animaux (ce nombre moyen est de 1.2 dans les données du projet OBSCAME avec un maximum à 6 individus). Toutefois, les ordres de grandeurs sont respectés. Il est recommandé d'utiliser l'approche exacte au vu du statut d'espèce protégée dont bénéficie le dauphin commun afin de se conformer au principe de précaution, et ainsi éviter de sous-estimer l'ampleur des captures accidentelles en nombre d'individus prélevés.

III.6. Cartographie

Le modèle M_{12} qui est ajusté à l'échelle du rectangle statistique CIEM (Fig. 6) permet de spatialiser à cette échelle les estimations obtenues par post-stratification. Ces dernières, pour les hivers 2022 et 2023 sont représentées ci-dessous (Fig. 9).

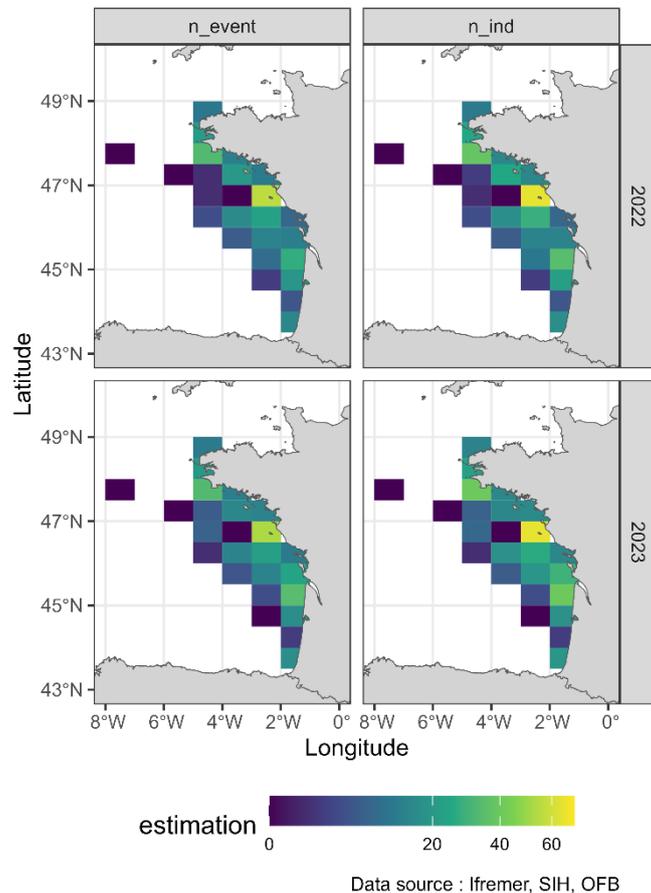


Fig. 9 — Estimations du nombre d'événements de captures accidentelles (n_{event}) et de dauphin communs (n_{ind}) capturés dans les filets pendant l'hiver 2022 et 2023 par rectangle statistique (estimations issues du modèle M_{12}).

Les estimations par rectangle statistique montrent une distribution côtière de la mortalité estimée, avec estimations élevées pour les rectangles jouxtant le département de la Vendée (85) et le sud du Finistère (29) pour les hivers 2022 et 2023.

Le modèle M_{12} autorise de faire des cartographies pour chaque semaine d'une année (sous l'hypothèse que les données du projet OBSCAME collectée sur deux années complètes peuvent être utilisées pour des estimations sur les années antérieures).

IV. Conclusion

Une analyse exploratoire des données du projet OBSCAMe (Vignard & Tachaires 2023) a été réalisée en vue de les utiliser pour estimer le nombre de captures accidentelles de dauphins communs dans les filets (engins GNS/GTR) avec la méthode dite de régression multi-niveau avec post-stratification. Cette exploration a permis de continuer le développement du modèle décrit par Rouby et al. (2022) afin de l'adapter aux données des fileyeurs :

- ce modèle peut être simplifié pour n'inclure que les données relatives aux événements de captures accidentelles et du nombre d'opérations de pêche par jour de mer;
- ce modèle peut être complexifié pour tenir compte de la distribution spatiale des données à l'échelle du rectangle statistique pour une plus grande précision spatiale;
- l'étape de post-stratification à l'échelle du rectangle statistique est coûteuse en calcul si la distribution prédictive *a posteriori* est recherchée mais il existe une approximation rapide et peu coûteuse en temps de calcul avec une précision moindre mais toujours acceptable; et
- il est important de tenir compte dans la modélisation du faible nombre d'évènements par rapport au nombre total d'opération de pêche. L'approche retenue avec une fonction de lien asymétrique est prometteuse.

