

Année universitaire : 2020 - 2021

Diplôme : Ingénieur agronome

Spécialisation :

« *Sciences Halieutiques et Aquacoles* »
préparé à Agrocampus Ouest Rennes

Option :

« *Ressources et Ecosystèmes Halieutiques* »

Mémoire de fin d'études

- d'ingénieur d'AGROCAMPUS OUEST (École nationale supérieure des sciences agronomiques, agroalimentaires, horticoles et du paysage), école interne de L'institut Agro (Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement)
- de master d'AGROCAMPUS OUEST (École nationale supérieure des sciences agronomiques, agroalimentaires, horticoles et du paysage), école interne de L'institut Agro (Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement)
- de Montpellier SupAgro (étudiant arrivé en M2)
- d'ingénieur d'AgroSup Dijon (Institut national supérieur des sciences agronomiques, de l'alimentation et de l'environnement)

Modélisation de la distribution spatiale de la sardine du Golfe de Gascogne (*Sardina pilchardus*) par intégration de données commerciales et scientifiques : enjeux et limites.

Par : Florian QUEMPEL



Soutenu à Rennes le 14/09/2021

Devant le jury composé de :

Président : Olivier Le Pape

Maître de stage : Baptiste Alglave, Etienne Rivot,
Marie-Pierre Etienne

Enseignant référent : Olivier Le Pape

Autres membres du jury (Nom, Qualité) :

Hélène Peltier – Ingénieure de recherche à Pelagis

Nicolas Bez – Directeur de Recherche, IRD

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

Confidentialité

Non Oui si oui : 1 an 5 ans 10 ans

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

Date et signature du maître de stage ⁽²⁾ : 28/09/2021
(ou de l'étudiant-entrepreneur)



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

Droits d'auteur

L'auteur ⁽³⁾ Nom Prénom Quimper Florian
autorise la diffusion de son travail (immédiatement ou à la fin de la période de confidentialité)

Oui Non

Si oui, il autorise

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

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

Date et signature de l'auteur :

 28/09/2021

Autorisation de diffusion par le responsable de spécialisation ou son représentant

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

Oui Non

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

Si oui, il autorise

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

Date et signature de l'enseignant :

28 Sep 2021



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

(2) Signature et cachet de l'organisme

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

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

« Si les mouettes suivent le chalutier, c'est parce qu'elles pensent que des sardines seront jetées à la mer »

E. Cantona – 1995

Remerciements

Je tiens dans un premier temps à remercier sincèrement Baptiste Alglave pour l'aide précieuse qu'il m'a apporté tout au long de ce stage. Toujours présent pour répondre à mes questions (même les plus idiotes, surtout les plus idiotes), ça aurait sans aucun doute été d'une bien plus grande difficulté sans toi. Encore merci Baptiste. Franchement j'insiste sans plaisanter, c'est vraiment super agréable de travailler avec toi, t'es patient, humble et quand même vachement doué, même au rugby tu donnes de bons conseils !

Je souhaite ensuite remercier Etienne Rivot et Marie Pierre Etienne, également pour leur suivi, leurs propositions et conseils avisés qui nous ont permis à moi et Baptiste d'avancer dans le stage le plus sereinement possible. De même, je remercie les différentes personnes qui ont pu nous apporter leur aide tout au long de ce stage ainsi que ceux nous ayant fourni les différents jeux de données : Mathieu Doray, Thomas Opitz, Youenn Vermard, Mathieu Woillez et Guillermo Boyra.

Vient ensuite l'ensemble des stagiaires, on ne se sera pas beaucoup vu les 3 premiers mois grâce à ce bienheureux confinement. Malheureusement nous avons ensuite été forcé de nous côtoyer, de sortir et d'aller au bar. On aura même un peu dansé.

Pas toujours facile de vous sortir de votre timidité, en particulier toi Gasparccio, franchement t'abuses il faut tout le temps te pousser. En plus depuis que je t'ai montré « Me llaman calle » tu t'arrêtes plus de la chanter. Je reparle de toi un peu plus loin quand je cause de la Cocoflo'c et de La Rochelle.

Un p'ti coucou à Lulu d'ses morts, j'ai hâte de recevoir un autre de tes messages le 14 juillet 2022 à midi, jour béni de la naissance de ton petit frère, Paul, et nous expliquant ô combien tu es farcie comme une tomate après avoir gaiement festoyer en admirant un feu d'artifice, à tel point qu'il t'est désormais impossible d'avaler le moindre aliment au restaurant. Dommage.

La bizette à Daphneux, à qui je décerne le titre de « révélation alcoolisée de ces 6 mois de stage » tellement ta furtive mais ô combien remarquable apparition aura été marquante dans la courte période de ce stage. Franchement c'est avec plaisir qu'on remet le couvert, d'ailleurs je vais de ce pas t'envoyer un message pour te motiver à venir claquer une canouche chez les amis de Lulu demain soir. Allez Zag comme dirais l'autre.

Plus brièvement, je fais un petit salut depuis cette page aux autres stagiaires, Florian, Jérémy et Sully (je sens que je vais bientôt plus avoir de place, en fait il faudrait que je fasse un rapport de remerciements)

Ensuite je serre chaleureusement la main à mes 2 collocs de la « Versailles sans bouchon des chiottes loc' », dont l'un a été mon collègue aux archives du labo, le bien sus-nommé Thomas. En vrai c'était sympa d'être avec toi dans le cagibi du bat' Halieut, on aura quand même pas mal discuté et tu m'auras aussi filé quelques conseils et idées, c'était chouette ! Dommage que tu aies trouvé l'amour durant ce stage, on se sera moins cotoyé que prévu. M'enfin je rentrais tous les week-ends aussi donc ça n'a pas aidé. Bref ! Une grosse bise au second colloc', l'heureux Quenting, je ne dirais rien de plus que : « La Carambole ! ». Si je vais dire un peu plus en fait, c'était là encore un plaisir de passer du temps à discuter avec toi (je me souviens notamment de cette soirée où je suis rentré à 1h du mat de l'agro et tu es sorti de ta chambre pour discuter, sacré toi tiens). Je pense qu'il m'arrivera de revenir toquer à ta porte en disant « Oh Quenting... ? ».

Je garde un dernier petit paragraphe, parce que j'ai été trop – Bordel Mikaëla j'allais t'oublier ! Enfin bon de toute façon tu restes dans le couloir – je disais donc trop exhaustif peut-être, pour les copains de l'agro Rennes et Dijon - en particulier mon chto Rem's et mon Coco qui comme moi ont quitté les terres Dijonnaises pour rejoindre ACO, avec qui on aura fait les loubards et qui ont été des véritables soutiens sans s'en rendre compte, j'ai plus de place mais vous méritez une page chacun pour toutes nos aventures - de prépa, de PeB, et ma famille évidemment, et en particulier mon père, qui m'a épaulé tout au long de ce stage, ma mère, mes 3 frères et mes grands-parents (mention gâteaux pour Mamignonne !). Je ne veux ni exagérer et faire plus d'une page de remerciements ni trop détailler mes sentiments pour ma famille et mes amis proches ici, mais sincèrement, merci. Je vous porte dans mon cœur.

Alllezzz La Rochelllllee !

Une pensée pour Jérôme Quinquis.

Table des matières

1. Introduction	1
1.1. Généralités : distribution d'espèce et SDM.....	1
1.2. Données scientifiques et données commerciales.....	2
1.3. Des modèles hiérarchiques pour intégrer différentes sources de données	3
2. Matériel et méthode.....	5
2.1. Description du cas d'étude : la sardine du Golfe de Gascogne.	5
2.2. Données	6
2.2.1. Données scientifiques issues de campagnes acoustiques	6
2.2.2. Données commerciales (logbooks & VMS).....	7
2.3. Spécificités du cas d'étude et conséquence pour la modélisation	9
2.4. Traitement des données	10
2.5. Modèle intégré hiérarchique.....	10
2.5.1. Champ latent de présence-absence.....	11
2.5.2. Modèle d'observation de la présence/absence.....	12
2.5.3. Modéliser l'échantillonnage préférentiel.....	12
2.5.4. Simplification induite par les contraintes d'INLA	13
2.5.5. Outil d'inférence : R-INLA	13
2.5.6. Métriques de validation	15
2.5.7. Démarche de valorisation du modèle	15
3. Résultats	16
3.1. Analyse à l'échelle d'un mois : mai 2018	16
3.1.1. Apport des différentes sources de données à l'inférence.....	16
3.1.2. Intensité de l'échantillonnage préférentiel	19
3.1.3. Ajout des co-variables environnementales.	20
3.1.4. Comparaison des capacités prédictives selon les différents modèles.....	22
3.2. Analyse spatio-temporelle	23
4. Discussion	28
4.1. Une inférence à l'échelle du mois dirigée par la donnée scientifique	28
4.2. Capacité prédictive du modèle	29
4.3. Mise en perspective des cartes de distribution avec les connaissances déjà disponibles	29
4.4. Limites de l'étude.....	30
4.4.1. Spécificités de l'activité de pêche à la sardine dans le Golfe de Gascogne.....	30
4.4.2. Limites de la modélisation de l'échantillonnage préférentiel.....	31
4.4.3. Estimation de l'effet des covariables environnementales.....	31
4.4.4. Biais liés aux données.	32
5. Conclusion.....	33
Références	35
Annexes.....	43

Liste des Figures

Figure 1 : Carte du Golfe de Gascogne et délimitation du domaine étudié.

Figure 2 : Observations (présence/absence) de *S. pilchardus* par la campagne scientifique PELGAS au cours du mois de Mai 2018 et par la campagne scientifique JUVENA au cours du mois de Septembre 2018.

Figure 3 : Observations (présence/absence) de *S. pilchardus* par les 3 flottilles commerciales entre janvier et décembre 2018

Figure 4 : Diagramme du modèle spatio-temporel hiérarchique intégré (d'après Alglave et al ; under review)

Figure 5 : Maille d'interpolation triangulaire et points de données dans le GdG (Mai 2018).

Figure 6 : Cartographies des probabilités de présence et des observations par les flottilles commerciales et la campagne scientifique PELGAS en Mai 2018.

Figure 7 : Cartographies des probabilités de présence (colonne de gauche) et écarts-types associés (colonne de droite) au mois de Mai 2018 obtenues à partir des données scientifiques, des données commerciales et dans un modèle intégré les combinant.

Figure 8 : Cartes d'inférence du modèle avec prise en compte d'un échantillonnage préférentiel dans le fonctionnement des flottilles commerciales, les écarts types associés et de la différence avec le modèle intégré ne prenant pas en compte d'E.P. (E.P – Intégré)

Figure 9 : Effet des co-variables environnementales (Bathymétrie, SST et Chlorophylle A) dans un modèle prenant en compte des données satellitaires ou issues des sorties du modèle physico-biogéochimique POLCOM-ERSEM.

Figure 9 : Cartes d'inférences en prenant en compte les co-variables environnementales (Bathymétrie, SST et Chlorophylle A issue des données satellitaires), les écarts types associés et de la différence avec le modèle intégré 'simple'. (Covariables – Intégré).

Figure 11 : Capacités prédictives des différents modèles par source de données, AUC à gauche et CPO à droite.

Figure 12 : Cartographie mensuelle de la probabilité de présence de la sardine dans le GdG en 2018

Figure 13 : Cartographie mensuelle de l'écart-type de la probabilité de présence de la sardine dans le GdG en 2018.

Figure 14 : Score AUC du modèle spatio-temporel par mois et source de donnée

Figure 15 : AUC moyenné par mois et source de donnée (2009-2018) et écarts types. Le trait rouge correspond à une AUC de 0.70.

Figure 16 : Cartographies de la probabilité de présence de la sardine et d'écarts-types associés dans le GdG entre mai et octobre 2016

Liste des Tableaux

Tableau 1 : Intérêts et limites de l'approche intégrée

Tableau 2 : Différences entre cas d'étude pélagique et benthodémersaux

Liste des Annexes

Annexe 1 : Analyse exploratoire des données commerciales : Spatialisation des CPUE en Mai 2018 de la flottille PTM_SPF.

Annexe 2 : Analyse exploratoire des données commerciales : Spatialisation des CPUE en Mai 2018 de la flottille PTM_DEF.

Annexe 3 : Analyse exploratoire des données commerciales : Spatialisation des CPUE en Mai 2018 de la flottille PS_SPF.

Annexe 4 : Séries temporelles des profils de débarquements entre 2008-2018 à un pas de temps mensuel pour les 3 flottilles commerciales.

Annexe 5 : Cartographie de la SST en Mai 2018 (données issues du modèle biogéochimique POLCOM-ERSEM à gauche et satellitaires à droite).

Annexe 6 : Cartographie de la chlorophylle A en Mai 2018 (données issues du modèle biogéochimique POLCOM-ERSEM à gauche et satellitaires à droite).

Annexe 7 : Cartographie de la bathymétrie en Mai 2018.

Annexe 8 : Fonctionnement des PC priors et récapitulatif des paramétrisations testées

Annexe 8 : Tableau récapitulatif des différentes modélisations testées (toutes les combinaisons n'ont pas été testées).

Annexe 9 : Maillages utilisés

Annexe 10 : Inférences obtenues pour chaque variables environnementale (SST, chlorophylle A, bathymétrie).

Annexe 11 : Comparaison effet des conditions environnementaux avec ou non une prise en compte de l'effet aléatoire (données satellitaires).

Annexe 14 : Courbes ROC et AUC pour chaque source de données et modèles.

Annexe 15 : Cartographies des CPO pour chaque source de donnée issue du modèle « donnée seule ». Les CPO sont représentées sur les cartes d'inférences respectives.

Annexe 16 : Cartographies des CPO pour chaque source de donnée et issue du modèle « donnée seule », et différence avec la CPO de ce modèle.

Annexe 18 : Scores AUC des modèles spatio-temporels par mois et source de donnée entre 2009 et 2018. Le trait rouge correspond à une AUC de 0.70

Annexe 19 : Cartographies de probabilité de présence de la sardine dans le GdG en 2009.

Annexe 20 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2009.

Annexe 21 : Cartographies de probabilité de présence de la sardine dans le GdG en 2010

Annexe 22 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2010.

Annexe 23 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2011.

Annexe 24 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2011.

Annexe 25 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2012.

Annexe 26 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2012.

Annexe 27 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2013

Annexe 28 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2013.

Annexe 29 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2014

Annexe 30 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2014.

Annexe 31 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2015.

Annexe 32 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2015.

Annexe 33 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2016

Annexe 34 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2016.

Annexe 35 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2017.

Annexe 36 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2017.

Annexe 37 : Cartographies moyennes de la probabilité de présence de la sardine dans le GdG entre 2009-2018

Annexe 38 : Cartographies moyennes des écarts types de la probabilité de présence de la sardine dans le GdG entre 2009-2018.

Annexe 39 : Cartographies de probabilités de présence d'œufs de sardines entre 2000 et 2004 (Bellier et al, 2007), à gauche cartes de présence moyenne et coefficients de variations à droite.

Annexe 40 : Cartographies d'abondance d'œufs de sardines (2000-2016) (Huret et al, 2018), à gauche cartes de d'abondances moyennes et coefficients de variations à droite.

Annexe 41 : Cartographies de biomasse de sardines (2000-2015) (Doray al, 2017a), à gauche cartes de présence moyenne et coefficients de variations à droite.

Annexe 42 : Distribution moyenne des abondances d'œufs de sardines réparties en 2 groupes à partir d'un cluster sur 13 MAFS et exemples de 4 cas représentatifs des 2 groupes (G1 au-dessus : 2000 et 2008 ; G2 en-dessous : 2010 et 2017) après krigeage à partir du modèle MAF (Données PELGAS 2000- 2017) (Petitgas *et al.*, 2020).