L'analyse réalisée permet de souligner les modifications à faire et à poursuivre en vue de développer un modèle adapté aux données des fileyeurs. En outre, les deux années 2022 et 2023 ont été agrégées dans cette analyse préliminaire et pourrait être distinguées dans une analyse plus poussée. Le modèle développé à terme pourra être tester sur les données OBSMER qui offre un autre jeu de données, indépendant de celui collecté lors du projet OBSCAMe.

Les premières estimations obtenues ici pourraient aussi servir à quantifier, sous plusieurs hypothèses, le nombre de captures accidentelles de dauphins communs qui auraient pu se produire pendant la période de fermeture spatio-temporelle de l'hiver 2024.

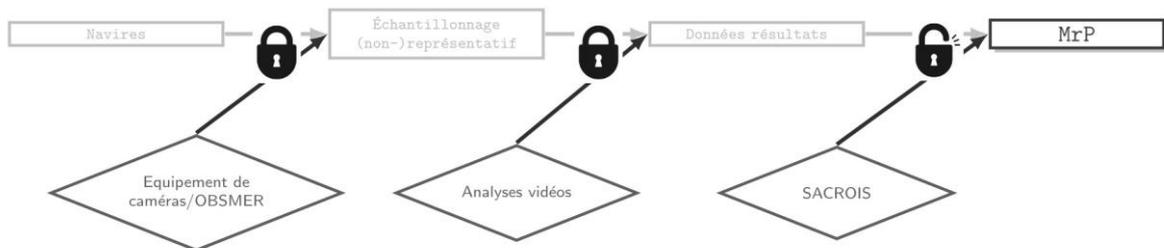
Enfin, il est important de noter que les résultats présentés ici sont préliminaires et reposent sur plusieurs hypothèses, dont que les données collectées sur les navires volontaires pour participer au projet OBSCAMe sont informatives sur les autres navires (*i.e.* pas de différences dans les styles de pêches entre navires volontaires et les autres). Sous l'hypothèse d'un biais de désirabilité sociale (c'est-à-dire un changement de comportement subséquent à la présence de caméras ; *e.g.* Benoît & Allard 2009) ou d'auto-sélection dans le projet OBSCAMe (c'est-à-dire que participer au projet reflète également la propension à avoir plus ou moins de captures accidentelles), les navires volontaires ne seraient pas représentatifs du reste des navires utilisant les engins GNS/GTR. Les estimations du risque de capture (Fig. 7) ne seraient alors pas généralisables à l'ensemble de la population de navires. Cette hypothèse est très plausible. Ainsi, la flottille des fileyeurs du large est sous-représentée au sein de l'échantillon OBSCAMe 2022-2023 : un seul fileyeur du large à merlu (*Merluccius merluccius*) est inclus dans l'échantillon alors que cette flottille représente un nombre de jours de mer conséquent ; et que cette pratique a été identifiée comme particulièrement à risque (Brevet et al. 2024). En outre, l'échantillonnage OBSCAMe est déséquilibré au niveau spatial avec une forte fréquentation du port de Cap-Breton par les navires volontaires par rapport aux autres ports

Estimer le nombre de captures accidentelles de dauphins communs dans les engins GNS/GTR

de la façade atlantique. Enfin, les estimations issues de la méthode MrP ici développées sont étroitement dépendantes du nombre de jours de mer fourni par SACROIS pour les fileyeurs. Si ce nombre de jour de mer n'est pas juste, cela se répercutera mécaniquement sur les estimations d'évènements de captures accidentelles et d'animaux capturés.

L'analyse ici présentée sur les données OBSCAME est préliminaire mais prometteuse. Elle a permis d'identifier les améliorations possibles du modèle en vue d'une application. A ce titre les estimations ici présentées ne sont donc pas encore utilisables. Pour cela il faudra lever certains verrous (Fig. 10).

Fig. 10 — Schéma des verrous à lever en vue d'une mise en œuvre analytique robuste pour estimer les captures accidentelles. On peut ainsi séquentiellement envisager une amélioration de la couverture d'échantillonnage en augmentant le nombre de caméras ou d'observateurs embarqués, une automatisation de l'analyse des flux vidéos pour pérenniser des résultats fiables et facilement accessibles, et s'assurer du rapprochement de ces derniers avec la base de données centralisé SACROIS afin d'assurer le bon fonctionnement du cadre analytique en vue d'une estimation avec la méthode MrP.