Liste des abréviations

ADNe : Acide Désoxyribo-Nucléique environnemental

AUC : Area Under the Curve

AMP : Aires Marines Protégées

CPO : Conditional Predictive Ordinate

CPUE : Captures Par Unité d'Effort

EMR : Energies Marines Renouvelables

ESDU : Unité horizontale d'échantillonnage élémentaire, "Elementary Sampling Distance Unit", un ESDU correspond à un mile nautique d'échantillonnage acoustique

GdG : Golfe de Gascogne

GPS : Système de Positionnement Global

IFREMER : Institut Français de Recherche pour l'Exploitation de la Mer

INLA : Integrated Nested Laplace Approximation

MCMC : Markov Chain Monte Carlo

PELGAS : PELagique GAScogne

PTM_DEF : Pelagic Pair Trawl Demersal pelagic Fish

PTM_SPF : Pelagic Pair Trawl Small Pelagic Fish

PS_SPF : Purse Seine Small Pelagic Fish

ROC : Receiver operating characteristic

SDM : Species Distribution Model

SST : Sea Surface Temperature

VMS : Vessel Monitoring Systems

1. Introduction

1.1. Généralités : distribution d'espèce et SDM

La bonne compréhension de la distribution des espèces marines est fondamentale pour la compréhension du fonctionnement des populations et des écosystèmes. Elle présente des enjeux majeurs pour la gestion et la conservation des espèces marines (Gaston, 2000; Lamoreux *et al.*, 2006). L'établissement de plans de gestion spatialisés (Wright and Kyhn, 2015), notamment les AMP, les parcs éoliens en mer (EMR) ou encore la délimitation de zones de pêches (Hunter *et al.*, 2006; Meyer *et al.*, 2007) nécessitent une connaissance fine de la distribution spatiale et saisonnière des organismes marins de manière générale et des stocks exploités en particulier (Booth, 2000).

Dans les écosystèmes marins, de nombreux facteurs affectent la répartition des poissons dans le temps et dans l'espace (Whittaker *et al.*, 1973) et rendent la détermination des aires de distribution complexe. En effet, les besoins d'une espèce varient selon un ensemble de conditions biotiques et abiotiques (habitat physique, relations trophiques, conditions physico-chimiques) desquelles vont dépendre leur distribution spatiale. Par ailleurs, leur cycle de vie se divise en différentes phases, correspondant chacune à des besoins précis et peut s'accompagner de changements d'habitats au cours de l'ontogénie (Delage et Le Pape, 2016). Enfin, l'homme est l'un des principaux prédateurs des espèces marines et les impacts liés aux activités anthropiques peuvent considérablement modifier la structure des communautés de poissons et leur répartition (Worm *et al.*, 2009). Notamment, l'exploitation par la pêche n'a pas lieu au hasard dans l'espace et le temps. Les pêcheurs sélectionnent le lieu et le moment de la pêche en fonction de leurs connaissances de l'écologie des espèces et des règles qui régulent l'activité de pêche (Branch *et al.*, 2006; Vermard *et al.*, 2008; van Putten *et al.*, 2012). En conséquence, la distribution des espèces marines peut potentiellement changer sur des périodes relativement courtes, le changement de distribution reflétant l'effet conjoint de l'activité de pêche, de l'environnement et du cycle biologique des espèces.

Les modèles de distribution d'espèces (SDM) sont des outils essentiels en écologie marine pour améliorer notre connaissance de la distribution des espèces et de la relation entre espèces et habitats. Ils permettent notamment de comprendre l'influence de l'environnement et des pressions anthropiques sur la distribution des espèces (Marshall *et al.*, 2014). Aujourd'hui, le développement des nouvelles technologies et techniques d'informations comme les ADNe (Bálint *et al.*, 2018), les suivis par acoustique passive (Gibb *et al.*, 2019) ou encore la multiplication des campagnes de science participative (August *et al.*, 2015), permet d'accéder à une grande diversité de sources de données, et présente de nouvelles opportunités pour cartographier les espèces. Cependant, la diversité des sources de données peut rendre leur intégration dans un seul et même modèle complexe (i.e. les données ne sont pas nécessairement de même nature – e.g. données de comptage, de présence-absence, de présence seulement). Un challenge important consiste à développer de nouvelles méthodologies permettant d'intégrer ces différentes sources de données afin d'en tirer une information synthétique sur la distribution des espèces (Isaac *et al.*, 2020).

1.2. Données scientifiques et données commerciales

En halieutique, les données de référence pour cartographier les espèces marines sont généralement issues de campagnes océanographiques. Ces campagnes scientifiques ont pour but de récolter des données représentatives des populations étudiées via un plan d'échantillonnage et une capturabilité standardisée. Cependant, ces campagnes ont un coût élevé et par conséquent, ont généralement une couverture temporelle restreinte (en général, elles ont lieu une à deux fois par an toujours à la même période) (Hilborn and Walters, 1992; ICES, 2005; Nielsen, 2015). Elles ont donc une utilité limitée pour étudier la dynamique intra-annuelle des espèces étudiées.

De nouvelles sources de données sont aujourd'hui rendues disponibles par le biais des dispositifs de surveillance des navires de pêche (Vessel Monitoring Systems – VMS). D'abord mis en place au Portugal en 1988 afin de suivre l'activité des pêcheurs suite à la dégradation des stocks (Navigs, 2005), le système VMS s'est rapidement généralisé depuis le début du 21^{ème} siècle (Marzuki, 2017). Ces dispositifs, aujourd'hui installés à bord de tous les navires sous pavillon européen de plus de 12 m, transmettent des données (position GPS, cap, vitesse, identification du navire...) par communication satellitaire toutes les heures aux centres de surveillance des pêches. Les points VMS (ou pings) peuvent ensuite être utilisés pour calculer un effort de pêche en identifiant les points VMS émis lors de l'activité de pêche sur la base de différents critères ou méthodes (Vermard *et al.*, 2008; Bez *et al.*, 2011; Gerritsen and Lordan, 2011; Hintzen *et al.*, 2012; Murray *et al.*, 2013).

En parallèle à l'acquisition de ces données VMS, les données de déclaration de captures sont déclarées dans les livres de bord des navires de pêche (données « logbooks »), dans lesquels les pêcheurs indiquent les quantités capturées par espèce à l'échelle d'un carré statistique (de 1° longitude par 0.5° de latitude) pour chaque journée de pêche, chaque engin utilisé et chaque marée. Cependant, la résolution spatiale de ces données (le carré statistique) demeure assez grossière ce qui limite leur utilisation pour des analyses à fine échelle spatiale (Marrs *et al.*, 2002; Lee *et al.*, 2010).

Chaque donnée de déclaration de pêche (quantités capturées par espèce) peut être associée à un ensemble de pings VMS identifiés en pêche (position de pêche des navires de pêche). Ainsi, en ventilant les données logbooks sur les pings VMS correspondant, il est possible de passer d'une image de la distribution des captures et des CPUE grossière, définie sur un domaine discret (i.e. les carrés statistiques) à une résolution spatio-temporelle fine et continue dans l'espace (Gerritsen and Lordan, 2011 ; Murray *et al.*, 2013). Ainsi, le croisement des données VMS et des données logbooks constituent une source de données additionnelle pour décrire la distribution des ressources exploitées. Elle est notamment susceptible de fournir de l'information en dehors des périodes couvertes par la donnée scientifique.

Les données commerciales présentent cependant un inconvénient majeur : les pêcheurs vont avoir tendance à cibler des zones poissonneuses où les abondances sont les plus élevées (Conn *et al.*, 2017; Pennino *et al.*, 2019; Rufener, 2020) afin de maximiser leurs captures et leurs revenus (Vermard *et al.*, 2008). Ce phénomène, appelé "échantillonnage préférentiel", introduit un biais systématique dans les estimations et les prédictions de modèles qui peuvent entraîner une surestimation de l'abondance dans les zones où l'échantillonnage est faible (Conn *et al.*, 2017; Pennino *et al.*, 2019; Alglave *et al.*, under review).

Par ailleurs, d'autres composantes peuvent intervenir dans les prises de décisions des pêcheurs et la distribution de l'effort de pêche (Salas and Gaertner, 2004; Abbott *et al.*, 2015; Girardin *et al.*, 2017; Stephenson *et al.*, 2018). En particulier, leur activité de pêche peut être fortement structurée par les exigences du marché et des effets filières limitant la dispersion de l'activité de pêche à des saisons et/ou des zones plus ou moins restreintes.

1.3. Des modèles hiérarchiques pour intégrer différentes sources de données

L'intérêt de développer des modèles hiérarchiques intégrés pour combiner différentes sources de données est maintenant bien établi en écologie (Schaub *et al.*, 2007; Zipkin *et al.*, 2019) et en sciences halieutiques en particulier (Maunder and Punt, 2013; Rochette *et al.*, 2013; Archambault *et al.*, 2018; Punt *et al.*, 2020; Rufener, 2020). En effet, en considérant que différentes sources de données sont des réalisations d'un même processus sous-jacent modélisé sur la base de variable latentes non directement observées, il est possible de combiner différentes sources de données dans un même modèle et ainsi d'améliorer l'estimation des paramètres du modèle, d'étendre des séries temporelles, d'élargir l'aire d'étude par rapport à une analyse séparée de différentes sources de données (Moriarty *et al.*, 2020).

Plusieurs auteurs ont déjà développé des modèles combinant données commerciales et données scientifiques pour tirer profit des deux sources de données - et ainsi prédire la distribution des espèces d'intérêt halieutique à une résolution spatio-temporelle plus fine (Alglave *et al.*, under review; Rufener, 2020; Gonzalez *et al.*, 2021). En particulier, le modèle développé par Alglave *et al.* (under review) permet de combiner les données scientifiques issues de campagnes de chalutage et les données 'logbooks x VMS', tout en prenant en compte le comportement de ciblage des pêcheurs. Cette approche a été appliquée à plusieurs espèces démersales du Golfe de Gascogne (en particulier la sole ainsi que d'autres espèces telles que la baudroie, le merlu, l'encornet). Il en ressort que les inférences dépendent principalement des données commerciales, en comparaison à la donnée scientifique qui n'a qu'une faible influence du fait du nombre restreint d'échantillons (en ordre de grandeur, 3000 échantillons commerciaux contre 50 échantillons scientifiques). De plus, il apparaît que les engins filtrés pour décrire la distribution des espèces d'intérêt (ici les chalutiers démersaux) ont un comportement opportuniste et n'ont pas un comportement de ciblage élevé pour une espèce en particulier (la proportion des espèces majoritaires dans les débarquements n'excède pas 20 – 25 % en général). Ainsi, le comportement de ciblage des pêcheurs, et en particulier l'effet de l'échantillonnage préférentiel vis-à-vis de la ressource, n'a que peu d'impact sur les prédictions spatiales des modèles.

Dans ce travail, nous proposons d'explorer l'intérêt du développement de modèles de distribution d'espèces intégrés combinant données scientifiques et donnée 'VMS x logbooks' pour des espèces de petits pélagiques comme l'anchois ou la sardine. Les petits poissons pélagiques (e.g. sardine, anchois, sprat) jouent un rôle écologique clé dans les écosystèmes côtiers, en transférant l'énergie du plancton aux niveaux trophiques supérieurs (Cury *et al.*, 2000; Spitz *et al.*, 2018). Par ailleurs, leur position trophique relativement basse dans le réseau alimentaire marin, ainsi que leur courte durée de vie et leur stratégie de reproduction consistant à produire de grandes quantités d'œufs pélagiques au cours d'une saison de frai prolongée, les rendent très dépendants de l'environnement (Hunter, and Alheit, 1995; Bakun, 1996).

Par ailleurs, la caractérisation de leur distribution soulève des enjeux méthodologiques importants. Les petits poissons pélagiques sont très mobiles par rapport aux espèces démersales et benthiques et effectuent des migrations plus fréquentes et plus importantes, à des échelles

allant de quelques jours à quelques semaines et de quelques dizaines de mètres à quelques dizaines de kilomètres, afin de faire face aux fluctuations plus importantes des conditions hydrobiologiques de l'écosystème pélagique (Mackinson, 1999; Robinson, 2004; Barange *et al.*, 2005). Enfin, ils font l'objet d'un ciblage fort de la part des flottilles qui les exploitent, ce qui renforce l'intérêt de développer des méthodes permettant de prendre en compte explicitement l'échantillonnage préférentiel pour l'intégration des données commerciales.

Le cas d'étude abordé dans ce travail est la sardine européenne (*Sardina pilchardus*) dans le Golfe de Gascogne. La sardine est un clupéidé sténotherme au comportement grégaire qui habite généralement des eaux dont la température varie de 8°C à 24°C et la salinité de 30 à 38 psu (Coombs *et al.*, 2006; Petitgas *et al.*, 2006; Stratoudakis *et al.*, 2007). Suivant son âge et son état physiologique elle recherche des conditions environnementales différentes, rendant sa répartition dans le GdG discontinue et variable.

L'étude des populations de sardine en GdG a commencé au début du siècle précédent, mais n'a été l'objet d'un échantillonnage systématique le long du plateau continental qu'à partir des années 1960 (cartographies de larves et des œufs de sardine - L'Herrou, 1967; Arbault and Lacroix, 1970). Ces premières campagnes ont d'abord été utilisées par Arbault et Lacroix (1970, 1977) pour réaliser les premières cartographies des zones de frayères de sardine en apportant des indications sur les secteurs de ponte, la saison de frai ainsi que les conditions favorables du milieu. Ces premières études présentent des résultats sur l'ensemble de l'année (une carte par saison) et sont les seules disponibles à ce jour. Par la suite les cartographies ont été réalisées à partir des données issues de la campagne PELGAS qui a lieu au printemps de chaque année (la campagne est centrée sur le mois de mai). La donnée scientifique pour décrire la répartition de la sardine est donc essentiellement limitée au mois de mai et, à notre connaissance, aucun travail n'a cherché à inférer la distribution de la sardine en dehors de cette période depuis les années 1970.

Ainsi ce travail a pour objectif de construire et d'ajuster un modèle combinant données scientifiques (PELGAS au printemps et JUVENA en automne) et données 'VMS x logbooks' pour inférer la distribution de la sardine du GdG sur l'ensemble de l'année. Ce travail doit nous permettre :

- D'appliquer un modèle intégré à des cas de d'étude différents des espèces étudiées jusqu'à maintenant (i.e. des espèces benthodémersales) et de voir dans quelle mesure les conclusions sont modifiées pour le cas d'espèces pélagiques.
- D'explorer l'apport des données commerciales pour inférer la distribution de la sardine en dehors de la période de campagne.

Le modèle est développé sur la base du cadre de modélisation intégré décrit par Alglave *et al.* (under review), et adapté pour son application à un cas d'étude de petit pélagique.

Dans un premier temps, ce modèle est développé pour inférer la distribution sur un seul pas de temps (pas de temps mensuel) afin d'en évaluer plusieurs caractéristiques :

- (i) la contribution des différentes sources de données dans l'inférence ;
- (ii) l'effet de la prise en compte de l'échantillonnage préférentiel des flottilles commerciales ;
- (iii) l'apport de covariables environnementales sur les prédictions du modèle.

Dans un second temps, nous avons étendu les analyses à la période 2009-2018 (toujours sur un pas de temps mensuel) pour explorer la variabilité temporelle de la distribution spatiale de la sardine.

2. Matériel et méthode.

2.1. Description du cas d'étude : la sardine du Golfe de Gascogne.

La sardine européenne, *Sardina pilchardus* (Walbaum, 1792) est une espèce pélagique appartenant à l'ordre des clupéiformes, la famille des *Clupeidae*, et la sous-famille des *Alosinae*. L'aire de répartition de la sardine s'étend des côtes mauritaniennes (Furnestin and Furnestin, 1959; Ettahiri *et al.*, 2003) jusqu'aux eaux de la Norvège en incluant la mer Méditerranée et la Mer Noire (Parrish *et al.*, 1989). Les sardines vivent en bancs très denses. Elles sont proches du fond le jour et elles remontent en surface en se dispersant la nuit (Whitehead, 1985). On la retrouve au-dessus du plateau continental jusqu'à des profondeurs de 120 m, mais elle abonde surtout le jour entre 30 et 55 m et la nuit entre 15 et 40 m de la surface (Quero *et al.*, 1989). La sardine effectue des déplacements saisonniers de faible amplitude régies par des besoins nutritionnels, reproducteurs ou en lien avec les conditions thermiques. Elle migre en automne vers le large et se rapproche des côtes au printemps. Selon la saison, l'âge et l'état sexuel elle réalise également des déplacements le long des côtes (Tosello-Bancal, 1994).