Si l'approche est prometteuse, elle doit néanmoins être encore explorée, ainsi que d'éventuelles alternatives analytiques en vue d'estimer le risque de capture à partir des données. Les explorations et développements en cours au sein du projet Delmoges restent nécessaires et pertinents. Par ailleurs, si des améliorations en termes de modélisation peuvent être recherchées, cette recherche ne doit pas éclipser qu'un échantillonnage représentatif faciliterait grandement les analyses. La représentativité des échantillons à collecter dans le futur est un enjeu crucial afin de lever des verrous.

Remerciements

Ce travail n'aurait pas été possible sans l'implication des fileyeurs volontaires ayant participé au projet OBSCAMe: ils sont ici remerciés. L'ensemble des ajustements de modèles n'aurait pas été possible dans les temps impartis sans les machines de calcul du mésocentre Aquitain (MCIA).

V. Bibliographie

Ben Hacine, M. & Bouallegue, R. (2015) Fitting the Log Skew Normal to the Sum of Independent Lognormals Distribution. ArXiv preprint <https://arxiv.org/pdf/1501.02344>

Benoît, H. P. & Allard, J. (2009) Can the Data from at-Sea Observer Surveys be Used to Make General Inference about Catch Composition and Discards? *Canadian Journal of Fisheries and Aquatic Sciences*, Vol. 66, p. 2025-2039

Brevet, M., Demanèche S., Peltier, H., Authier, M. & Dubroca, L. (2024) Description des activités de pêches, cartographie et typologie des flottilles opérant dans le golfe de Gascogne. Ref. WP3. Livrable L.3.11. Delmoges. <https://doi.org/10.13155/99926>

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A. (2017) *Stan*: A Probabilistic Programming Language. *Journal of Statistical Software*, Vol. 76, No. 1, p. 1-32

Gelman, A. & Little, T. (1997) Poststratification Into Many Categories Using Hierarchical Logistic Regression. *Survey Methodology*, Vol. 23, p. 127-135

Gelman, A. & Price, P. N. (1999) All Maps of Parameter Estimates are Misleading. *Statistics in Medicine*, Vol. 18, p. 3221-3234

Gelman, A., Hwang, J., Vehtari, A. (2014) Understanding Predictive Information Criteria for Bayesian Models. *Statistics and Computing*, Vol. 24, No. 6, p. 997-1016

Louis, T. A. & Zeger, S. L. (2009) Effective Communication of Standard Error and Confidence Interval. *Biostatistics*, Vol. 10, No. 1, p. 1-2

Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S. J., Gelman, A., DiMaggio, C. (2019) Bayesian Hierarchical Spatial Models: Implementing the Besag York Mollié Model in *Stan*. *Spatial and Spatio-Temporal Epidemiology*, Vol. 31, p. 100301

Pebesma, E., Bivand, R. (2023) **Spatial Data Science: With Applications in R**. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429459016>

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Rouby, E., Dubroca, L., Cloâtre, T., Demanèche, S., Genu, M., Macleod, K., Peltier, H., Ridoux, V. & Authier, M. (2022) Estimating Bycatch from Non-Representative Samples (II): a Case Study on Common Dolphins in the Bay of Biscay. *Frontiers in Marine Science*, Vol. 8, No. 795942

Shyalika, C., Wiwkramarachchi, R., Sheth, A. (2023) A Comprehensive Survey on Rare Event Prediction. University of South Carolina, University of South Carolina. preprint

Estimer le nombre de captures accidentelles de dauphins communs dans les engins GNS/GTR

Van der Paal, B. (2014) A Comparison of Different Methods for Modelling Rare Events Data. Master of Science Thesis. Faculty of Sciences, Department of Applied Mathematics, Computer Science and Statistics.

Vignard, C. & Tachouires, S. (2023) Projet OBSCAMe « *OB*Servation pour mieux comprendre les *Captures Accidentelles de Mammifères marins par les fileyeurs du golfe de Gascogne, au travers de l'observation électronique à distance* ». Brest : Office Français de la Biodiversité. <https://www.documentation.eauetbiodiversite.fr/notice/0000000018b43b4c5ce84ac876a2434>

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemond, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, P. D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C.,

Woo, K., Yutani, H. (2019) Welcome to the tidyverse. *Journal of Open Source Software*, Vol. 4, No. 43, p. 1686