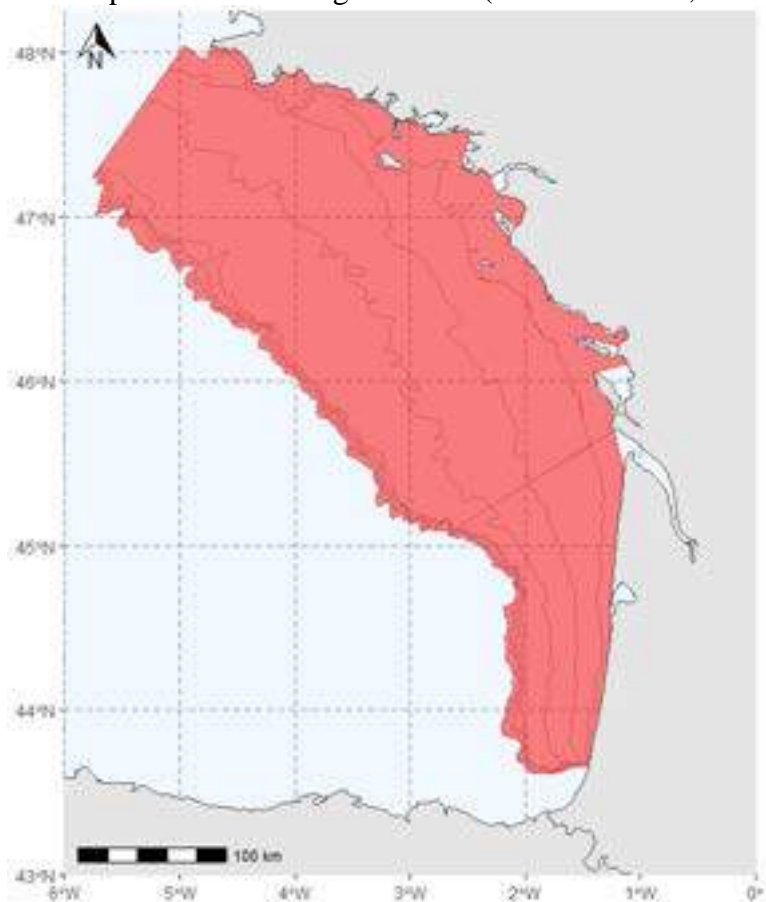


Figure 1 : Carte du Golfe de Gascogne et délimitation du domaine étudié.

A ce jour, les études suggèrent des populations auto-suffisantes (uniques et sans échanges avec les autres) distinctes en Manche/Mer Celtique et dans le Golfe de Gascogne (Gatti *et al.*, 2018; Laviolle *et al.*, 2019). En revanche la comparaison des tailles et structures en âge semblent indiquer un lien entre les cohortes observées en Mer Cantabrie et les recrutements dans le GdG, suggérant ainsi l'existence d'une connectivité entre ces zones (Silva *et al.*, 2009; Silva *et al.*, 2019). Néanmoins, en l'absence de certitudes sur l'intensité des flux entre ces zones, le

CIEM réalise une évaluation en considérant un stock de sardine indépendant pour le GdG (ICES, 2016). Elle n'est actuellement soumise à aucune réglementation commerciale, en dehors d'une taille minimale de capture fixée à 11 cm. Surpêchée depuis 2014, sa dernière évaluation indique que ce stock a enfin atteint le bon état écologique en 2019, avec 24 400 T capturées. (ICES, 2020)

Cette étude se limite à la zone dite française du GdG (Figure 1). Cette zone correspond au domaine échantillonné par la campagne scientifique PELGAS. Elle comprend le plateau continental situé entre la pointe du Raz et la fosse du Cap Breton.

2.2. Données

2.2.1. Données scientifiques issues de campagnes acoustiques

Deux campagnes scientifiques ont lieu chaque année dans le GdG (Doray *et al.*, 2018b; Boyra *et al.*, 2020). Leur objectif principal est d'estimer la biomasse et de prédire le recrutement des petits pélagiques. PELGAS a lieu au printemps depuis 2000. Elle est centrée sur le mois de mai et est réalisée par l'IFREMER à bord du RV "Thalassa". La campagne JUVENA a lieu depuis 2003 et est centrée sur le mois de septembre (Figure 2). Elle a un focus sur les populations de juvéniles d'anchois et couvre les côtes espagnole et française. Seuls les échantillons situés sur le domaine d'étude ont été utilisés dans l'analyse (i.e. les échantillons situés dans le GdG). Son plan d'échantillonnage est adaptatif et est donc susceptible de varier entre les années : les limites de la zone échantillonnée sont redéfinies chaque année afin de couvrir l'ensemble de l'aire de distribution de l'anchois, connue via les données commerciales et les précédentes campagnes scientifiques.

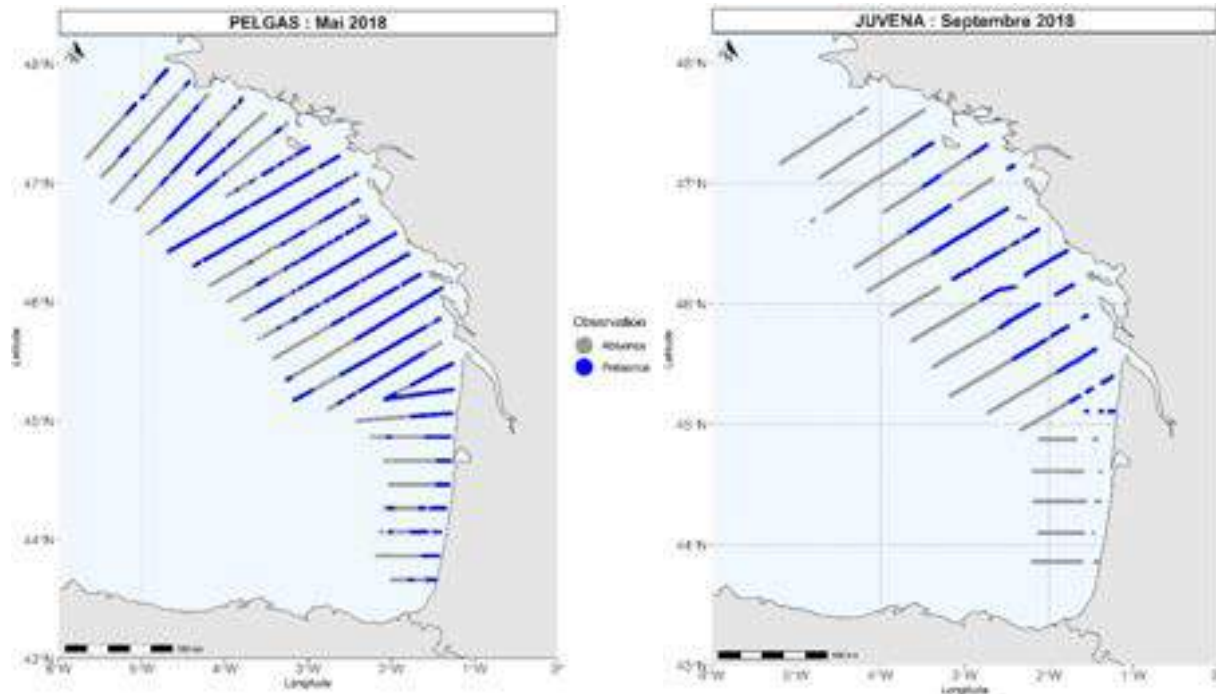


Figure 2 : Observations (présence/absence) de *S. pilchardus* par la campagne scientifique PELGAS au cours du mois de Mai 2018 et par la campagne scientifique JUVENA au cours du mois de Septembre 2018.

PELGAS et JUVENA sont des campagnes acoustiques qui utilisent un écho-sondeur pour émettre de courtes impulsions électriques, transmises sous forme d'impulsions ultrasonores, vers les fonds marins. En présence de poissons, une partie de l'énergie est rétrodiffusée vers l'émetteur. L'écho-intégration permet d'évaluer la biomasse présente dans une zone à partir du

cumul de l'énergie acoustique rétrodiffusée par l'ensemble des cibles présentes (Maclennan *et al.*, 2002). Ces écho-intégrations sont réalisées à l'échelle d'une ESDU (Unité horizontale d'échantillonnage élémentaire, "Elementary Sampling Distance Unit") qui correspond à un mile nautique d'échantillonnage acoustique. Les densités acoustiques, issues de l'écho-intégration, correspondent à un assemblage d'espèces.

Pour déterminer la composition en espèce et en taille des ESDU, des chalutages pélagiques sont effectués régulièrement (2 à 3 chalutages par jour) et la composition en espèce et en taille des coups de chaluts sont réalloués aux écho-intégrales par ESDU soit : (i) par utilisation d'un trait de chalut de référence, (ii) en définissant des régions où les assemblages espèces/tailles sont homogènes, (iii) ou à l'aide d'un modèle géostatistique, une estimation est obtenue aux nœuds d'une grille et est suivie d'une interpolation (Doray *et al.*, 2010).

La majorité des poissons pélagiques étant souvent hors de portée du sondeur (dans une couche d'eau entre la surface et 10 mètres d'immersion) pendant les périodes de nuit, la prospection acoustique se réalise essentiellement de jour (Doray *et al.*, 2018b; Boyra *et al.*, 2020).

Pour les données PELGAS, on obtient ainsi des biomasses et des abondances par espèce, classe de taille, et ESDU. Pour JUVENA seule une biomasse par espèce et ESDU est indiquée.

La taille réglementaire de capture de la sardine étant fixée à 11 cm, seules les observations scientifiques supérieures à cette taille sont conservées pour la suite des analyses si le jeu de données le permet.

2.2.2. Données commerciales (logbooks & VMS)

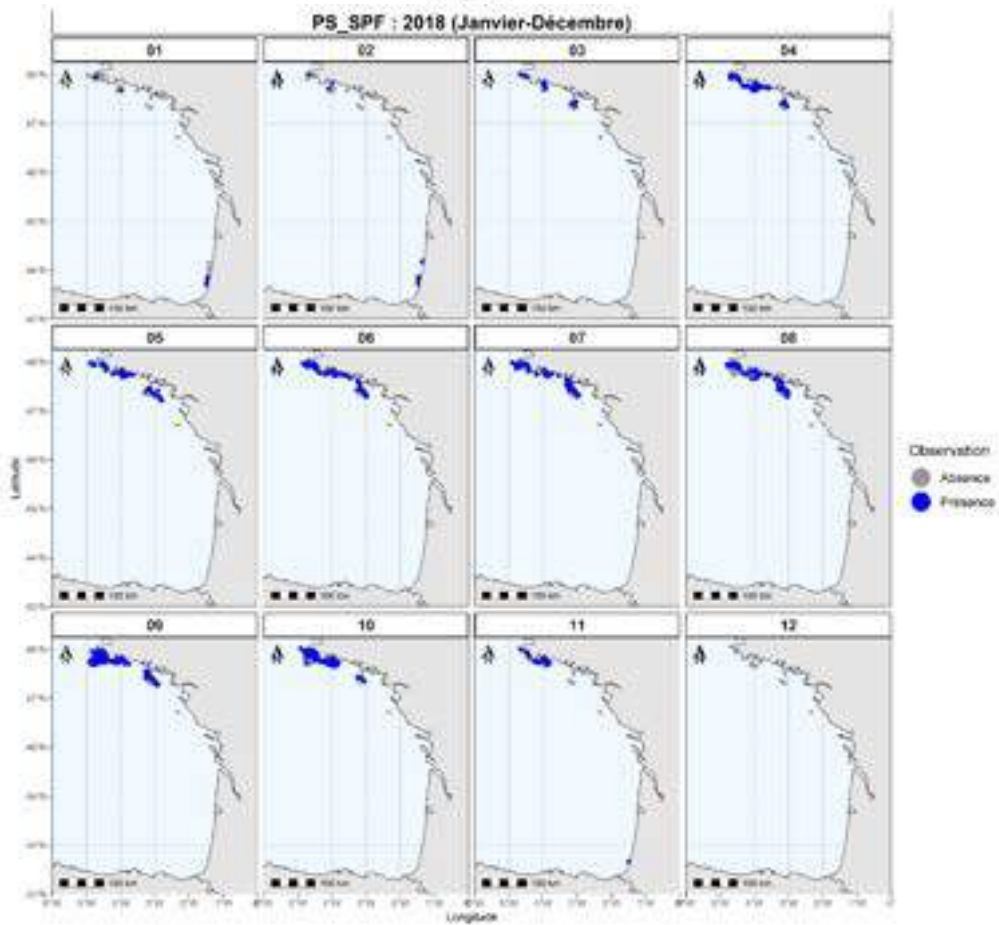
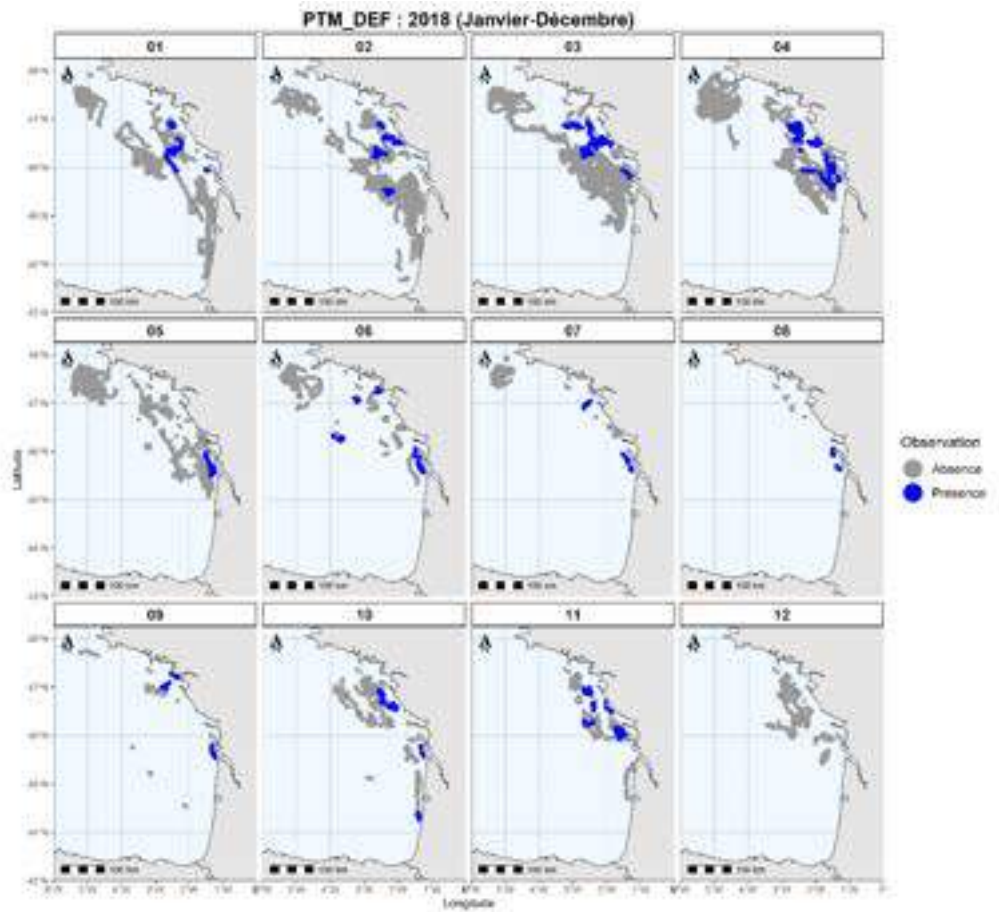
Le système de surveillance des navires par satellite (VMS) fournit à intervalles réguliers (toutes les heures) des données sur la position, la direction et la vitesse des navires aux autorités de pêche. Il est obligatoire pour les navires de pêche professionnelle de plus de 12 mètres, sous pavillon de l'Union européenne, depuis le 1er janvier 2012, et permet un suivi de l'activité de pêche à une résolution spatio-temporelle très fine.

Ces informations spatialisées sont couplées aux données de débarquements (logbooks) suivant la méthodologie décrite par Hintzen *et al.* (2012). Afin de classifier les navires en flottilles homogènes, une analyse exploratoire des captures, des efforts et des CPUE a été réalisée sur la période 2009 – 2018 à un pas de temps mensuel pour tous les métiers ayant pêchés au moins une fois la sardine (Annexes 1 à 4).

A la suite de cette analyse, les données commerciales sont ainsi réunies en groupes de navires utilisant les mêmes engins et ayant un comportement de ciblage relativement homogène (i.e. groupe d'espèce pêchée similaire). Ces groupes de navires sont appelés flottilles dans la suite du rapport. Seules les flottilles pêchant sur le domaine d'étude et présentant une proportion de présence de sardine non négligeable dans les captures (> 1% de présence) sont conservées pour l'analyse (Annexe 4). Finalement, les données commerciales ont été regroupées en 3 flottilles :

- les chaluts à bœufs pélagiques ciblant les petits pélagiques (PTM_SPF) ;
- les chaluts à bœuf pélagiques ciblant les espèces démersales (PTM_DEF) ;
- les senneurs ciblant les petits pélagiques (PS_SPF).

Ces 3 flottilles sont marquées par une forte saisonnalité et présentent une activité essentiellement côtière (Figure 3). La conséquence du regroupement de PTM_SPF et PTM_DEF en une seule flottille a été étudiée au cours du stage.



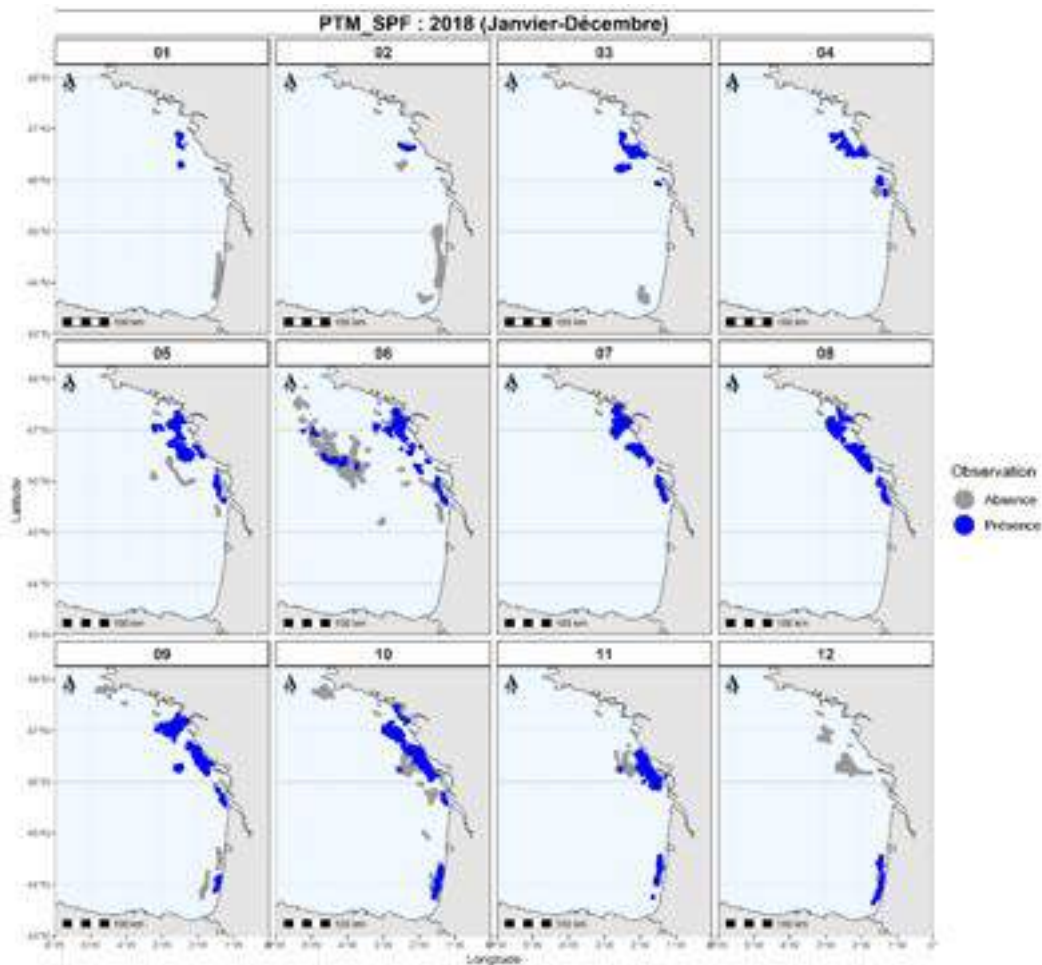


Figure 2 : Observations (présence/absence) de *S. pilchardus* par les 3 flottilles commerciales entre janvier et décembre 2018

2.3. Spécificités du cas d'étude et conséquence pour la modélisation

Dans le cas des espèces pélagiques, les CPUE issues des flottilles commerciales ne peuvent pas être considérées comme des indices d'abondance proportionnels à la biomasse. En effet, les poissons ont des comportements plus ou moins grégaires, durant les phases de fraie ou d'alimentation par exemple, et certaines espèces se déplacent en bancs de tailles importantes et très denses. Ces comportements d'agrégation déterminent en partie l'espace occupé par les individus d'un stock (Pitcher, 1980). L'organisation en bancs est commune chez les poissons, Burgess et Shaw (1979) estiment que 80% des poissons ont un tel comportement au cours de leur cycle de vie, en particulier au stade juvénile. Ce comportement est prévalant chez les espèces pélagiques, comme les scombridés, les carangidés ou encore les clupéidés, et cela à des conséquences majeures sur le fonctionnement des pêcheries associées à leur exploitation. En effet, dans ce cas, la relation entre biomasse et captures-par-unité d'effort n'est plus linéaire. Du fait du comportement d'agrégation de la ressource, les CPUE restent relativement stables quel que soit le niveau de biomasse sous-jacent (ce phénomène est aussi connu sous le nom d'hyperstabilité des CPUE - Pitcher, 1995; Fréon et Misund, 1999).

Dans ce cas, les données de CPUE commerciales ne peuvent pas être utilisées directement pour modéliser la distribution de la biomasse de sardine. Par conséquent, les données commerciales et les données scientifiques sont dégradées en données de présence-absence et le cadre de modélisation développé pour intégrer ces données doit être adapté en conséquence.

2.4. Traitement des données

Les données scientifiques sont filtrées pour conserver uniquement les sardines d'une taille supérieure à 11 cm (possible uniquement avec les données PELGAS). Elles sont ensuite groupées par 'Espèce x ESDU x Date' (à chaque ESDU correspond une unique position GPS). Si une biomasse non nulle est mesurée, on indique une présence (pres = 1), dans le cas contraire on indique une absence (pres = 0). Quand aucune information n'est spécifiée pour la sardine, mais que des biomasses sont mesurées pour d'autres espèces, on ajoute une absence pour cette ESDU. Enfin, toutes les observations sont considérées comme étant réalisées sur le mois de calibration de la campagne (i.e. mai pour PELGAS et septembre pour JUVENA)

Les données commerciales de captures sont traitées en suivant la méthode développée par Hintzen et al (2012). D'une part, on filtre les ping VMS pour conserver uniquement pings émis lors de l'activité de pêche (vitesse moyenne supérieur à 4.5 nœuds – algorithme AlgoPesca développé par l'IFREMER). Les déclarations de pêche (logbooks) sont déclarées à l'échelle d'une séquence de pêche (i.e. la combinaison 'carré statistique x engin utilisé x marée x jour'). Les quantités déclarées sont réallouées de façon uniforme sur les pings VMS de la séquence de pêche correspondante. Les présence/absence sont calculées de la même manière que précédemment (biomasse non nulle/nulle).

2.5. Modèle intégré hiérarchique

Le modèle hiérarchique est structuré en 3 couches de variables reliées entre elles par des relations de dépendances probabilistes : (1) la couche des observations (scientifiques et commerciales) ; les observations sont conditionnelles au (2) champ latent (non observé) ; du fait de la nature des observations (présence-absence), le champ latent est un champ de présence-absence ; (3) les hyperparamètres contrôlent la structure du champ latent.

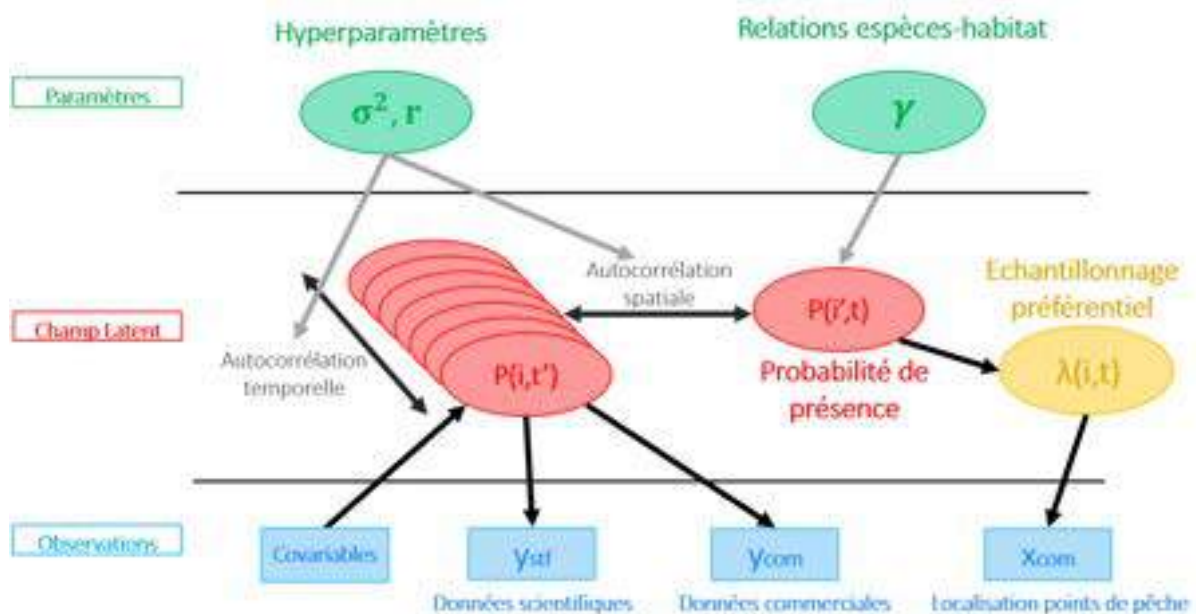


Figure 3 : Diagramme du modèle spatio-temporel hiérarchique intégré (d'après Alglave et al ; under review)

Dans ce modèle, les équations d'observation permettent de relier les différences sources de données (données de présence absence commerciales et scientifique) au champ latent de

présence absence. Le modèle permet de représenter explicitement le fait que la position spatiale des points d'échantillonnage commerciaux dépend du champ latent de présence/absence (Figure 4). En effet, les pêcheurs tendent à cibler des zones où la ressource est présente (i.e. l'échantillonnage est dit préférentiel – Diggle *et al.*, 2010) ce qui peut introduire un biais dans les prédictions si l'échantillonnage de la donnée commerciale et le lien avec la variable à modéliser n'est pas pris en compte dans la méthode d'inférence.

2.5.1. Champ latent de présence-absence

Soit une variable aléatoire $Z(i,t)$ décrivant la présence de sardine au site i au pas de temps t , elle prend la valeur 1 ou 0 selon qu'une sardine est présente ($Z(i,t) = 1$) ou pas ($Z(i,t) = 0$). La variable aléatoire $W_{i,t}$ suit donc une loi de Bernoulli de paramètre $P_{i,t}$ (Eq. 1).

$$Z(i,t) \sim \text{Bernoulli}(P(i,t)) \quad (\text{Eq. 1})$$

$P(i,t)$ représente la probabilité de présence de sardines au site i au temps t .

Cette probabilité de présence est modélisée par un champ aléatoire gaussien (dans l'échelle du logit) sur un domaine discret. La valeur du champ latent au point i est exprimée comme une combinaison linéaire qui dépend de :

- Un intercept μ , représentant la valeur moyenne du champ de présence/absence de sardine.
- Des co-variables environnementales C et un paramètre γ qui capture la relation espèce-habitat (effets fixes) ; Les k covariables $C_k(i,t)$ avec la fonction f_k modélisant la relation espèce-habitat (possiblement non-linéaire). Dans le cas présent, la bathymétrie, la chlorophylle a et la température de surface (SST) ont été inclus dans le modèle (Zwolinski *et al.*, 2010 ; Doray, *com. pers*). Les modèles ont été ajustés avec des données satellitaires de Copernic ainsi qu'avec les sorties du modèle physico-biogéochimiques POLCOM-ERSEM (SST et chlorophylle a) dont la résolution était plus fine. Les cartes de ces covariables au mois de mai 2018 sont disponibles en Annexes 5 à 7. Les covariables ont été normalisées pour l'ajustement afin de faciliter la convergence du modèle.
- Un effet fixe Mois $V(t)$
- Un effet aléatoire spatio-temporel $U(i,t)$ qui capture les autocorrélations spatiale et temporelle (Eq. 2). L'effet spatio-temporel est un champ aléatoire gaussien de moyenne zéro et dont la matrice de covariance suit une fonction de corrélation de Matérn caractérisée par un paramètre de portée r – la distance à laquelle la corrélation entre deux points est égal à 0.1 – et un paramètre de variance σ^2 (Eq. 3). La corrélation temporelle est modélisée à l'aide d'un processus autorégressif d'ordre 1 (Blangiardo *et al.*, 2013, Krainski *et al.*, 2019).

$$\text{Logit}(P(i,t)) = \mu + V(t) + \sum_k f_k(C_k(i,t)) + U(i,t) \quad (\text{Eq. 2})$$

$$U(i,t) \sim \text{GMRF}(0, \Sigma(\sigma^2, r)) \quad (\text{Eq. 3})$$

Afin d'assurer l'identifiabilité du modèle, les paramètres associés au mois de mai ($V(t_{mai})$ - mois de PELGAS) et à la campagne PELGAS ($\alpha(j_{pelgas})$) sont fixés à 0 et leur effet est capturé par l'intercept μ .

2.5.2. Modèle d'observation de la présence/absence

Soit la variable aléatoire $Y(i, j, t)$ représentant l'observation d'un individu au site i au temps t par la flottille (commerciale ou scientifique) j . Le processus d'observation $Y(i, j, t)$ est ainsi conditionnel au processus écologique $Z(i, t)$. Une sardine ne peut être observée au site i et au temps t que si elle est présente ($Z(i, t) = 1$), mais la probabilité qu'elle soit détectée quand elle est présente n'est pas forcément 1 mais dépend de la détectabilité de la flottille j qu'on modélise à l'aide d'une probabilité de détection noté $\delta(i, j, t)$ et dépendant éventuellement de co-variables C' (Eq. 4).

$$P(Y(i, j, t) = y_{i,j,t}) = \begin{cases} P'(i, j, t) & : \text{Si } y_{i,j,t} = 1 \text{ (présence)} \\ 1 - P'(i, j, t) & : \text{Si } y_{i,j,t} = 0 \text{ (absence)} \end{cases} \quad (\text{Eq. 4})$$

$$P'(i, j, t) = Z(i, t) * \delta(i, j, t) \quad (\text{Eq. 5})$$

$$\text{Logit}(\delta(i, j, t)) = C_k'(i, j) * \rho \quad (\text{Eq. 6})$$

2.5.3. Modéliser l'échantillonnage préférentiel

L'échantillonnage des données commerciales ne suit pas de protocole standardisé et, en particulier, les points de pêches peuvent être préférentiellement répartis dans les zones où la probabilité de présence des espèces ciblées est plus forte. Ce processus est appelé 'échantillonnage préférentiel' (Diggle *et al.*, 2010), il peut biaiser les sorties des méthodes d'inférence et conduire à la surestimation des prédictions dans les zones sous-échantillonnées. Pour intégrer l'échantillonnage préférentiel à l'inférence, nous modélisons de façon explicite la distribution des points de pêche dans le modèle au travers d'un processus ponctuel poissonnien. Cette modélisation fait intervenir un second terme dans la vraisemblance (en plus du terme de vraisemblance associé à l'observation de la présence/absence) qui intègre la position GPS des points de pêche au travers de ce processus ponctuel poissonnien.

Notons $x_{com j}$ les points où les bateaux de la flottille j sont identifiés en pêche (Eq. 9). $x_{com j}$ suit un processus ponctuel non homogène de Poisson décrit par une intensité $\lambda(i)$ variant dans l'espace. $\lambda(i)$ correspond à l'intensité de pêche au point i . Cette intensité est modélisée comme une combinaison log-linéaire de l'effet aléatoire U , multiplié par un paramètre b_j , et un effet spatial résiduel W de même paramétrisation que U capturant l'effet de covariables non pris en compte par le terme associé à l'échantillonnage préférentiel (Eq. 10). Le paramètre b_j quantifie la force de l'échantillonnage préférentiel. Ainsi, une valeur de $b_j = 0$ indique que l'échantillonnage de la donnée commerciale est indépendant du champ latent et $b_j > 0$ indique que l'échantillonnage est préférentiel.

$$x_{com j} \sim IPP(\lambda(i)) \quad (\text{Eq. 9})$$

$$\text{Log}(\lambda(i)) = \mu + b_j * U(i, j) + W(i, j) \quad (\text{Eq. 10})$$

L'échantillonnage préférentiel est analysé sur un seul pas de temps (en mai) afin d'identifier son effet sur les prédictions du modèle.

2.5.4. Simplification induite par les contraintes d'INLA

L'inférence a été réalisée grâce au logiciel INLA (voir ci-dessous). Les contraintes de codage liées à INLA rendent impossible la séparation explicite des équations de processus (Eq. 1-3) et des équations d'observation (Eq. 4-6). Ainsi, la variable latente intermédiaire de présence absence Z ne peut pas être modélisée explicitement. Le modèle final intègre les facteurs écologiques responsables de la variabilité du champ latent de présence/absence et les facteurs de la variation de la détectabilité en une seule équation :

$$P(Y(i, j, t) = y_{i,j,t}) = \begin{cases} P(i, j, t) & : \text{Si } y_{i,j,t} = 1 \text{ (présence)} \\ 1 - P(i, j, t) & : \text{Si } y_{i,j,t} = 0 \text{ (absence)} \end{cases} \quad (\text{Eq. 7})$$

$$\text{logit}(P(i, j, t)) = \mu + \alpha(j) + V(t) + \sum_k f_k(C_k(i, t)) + U(i, t) \quad (\text{Eq. 8})$$

où $\alpha(j)$ représente directement l'effet source de données capturant les différences de détectabilité entre les différentes sources de données ou les différentes flottilles

2.5.5. Outil d'inférence : R-INLA

L'inférence a été réalisé dans un cadre bayésien grâce à l'outil R-INLA et au package Inlabru. R-INLA (Integrated Nested Laplace Approximation) est un outil d'inférence bayésien introduit par Rue *et al.* (2009). Il offre une alternative aux méthodes d'estimation des distributions a posteriori par échantillonnage de type MCMC pour l'estimation de champs gaussiens avec pour principal avantage d'être notablement plus rapide.

R-INLA combine 3 approximations distinctes faisant appel à l'approche SPDE (Stochastic Partial Differential Equations) (Approximation 1 et 2 - Lindgren *et al.*, 2011) et à l'approximation de type INLA (Approximation 3 - Rue *et al.*, 2009) :

- (1) Il approxime un champ gaussien par un champ Gauss-Markov (lorsque le champ gaussien admet comme fonction de corrélation une fonction de Matérn). Cette approximation permet d'obtenir une représentation creuse de la matrice de précision Q (propriétés markoviennes) et de simplifier les calculs matriciels liés à l'estimation de l'effet aléatoire spatial (Lindgren *et al.*, 2011).
- (2) Il diminue la résolution spatiale à laquelle est estimée la structure spatiale du champ latent en le modélisant à l'échelle d'une maille triangulaire creuse (Figure 5). Le lien entre les points de données (définis sur un domaine continu) et l'effet aléatoire (définie aux sommets de la maille) est réalisé via une interpolation linéaire des valeurs de l'effet aléatoire sur les points de données (Krainski *et al.*, 2019). Cette approximation permet d'estimer l'effet aléatoire sur un nombre de points plus faible (les nœuds de la maille) que si l'effet aléatoire avait dû être estimé en chaque point de données (en particulier lorsque les points de données sont nombreux). Différentes mailles ont été testées afin d'évaluer la sensibilité des sorties du modèle à la résolution de la maille (Annexe 8).

- (3) Il simplifie l'estimation des distributions a posteriori conjointe du champ latent U et des hyperparamètres θ notée $\pi(U, \theta | y)$ en mobilisant des méthodes d'approximations de Laplace. Au lieu d'utiliser des méthodes d'échantillonnage de type MCMC souvent très longues notamment quand le nombre de dimensions augmente, R-INLA approxime les distributions marginales à posteriori du champ latent $\pi(U_i | y) = \int \pi(U_i | \theta, y) \pi(\theta | y) d\theta$ et des hyperparamètres $\pi(\theta_j | y) = \int \pi(\theta | y) d\theta_{-j}$ en simplifiant $\pi(\theta | y)$ et $\pi(U_i | \theta, y)$ via des approximations de Laplace emboîtées.

Dans le modèle spatio-temporel, la domaine spatial est défini par une SPDE et un modèle autorégressif d'ordre 1 (i.e. AR(1)) pour modéliser l'auto-corrélation temporelle (Blangiardo *et al.*, 2013, Krainski *et al.*, 2019). La relation espèce habitat, possiblement non-linéaire, est modélisée par une SPDE à une dimension.

Les sorties n'étant pas sensible aux choix des priors, ils sont fixés par défaut pour la suite de l'analyse (Annexe 8). Par ailleurs, différentes mailles ont été testées afin de sélectionner celles assurant le meilleur compromis qualité d'inférence/temps de calcul (Annexe 9).

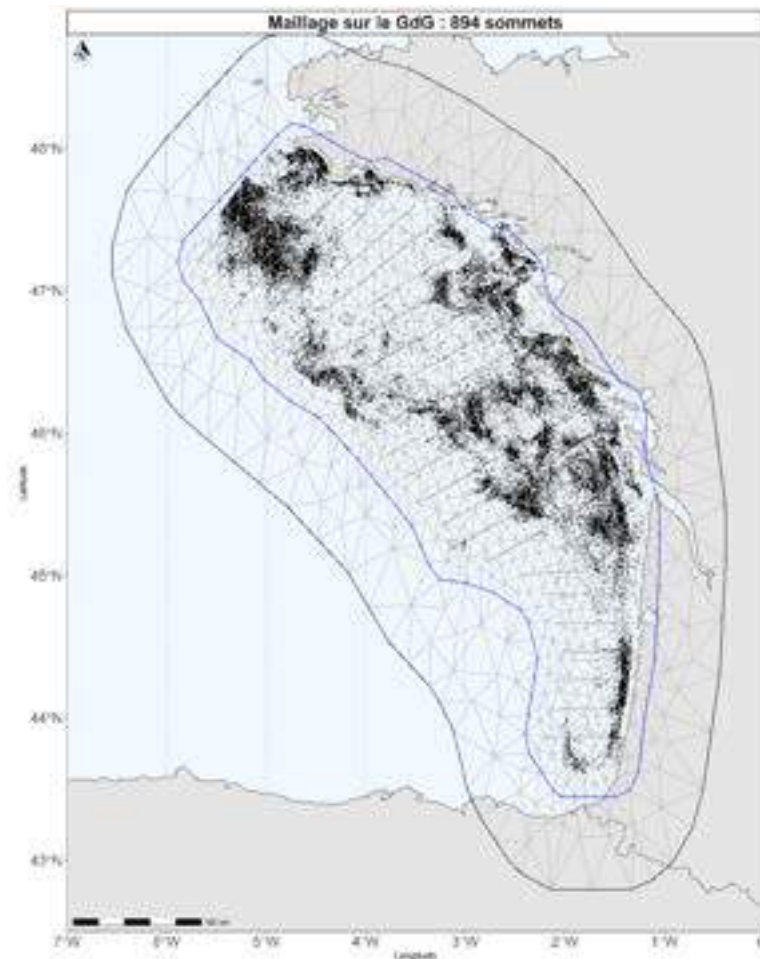


Figure 5 : Maille d'interpolation triangulaire et points de données dans le GdG (mai 2018).

2.5.6. Métriques de validation

La performance prédictive des modèles est évaluée pour chaque source de donnée j par analyse de la courbe ROC, l'AUC associée et la CPO.

Une courbe ROC (receiver operating curve characteristic) est un graphique représentant les performances d'un modèle de classification. Cette courbe trace le taux de vrais positifs (les présences observées prédites correctement) en fonction du taux de faux positifs (les absences observées faussement prédites par des présences). A partir de cette courbe on peut calculer l'AUC (« Area Under Curve » ou "aire sous la courbe ROC") qui correspond à l'intégrale sous la courbe ROC. L'AUC prend des valeurs entre 0 et 1. Un modèle qui prédit systématiquement l'inverse de ce qui est observé à une AUC de 0. Un modèle dont toutes les prédictions sont justes à une AUC de 1. Un modèle qui classe les observations au hasard aura une AUC de 0.5.

La CPO (Conditional predictive ordinate) correspond à un test de validation croisée basé sur la densité *a posteriori* de la donnée y_i quand le modèle est ajusté sur toute les données à l'exception de y_i (Gómez-Rubio, 2020).

$$CPO_i = \pi(y_i | y_{-i}) \quad (\text{Eq. 10})$$

Cette densité prend une valeur comprise entre 0 et 1 selon qu'elle est mal ($CPO_i = 0$) ou bien prédite ($CPO_i = 1$).

Nous résumons cette métrique en calculant la Log-CPO pour chaque source de donnée (i.e. en ne considérant que les n_j observations issues d'une des sources de données). Plus la valeur de la LCPO est faible, plus la capacité prédictive du modèle est élevée pour cette source de donnée.

$$LCPO_j = - \sum^{n_j} \log(CPO_i) \quad (\text{Eq. 11})$$

2.5.7. Démarche de valorisation du modèle

Dans un premier temps, l'approche de modélisation intégrée a été appliquée à un seul pas de temps de la campagne PELGAS (mai 2018) afin de comparer différentes configurations de modèles et d'évaluer (1) l'apport des différentes sources de données dans l'inférence, (2) l'impact de l'échantillonnage préférentiel sur l'inférence et (3) l'apport des co-variables environnementales dans le modèle.

- (1) La contribution des différentes sources de données à l'inférence est étudiée en comparant les sorties du modèle intégrant toutes les sources de données aux modèles ajustés soit à la donnée scientifique seule, soit à la donnée commerciale seule. Pour cette partie, les trois flottilles commerciales sont intégrées dans un modèle « commercial ».
- (2) Pour évaluer l'effet de l'échantillonnage préférentiel sur l'inférence, nous comparons un modèle prenant en compte l'échantillonnage préférentiel avec un modèle ne prenant pas en compte l'échantillonnage préférentiel. Dans ces configurations, toutes les sources de données sont intégrées.

- (3) L'ajout de covariables au modèle est réalisé en deux étapes. Dans un premier temps, nous ajustons le modèle avec les covariables sans prendre en compte l'effet aléatoire spatial. Dans un second temps, nous ajustons un modèle prenant en compte les covariables et l'effet aléatoire spatial. Cette approche en 2 étapes permet d'appréhender les problèmes d'identifiabilité de la relation espèce-habitat lorsque le modèle est potentiellement sur-paramétré.

La comparaison des différentes configurations de modèle sur un seul pas de temps a permis de choisir la configuration qui réalise le meilleur compromis entre qualité de l'inférence, temps de calcul et complexité du modèle pour l'ajuster sur l'ensemble des pas de temps des années 2009 à 2018.

3. Résultats

3.1. Analyse à l'échelle d'un mois : mai 2018

Pour chaque sous partie nous présentons deux types de cartes - la probabilité de présence prédite par le modèle et les écarts-types associés – ainsi que les performances prédictives de chaque modèle. Une comparaison de ces performances est réalisée dans une dernière partie. Un troisième type de carte, étudiant les différences absolues des probabilités de présence inférées, compare les configurations testées (ajout de l'échantillonnage préférentiel et des co-variables) avec le modèle intégré de référence. En annexe figurent : les courbes ROC, les cartes de capacités prédictives des CPO, les cartes d'inférences de 2009 à 2018 et les scores AUC associés et moyens sur la période (Annexes 14 à 36).

3.1.1. Apport des différentes sources de données à l'inférence

Le modèle intégré produit des cartes de présence-absence en cohérence avec les différentes sources de données (Figure 6). Dans les zones où les observations sont positives, les probabilités de présence sont élevées. Dans les zones où les observations sont nulles, les probabilités de présence sont faibles. La bonne concordance entre les observations de présence-absence et les prédictions du modèle apporte une première validation qualitative de la performance du modèle.

Les cartes de prédictions ajustées aux différentes sources de données montrent que les données scientifiques apportent le plus d'information à l'inférence (Figure 7). En effet, la carte de distribution obtenue à partir du modèle intégré est presque superposable à celle obtenue à partir des données scientifiques seules. Comme les données scientifiques couvrent l'ensemble du domaine d'étude et bénéficient d'un échantillonnage dense, elles sont suffisantes pour décrire les patrons de distribution de la sardine sur l'ensemble du GdG. Les écarts-types sont globalement faibles sur l'ensemble de l'aire d'étude avec des valeurs légèrement plus élevées entre les transects du plan d'échantillonnage ainsi que dans les zones de transitions de présence/absence.

Les données commerciales sont beaucoup plus concentrées dans l'espace et essentiellement restreintes à une bande très côtière. Lorsqu'elles sont la seule source de donnée utilisée, elles ne permettent pas de capturer les bancs de sardine au large et les prédictions en dehors du rayon d'action des flottilles commerciales ne sont pas fiables. Les écarts types du modèle ajusté à la donnée commerciale seule sont très élevés au large et faibles dans le rayon d'action des flottilles.

Dans le modèle intégré, les données commerciales apportent de l'information en complément des données scientifiques. Elles apportent de l'information sur les zones côtières, mais ne modifient que légèrement les patrons de distributions obtenus avec la donnée scientifique seule, en précisant certaines tâches de présences côtières mal décelées par les données scientifiques (e.g. au Nord de la carte).

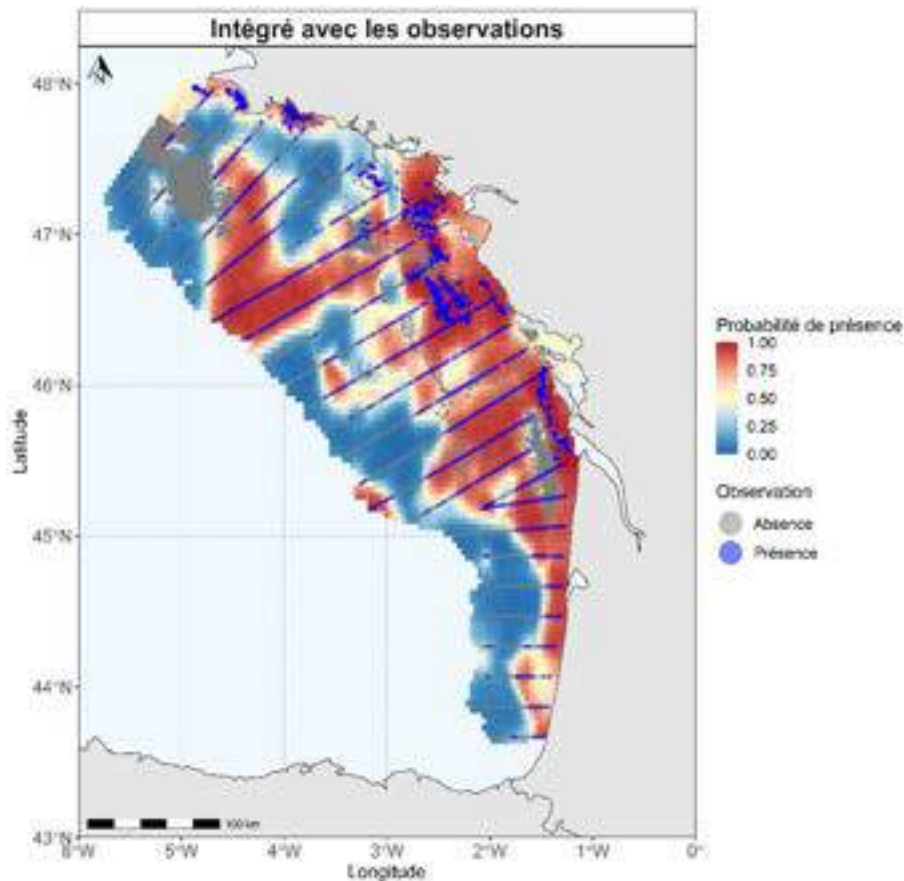


Figure 6 : Cartographies des probabilités de présence et des observations par les flottilles commerciales et la campagne scientifique PELGAS en Mai 2018.

Le modèle ajusté sur une seule source de donnée a systématiquement d'excellentes capacités prédictives avec des AUC toujours supérieures à 0.90 (Figure 11). De même, les modèles intégrés ont aussi d'excellentes performances prédictives, à l'exception des prédictions des données PTM_SPF pour lesquelles les scores sont autour de 0.8, ce qui demeure de très bons scores.

Concernant les LCPO, il apparaît ici que l'intégration des données commerciales dans le modèle en plus des données scientifiques améliore les prédictions de données commerciales au détriment des observations scientifiques (plus la LCPO est faible, meilleur est la capacité prédictive du modèle) (Figure 11).

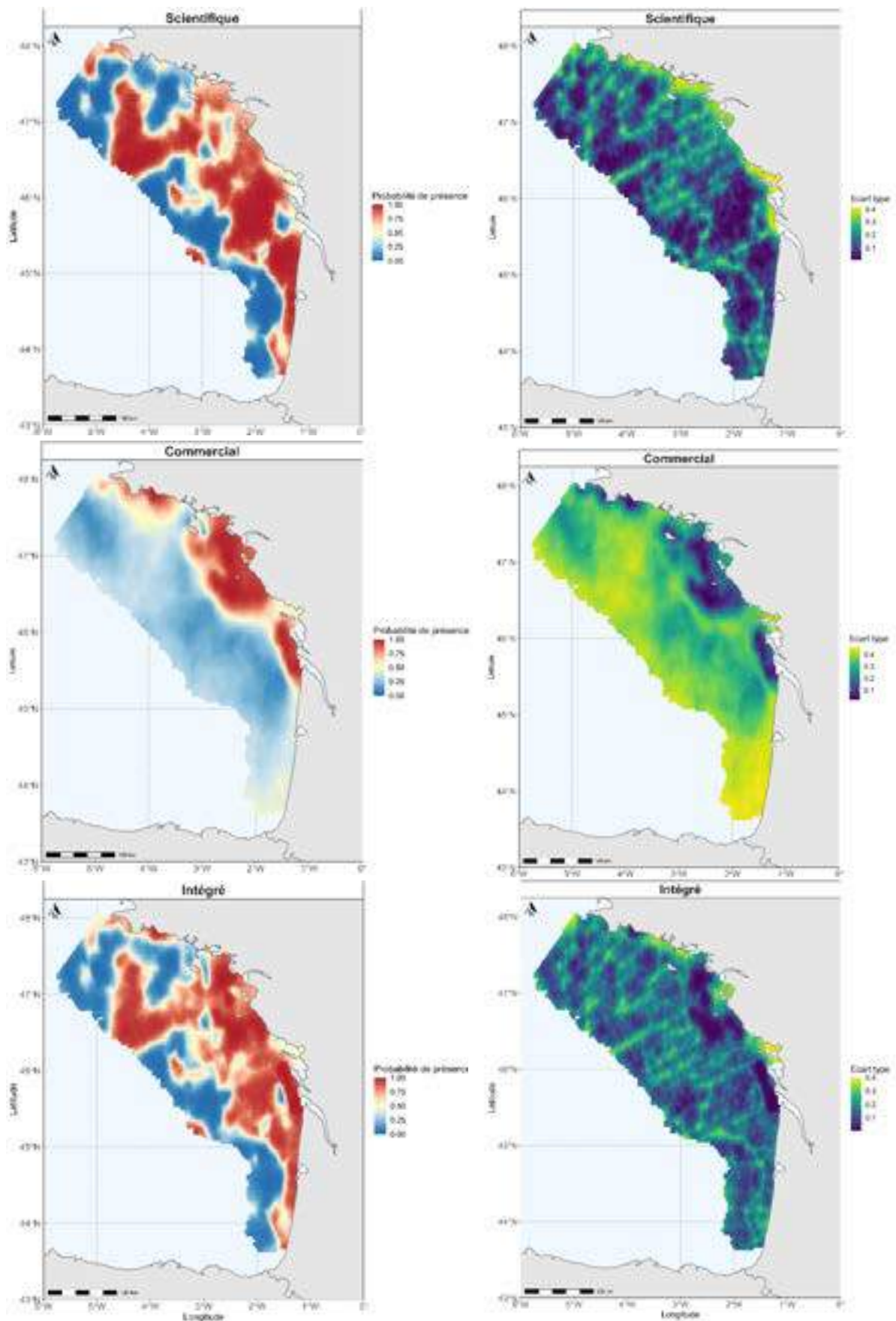


Figure 7 : Cartographies des probabilités de présence (colonne de gauche) et écarts-types associés (colonne de droite) au mois de Mai 2018 obtenues à partir des données scientifiques, des données commerciales et dans un modèle intégré les combinant.

3.1.2. Intensité de l'échantillonnage préférentiel

L'échantillonnage préférentiel semble avoir peu d'influence sur les inférences car les cartes d'inférences obtenues avec ou sans E.P. sont similaires. Cependant, le modèle avec E.P. rend mieux compte du fonctionnement des pêcheries.

Les résultats suggèrent des intensités de ciblage différentes en fonction des trois flottilles commerciales, avec un ciblage plus marqué pour les senneurs. Ainsi logiquement, la prise en compte d'un échantillonnage préférentiel modifie légèrement les inférences dans les zones côtières où sont présents les senneurs (au Sud de la Bretagne, de la Pointe du Raz à Belle-Île-en-Mer) (Figure 8). En effet, le paramètre b_j estimé (qui contrôle l'intensité de l'échantillonnage préférentiel) prend une valeur de 1.48 pour les senneurs (PS_SPF), tandis qu'il est estimé à 0.37 et -0.48 respectivement pour les flottilles PTM_SPF et PTM_DEF. Cela suggère l'existence d'un échantillonnage préférentiel uniquement pour les flottilles

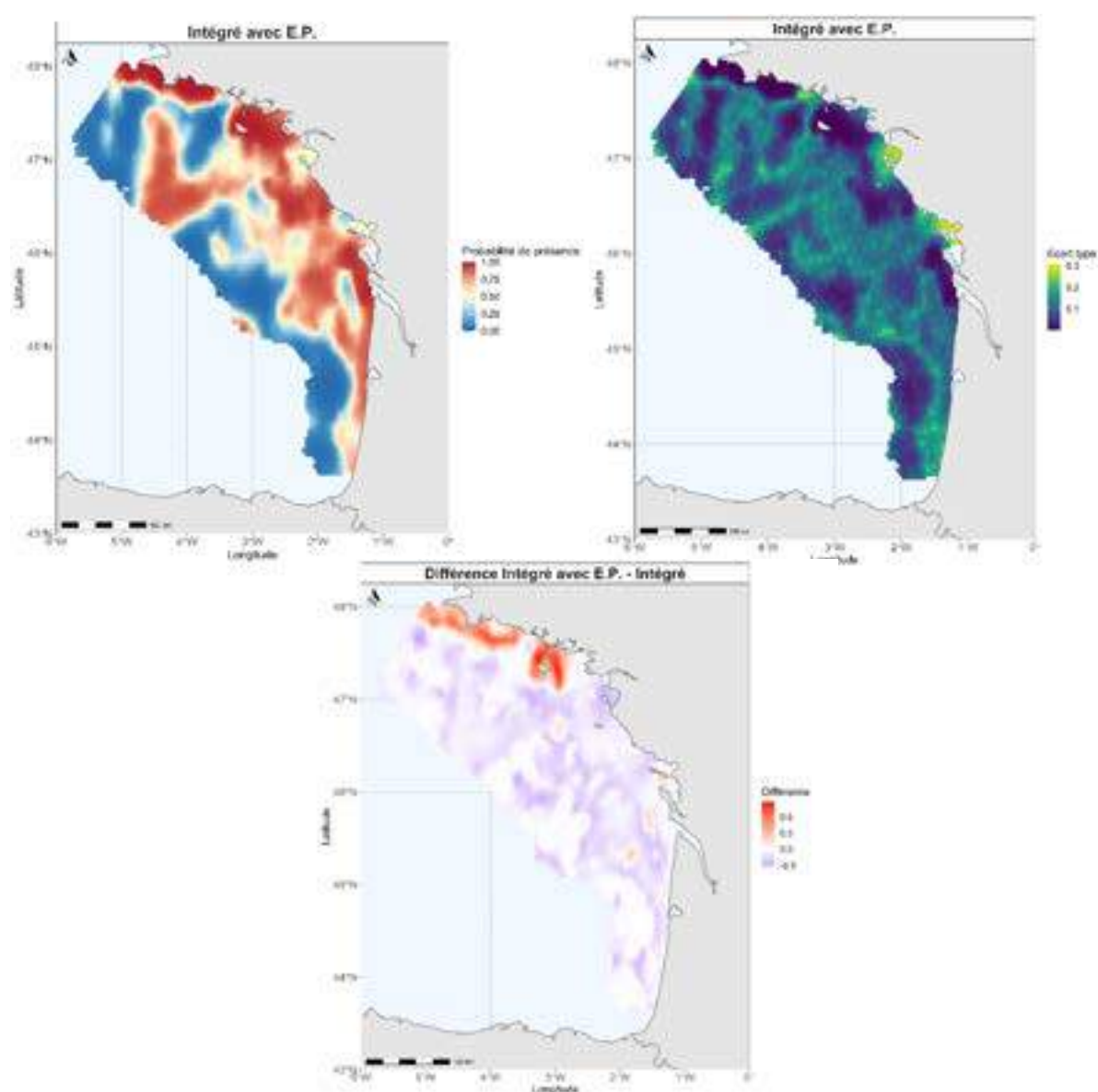


Figure 8 : Cartes d'inférence du modèle avec prise en compte d'un échantillonnage préférentiel dans le fonctionnement des flottilles commerciales, les écarts types associés et de la différence avec le modèle intégré ne prenant pas en compte d'E.P. (E.P. – Intégré)

commerciales ciblant explicitement les petits pélagiques (senneurs et PTM_SPF), et un échantillonnage plus fort pour les senneurs en comparaison aux chaluts en bœufs pélagiques. Ces résultats sont consistants avec le fonctionnement des pêcheries étudiées.

La prise en compte de l'E.P. permet de réduire les écarts-types dans le rayon d'action des flottilles. En effet, dans le cas où l'échantillonnage préférentiel est pris en compte dans l'inférence, la distribution des points de pêche contribue à l'inférence et ils apportent de l'information dans les zones échantillonnées, ce qui diminue l'écart-types dans ces zones.

De plus, la prise en compte de l'E.P. réduit les probabilités de présence sur l'ensemble du GdG, à l'exception des zones d'activités des senneurs, notamment dans les zones de transition entre les fortes et les faibles probabilités de présence (Fig. 8, carte du bas). Cette diminution est due à l'absence de points de pêche des bolincheurs qui concentrent leur activité en Bretagne Sud. Le modèle considère ainsi que les probabilités de présence sont plus faibles sur les autres secteurs.

Notons que l'analyse avec seulement deux flottilles (les flottilles PTM_DEF et PTM_SPF étaient regroupés en une seule flottille PTM) ne laissait apercevoir aucune modifications dans les inférences (les résultats ne figurent pas dans ce manuscrit). Les \mathbf{b}_j estimés prenaient des valeurs proches de zéro et la force de l'échantillonnage préférentiel estimé était relativement faible. La distinction de PTM en 2 segments plus homogènes permet de mieux estimer l'intensité du ciblage. Pour autant les patrons de distributions de la sardine sont globalement similaires en fusionnant ou non les flottilles PTM_DEF et PTM_SPF.

En réduisant le domaine à la frange côtière (50 km de distance du littoral) les estimations des paramètres de ciblage sont modifiées ($\mathbf{b}_{PS_SPF} = 1.05$; $\mathbf{b}_{PTM_SPF} = 0.49$; $\mathbf{b}_{PTM_DEF} = -0.17$). Cela suggère que d'autres facteurs, notamment la distance à la côte, entrent en compte dans la répartition des points de pêches \mathbf{x}_{com} . En particulier on observe que \mathbf{b}_{PS_SPF} diminue tandis que les autres valeurs augmentent : tous les points de pêche des senneurs étant côtiers, la réduction de la zone diminue l'influence de cette répartition sur la force du ciblage. Au contraire, les autres flottilles ont des activités qui s'éloignent plus vers le large et pour lesquelles les observations indiquent des absences plus nombreuses. En restreignant la zone d'étude on augmente l'importance de l'échantillonnage préférentiel.

Cependant, malgré l'intérêt de la prise en compte de l'E.P. pour une meilleure représentation de la distribution spatiale des captures commerciales, cette dernière entraîne une diminution de l'AUC et une augmentation des LCPO en comparaison au modèle intégré et donc à de moins bonnes capacités prédictives des observations de présence/absence (Figure 11). Ainsi, en considérant ces résultats et les temps de calculs nécessaires à l'ajustement des modèles prenant en compte l'échantillonnage préférentiel, nous choisissons de ne pas le mettre en place dans le modèle spatio-temporel.

3.1.3. Ajout des co-variables environnementales.

L'ajout de covariables environnementales dans le modèle a peu d'effet sur les prédictions du modèle et les résultats obtenus sont peu sensibles à la source des données considérée, qu'elles soient satellitaires ou issues du modèle physico-biogéochimique (Figure 9). Les relations espèces-habitat permettent de mettre en évidence un effet bathymétrie pour les faibles profondeurs et des effets SST et chlorophylle A relativement faibles. Les cartes d'inférences sont présentées en Annexe 11. Les relations espèces-habitat sont similaires pour toutes les covariables même en retirant l'effet aléatoire (Annexe 12). Les cartographies obtenues par le modèle intégré avec ou sans effet des covariables sont similaires (Figure 10).

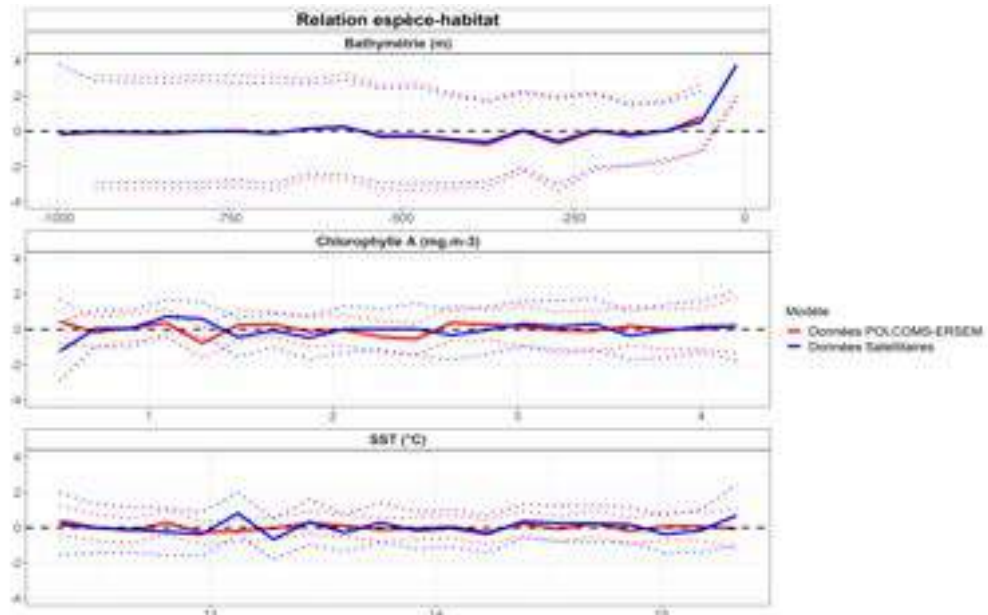


Figure 9 : Effet des co-variables environnementales (Bathymétrie, SST et Chlorophylle A) dans un modèle prenant en compte des données satellitaires ou issues des sorties du modèle physico-biogéochimique POLCOM-ERSEM.

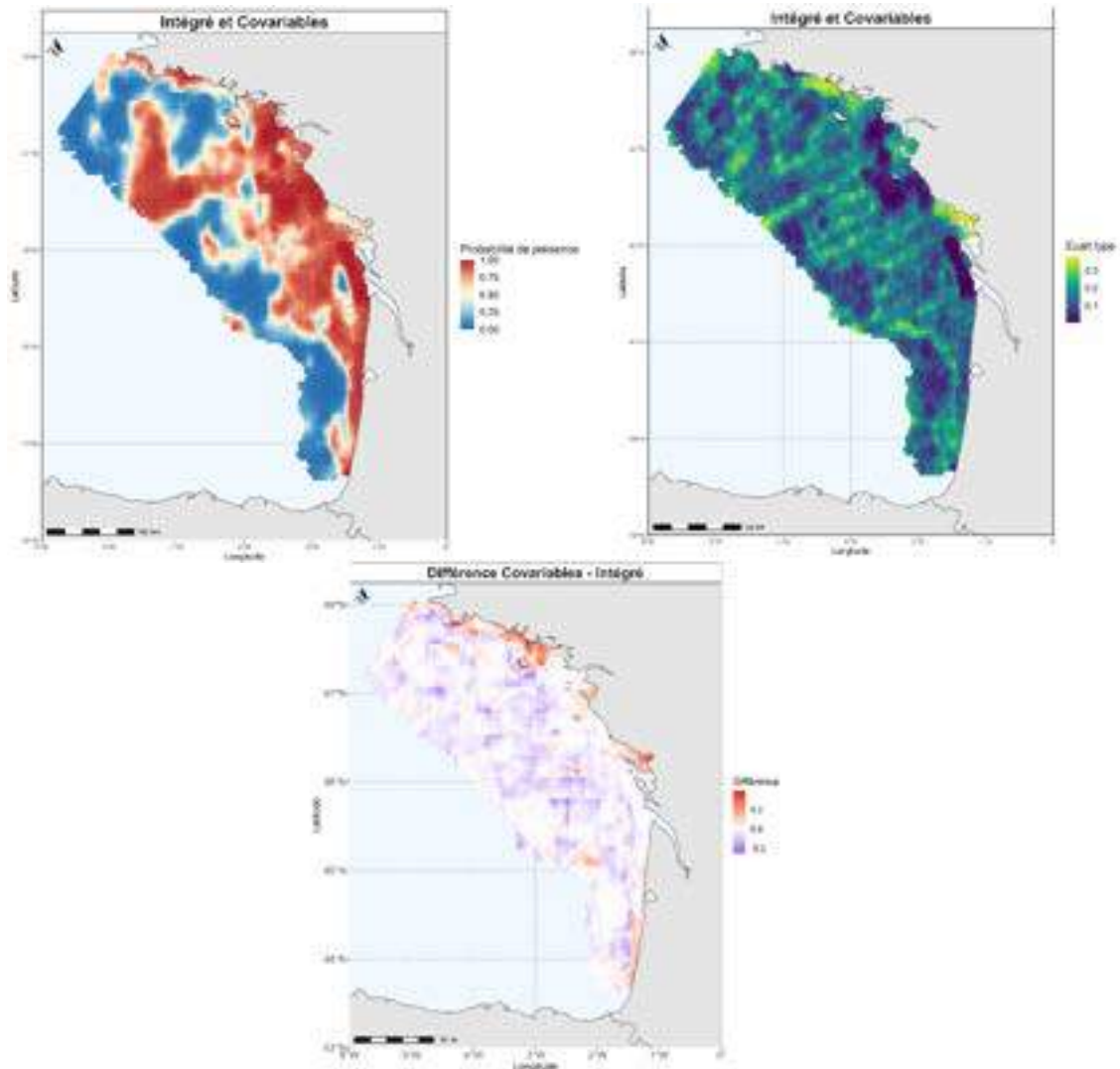


Figure 10 : Cartes d'inférences en prenant en compte les co-variables environnementales (Bathymétrie, SST et Chlorophylle A issue des données satellitaires), les écarts types associés et de la différence avec le modèle intégré 'simple'. (Covariables – Intégré).

La principale différence provient d'une frange littorale de forte probabilité, issue de la prise en compte de la bathymétrie. Finalement, l'ajout de covariables environnementales dans le modèle n'améliore pas les capacités prédictives du modèle (Figure 11).

Les difficultés d'identifiabilité du modèle, le faible apport des covariables dans l'inférence et le risque d'une confusion d'effet avec les facteurs affectant la distribution des points pêche nous ont conduit à laisser de côté les covariables lors de l'ajustement du modèle spatio-temporel. L'effet des covariables sur les prédictions du modèle spatio-temporel sont étudiées à posteriori de l'ajustement du modèle spatio-temporel. Les résultats sont présentés en annexe (13).

3.1.4. Comparaison des capacités prédictives selon les différents modèles.

Les modèles prenant en compte un seul jeu de données présentent le plus souvent les meilleures capacités prédictives (Figure 11). Ce résultat est attendu. En effet, le modèle intégré fait une synthèse de l'ensemble des données, ce qui explique le fait qu'il soit moins bien ajusté à chaque jeu de données pris en compte séparément.

La comparaison des AUC entre les différents modèles laisse apparaître qu'elles ont tendance à diminuer avec une augmentation de la complexité du modèle selon l'ordre suivant :

$$AUC(\text{Donnée seule}) > AUC(\text{Intégré}) > AUC(\text{Covariables}) > AUC(\text{E.P.})$$

Néanmoins, dans tous les cas l'AUC reste supérieure à 0.75, ce qui correspond à de bonnes capacités prédictives. Notons un changement dans cet ordre pour la prédiction des données issues des chaluts en bœufs pélagiques (PTM_SPF et PTM_DEF) lorsque les covariables environnementales sont prises en compte.

L'analyse des LCPO est plus nuancée : pour les données scientifiques et PS_SPF, la qualité prédictive diminue avec la complexité du modèle, de la même manière que pour les AUC. En revanche pour les flottilles PTM_SPF et PTM_DEF, la performance prédictive est meilleure avec un modèle intégré et les différences sont très faibles en prenant en compte un échantillonnage préférentiel ou des covariables (Figure 11).

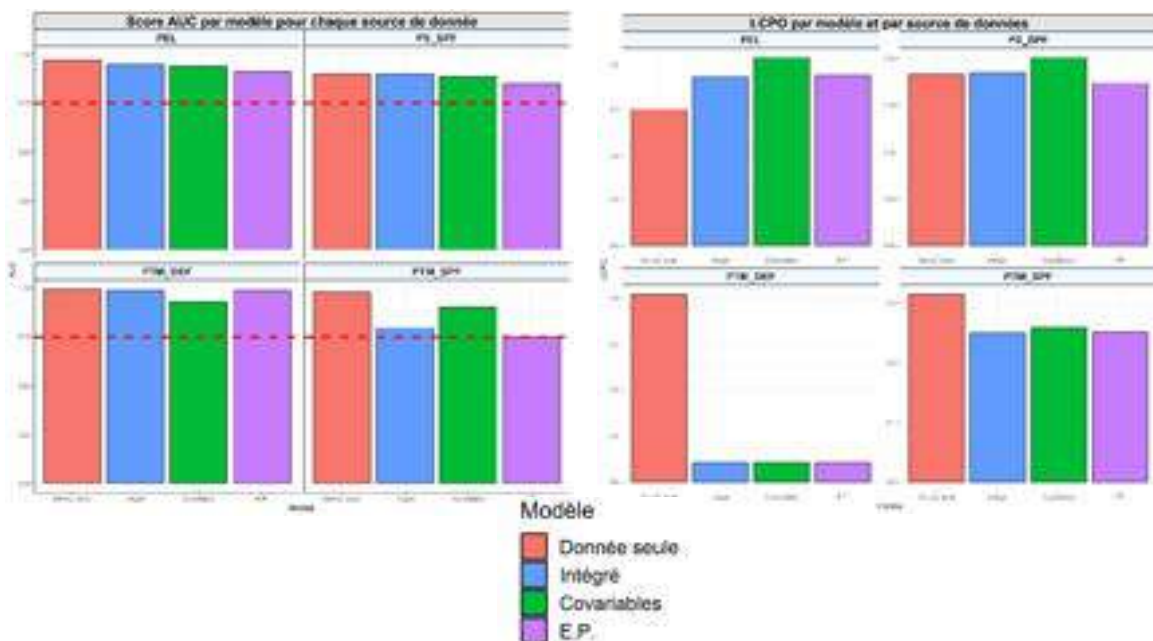


Figure 11 : Capacités prédictives des différents modèles par source de données, AUC à gauche et CPO à droite.

Au global, la prise en compte de plusieurs sources de données dans un modèle ne modifie que légèrement sa capacité prédictive, qui demeure néanmoins bonne. Dans le détail, cette réduction n'est pas systématique et varie selon la source de données (e.g. LCPO de PELGAS). La complexification du modèle via la mise en place de l'échantillonnage préférentiel ou de covariables accentue le plus souvent ces différences tout en réduisant globalement les performances prédictives du modèle.

3.2. Analyse spatio-temporelle

Le modèle spatio-temporel (Eq 7 et 8) est ajusté sur un modèle intégré « simple » combinant différentes sources de données sur une année à un pas de temps mensuel (campagnes PELGAS en mai, JUVENA en septembre, et trois flottilles commerciales disposant de données chaque mois), mais sans processus d'échantillonnage préférentiel et sans covariables.

Les cartographies inférées indiquent des présences de sardine principalement dans les zones côtières (Figure 12, pour l'année 2018 uniquement). Les zones de présences élevées sont situées entre la Baie de Vilaine et l'Estuaire de la Gironde ainsi qu'au niveau du Bassin de Parentis au Sud. Ces patrons de présences s'étendent sur le plateau continental en Mai et Septembre, qui correspondent aux périodes durant lesquelles sont réalisées les échantillonnages par les campagnes scientifiques. En hiver, les zones de présences de sardines sont très réduites dans l'espace du fait de la faible quantité de données commerciales disponibles sur cette période.

Pour les mois où seules les données commerciales sont disponibles (tous les mois sauf mai et septembre), les zones échantillonnées par les flottilles commerciales sont restreintes à la bande côtière et ne permettent pas d'obtenir des inférences fiables sur l'ensemble du GdG. La prédiction de la présence/absence au-delà de la frange côtière fréquentée par les pêcheurs ciblant ces espèces n'est pas interprétable. L'analyse des cartographies des écarts types associés révèlent des zones très vastes où les écarts-types atteignant des valeurs élevées (autour de 0.40 - Figure 13).

Concernant la capacité prédictive du modèle spatio-temporel, les AUC sont très élevées (presque systématiquement supérieures à 0.90, des valeurs minimales à 0.75) (Figure 14). Notons que pour certains mois aucune AUC n'a pu être calculée, les observations étant uniquement constituées soit de présences soit d'absences.

Cette approche est étendue à la période 2009-2018 (chaque année est traitée indépendamment des autres ; cartes en Annexes 19 à 36). Les analyses sont semblables et les observations commerciales sont localisées, la saisonnalité de l'activité de pêche est plus marquée sur le début de la série chronologique. Des scores AUC élevés sont également obtenus pour chaque année et source de données avec, dans 95% des cas, des valeurs d'AUC supérieures à 0.70. (Figure 15 et Annexes 17 et 18)

Cependant, il est intéressant de noter que la moitié des années étudiées (2010, 2012, 2013, 2014, 2016) n'identifient aucun patron de distribution de sardine au large malgré un échantillonnage dans cette zone (Figure 16 pour exemple avec l'année 2016). De plus, pour ces années, les patrons sont similaires à ceux décrits par la donnée commerciale seule sur la période printemps-automne (Figure 16). Ainsi, dans certains cas, la donnée commerciale semble permettre de décrire une partie de la distribution de la sardine malgré sa restriction à des zones côtières et identifier une stabilité des patrons de présence.

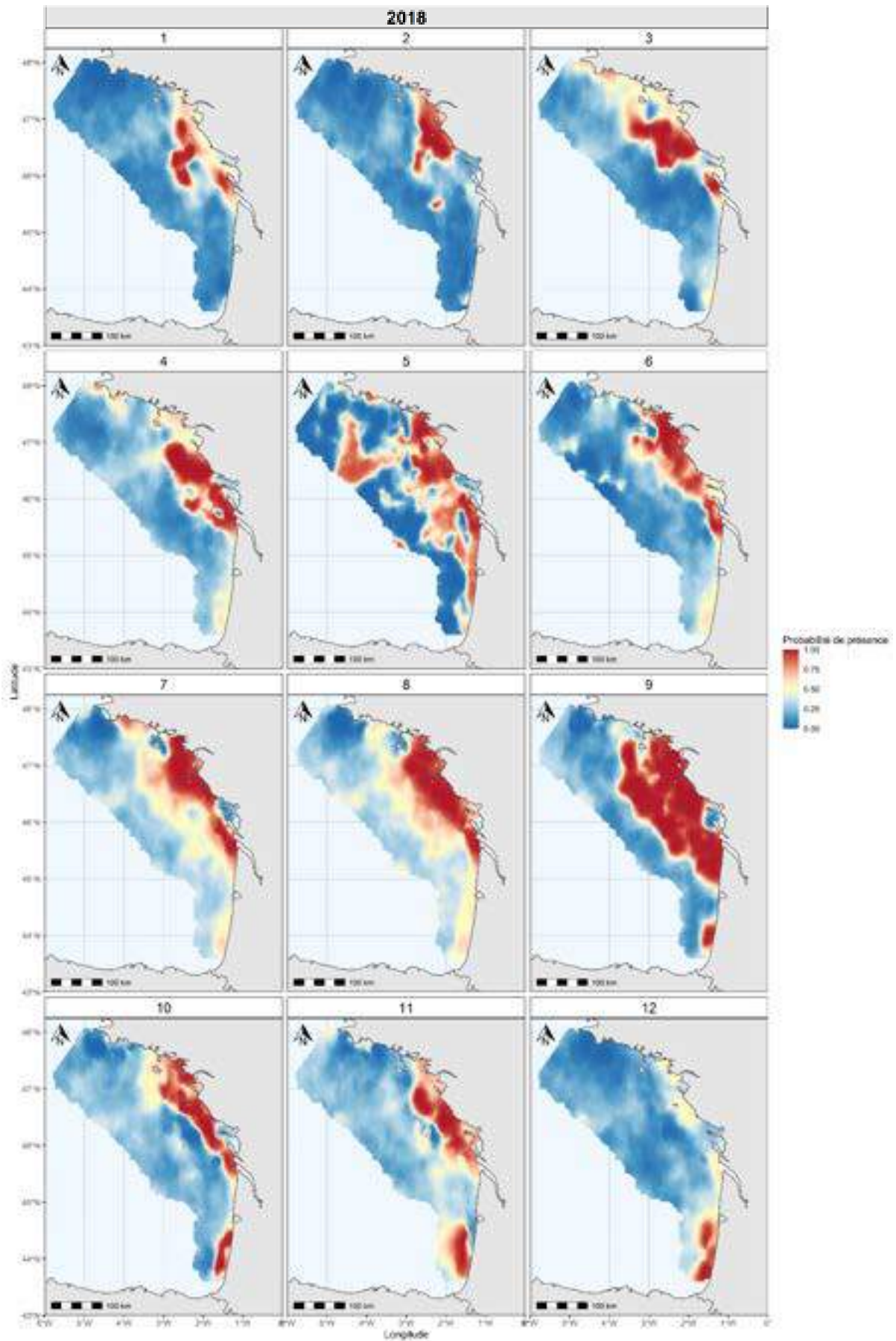


Figure 12 : Cartographie mensuelle de la probabilité de présence de la sardine dans le GdG en 2018

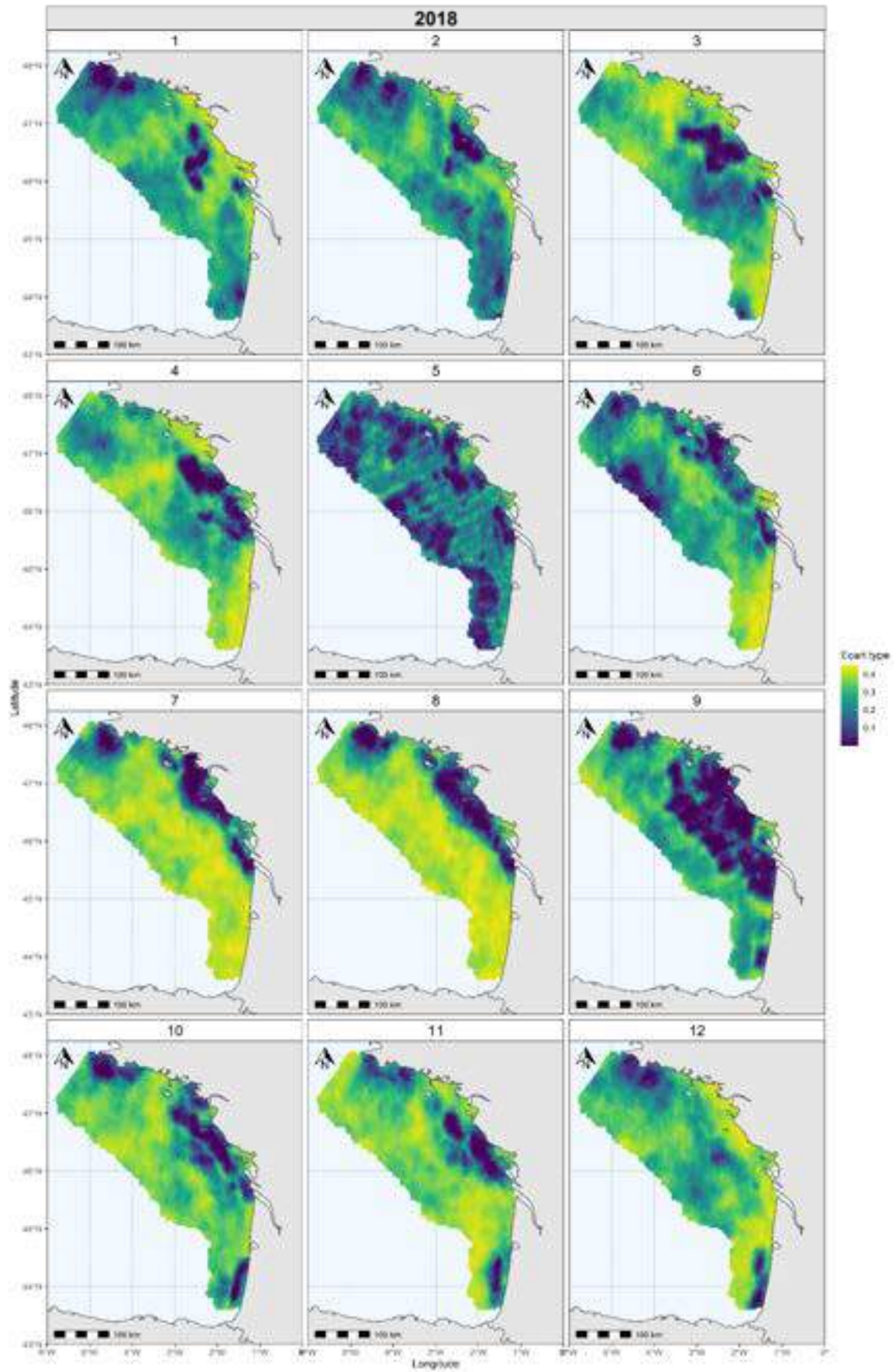


Figure 13 : Cartographie mensuelle de l'écart-type de la probabilité de présence de la sardine dans le GdG en 2018.

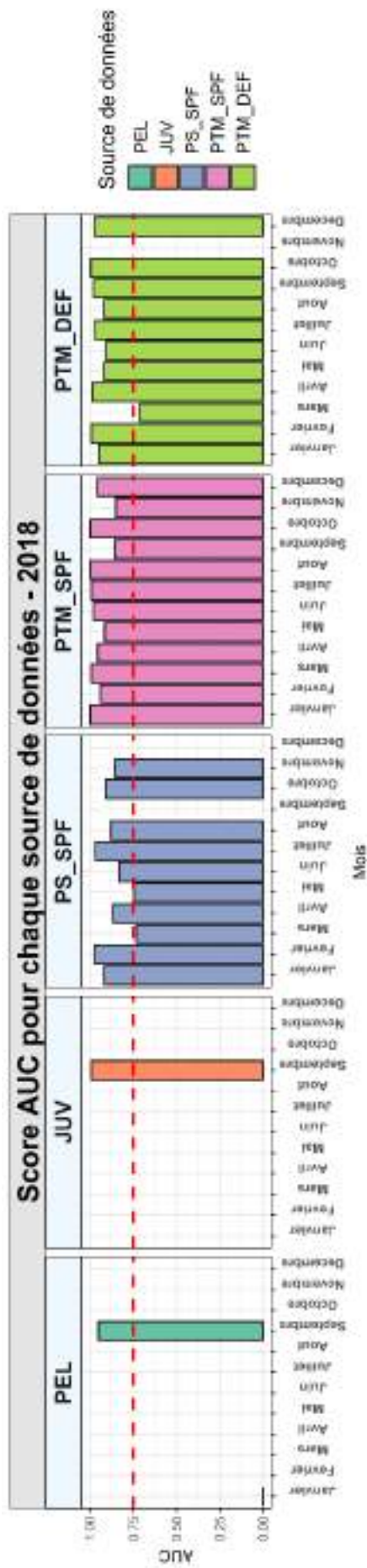


Figure 14 : Score AUC du modèle spatio-temporel par mois et source de donnée

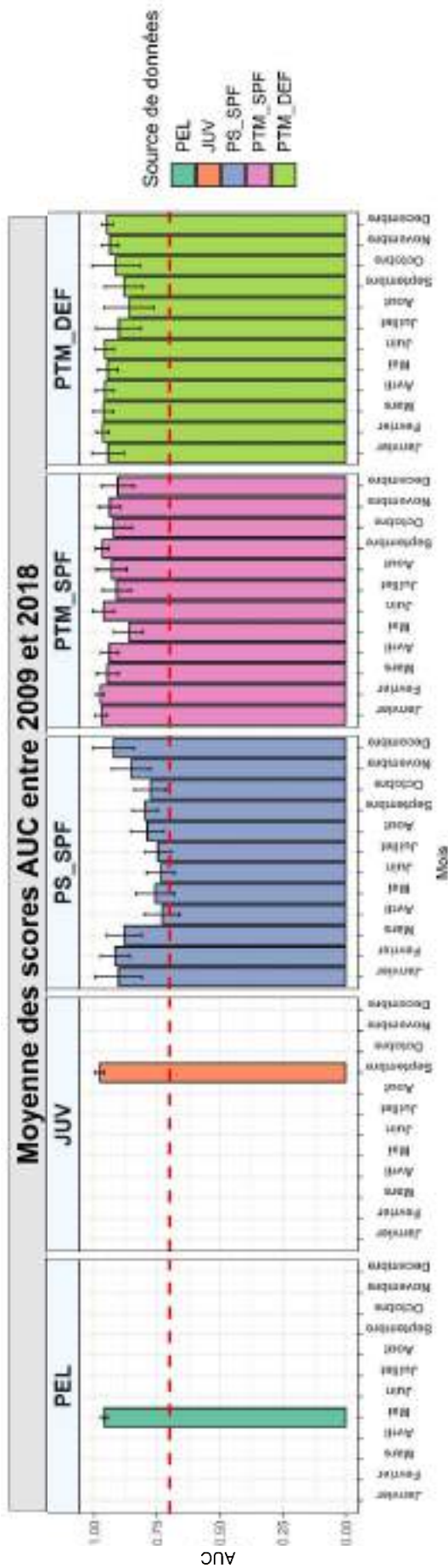


Figure 15 : AUC moyenné par mois et source de donnée (2009-2018) et écarts types. Le trait rouge correspond à une AUC de 0.70.

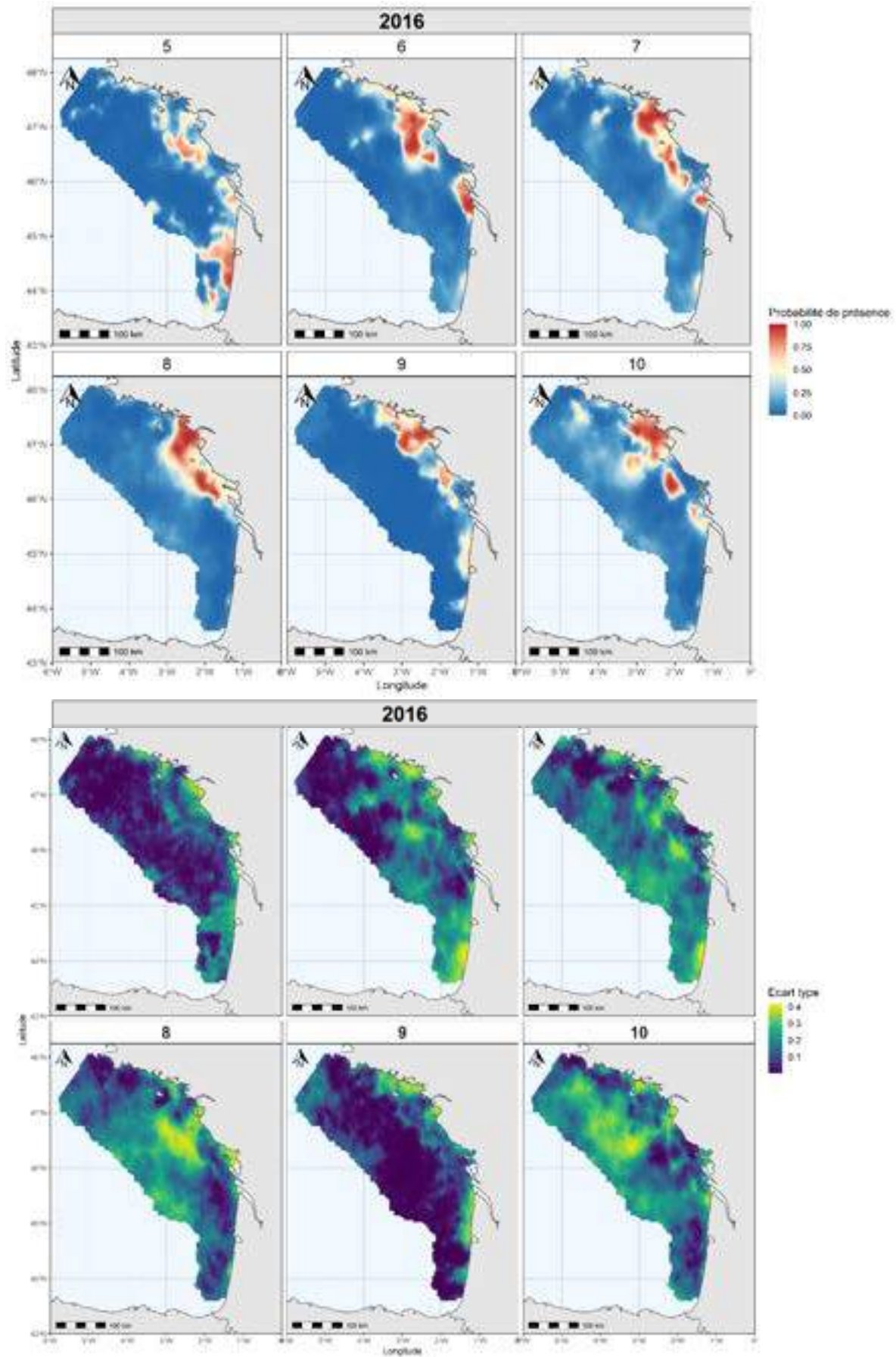


Figure 16 : Cartographies de la probabilité de présence de la sardine et d'écarts-types associés dans le GdG entre mai et octobre 2016

4. Discussion

Dans ce travail, nous avons développé un cadre de modélisation permettant de combiner des données issues de campagnes scientifiques et commerciales pour inférer la distribution de la sardine (carte de probabilité de présence) dans le Golfe de Gascogne. Le modèle a été construit en adaptant une approche de modélisation initialement développée pour des espèces démersales (Alglave *et al.*, under review). L'analyse a été conduite d'abord sur un seul pas de temps, puis à l'échelle d'une année entière dans un second temps en ajoutant une composante temporelle au modèle. Les résultats permettent de mettre en évidence des différences notables avec les cas d'étude benthodémersaux. Ils révèlent l'intérêt de combiner différentes sources de données pour inférer la distribution de la sardine sur l'ensemble de l'année, mais aussi les limites de l'approche liées en partie au fonctionnement des pêcheries de petits pélagiques.

4.1. Une inférence à l'échelle du mois dirigée par la donnée scientifique

Les analyses menées sur les mois où la donnée scientifique est disponible (i.e. les mois de mai et septembre) montrent que pour ces deux mois, les inférences sur la distribution de la sardine sont dominées par les données de campagnes scientifiques. L'intégration des données commerciales permet de compléter l'information fournie par les données scientifiques dans des zones côtières très localisées.

Ces résultats sont différents de ceux obtenus sur des espèces démersales par Alglave *et al.* (under review) qui indiquent au contraire que les données commerciales dominent les inférences. Ces différences sont liées à la taille relative des deux sources de données ainsi qu'à la couverture spatiale des données commerciales.

La première différence essentielle concerne la taille des jeux de données scientifique et commercial. Dans le cas de la sardine, les données scientifiques sont issues d'un échantillonnage acoustique (PELGAS et JUVENA), alors que les données scientifiques utilisées par Alglave *et al.* (under review) sont issues de données des campagnes EVHOE obtenues par traits de chaluts. Ainsi le nombre d'observations est très différent : les campagnes acoustiques fournissent plus de 1500 observations (1800 en 2018 pour PELGAS et 1600 pour JUVENA) pour un mois, contre 150 observations pour les campagnes EVHOE. En effet, la majorité des plans d'échantillonnages des campagnes scientifiques sont mis en œuvre pour fournir des indices d'abondance sans biais ou des données biologiques pour l'évaluation des stocks, plutôt que pour décrire précisément la distribution des espèces d'intérêts halieutique. Dans le cas des poissons pélagiques comme la sardine, les campagnes acoustiques permettent de couvrir de larges aires d'études avec un échantillonnage dense sur une courte période, fournissant des estimations de biomasses tout en renseignant précisément sur la répartition de la spatiale (Georgakarakos *et al.*, 2011).

Une seconde différence est liée à la répartition spatiale des données commerciales. Dans le cas des pêcheries démersales traitées dans Alglave *et al.*, les pêcheries commerciales couvrent un large domaine spatial comparable à la couverture spatiale des données de campagne. Dans le cas de la sardine, les coups de pêche se concentrent exclusivement dans la zone côtière (et sur des zones déjà bien couvertes par les données scientifiques). Ainsi, l'essentiel de la zone de distribution de la sardine n'est pas couvert par la pêche commerciale, alors qu'elle est bien couverte par les données scientifiques.

4.2. Capacité prédictive du modèle

Les cartographies de la répartition spatiale de la sardine semblent précises dans les zones où la donnée est disponible. Les écarts-types associés aux probabilités de présences sont relativement faibles sur les secteurs où la donnée est disponible. De plus, les scores AUC sont élevés (AUC > 0.70) pour toutes les sources de ce qui laisse penser que le modèle produit des prédictions fiables (Hosmer *et al.*, 2013).

Cependant l'utilisation de l'AUC pour justifier de la qualité des sorties du modèle est discutable puisque l'AUC est calculé à partir de la donnée ajustée au modèle. Par conséquent, l'AUC peut conduire à surestimer les performances du modèle (Jiménez-Valverde *et al.*, 2008; Lobo *et al.*, 2008). En particulier, Lobo *et al.* (2008) ont remis en question la fiabilité de l'AUC comme mesure unique de la performance du modèle.

4.3. Mise en perspective des cartes de distribution avec les connaissances déjà disponibles

La comparaison des patrons de distribution avec ceux obtenus par les différentes études révèlent une certaine cohérence (Annexe 37 à 40). En étudiant la distribution spatiale des zones de fraies de la sardine dans le GdG, Bellier *et al.*, (2007) identifient 3 sites récurrents : (1) du sud-ouest de la Bretagne, au nord de l'estuaire de la Gironde (2) au sud de l'estuaire de la Gironde dans le Bassin de Parentis et (3) sur le rebord du plateau continental entre 46°N - 47°30'N et -6°W - -4°50'W. Ces sites correspondent à des zones de plus fortes probabilités de présences d'œufs de sardines et de faibles variances. Les 2 premiers sites, les plus côtiers, sont également observés et décrits par Doray *et al.* (2018a) et Huret *et al.* (2018) en termes de biomasses de sardines et d'abondances d'œufs. De même, Petitgas *et al.* (2020) complètent et confirment ces analyses, d'une part en identifiant les mêmes zones de fortes abondances d'œufs et en suggérant un changement dans leur distribution, se rapprochant des zones côtières à partir de 2009. Au cours de cette étude nous avons pu identifier des zones semblables de fortes probabilités de présences : Pointe du Raz au Plateau de Rochebonne (du Sud-Ouest de la Bretagne au Nord de l'estuaire de Gironde (1)) et au niveau du Bassin de Parentis (Au nord de l'estuaire de la Gironde (2)) au mois de mai, mais également dans une certaine mesure durant la période printemps-automne pour chaque année étudiée.

Le troisième site identifié par Bellier *et al.* (2007) correspond également à une zone de fortes probabilités de présences, cette dernière n'est cependant observable qu'à partir des campagnes scientifiques, et uniquement certaines années (2009, 2015 et 2018). Ces résultats sont donc aussi en cohérence avec les conclusions de Petitgas *et al.* (2020) qui rapporte que les distributions de la sardine se rapprochent de ces zones côtières.

Pour autant, il est important de rappeler que ces différentes études portent sur les œufs de sardines (en abondance ou présence) ou sur la biomasse de sardines et pas sur la probabilité de présence.

4.4. Limites de l'étude

4.4.1. Spécificités de l'activité de pêche à la sardine dans le Golfe de Gascogne

L'exploitation de la sardine par les flottilles commerciales se caractérise par un lien très fort entre l'activité de pêche et les demandes du marché, avec pour conséquence une forte saisonnalité ainsi qu'une activité de pêche très localisée dans l'espace et côtière. A cela s'ajoute l'hyperstabilité des CPUE qui rend impossible le calcul d'indices d'abondances ou de biomasses à partir des captures commerciales. Ainsi, les captures de sardines par les professionnels de la pêche renseignent sur la **présence** du poisson **commercialisable** et sur des **zones restreintes géographiquement et temporellement**. Elles ne donnent aucune indications sur le reste de la population en dehors de leur période de pêche. Ces limites de la donnée commerciale ont des implications fortes sur les prédictions du modèle intégré et sur les interprétations qu'on peut en tirer.

La concentration de l'activité de pêche dans des zones côtières est une conséquence des demandes du marché. En effet, les sardines doivent être vendues fraîche le jour même pour être transformées dans la journée, ainsi les conserveries fonctionnent souvent par contrats ou de gré à gré, et commandent un tonnage précis dans un délai court. Les pêcheurs sont donc incités à ne capturer que la commande passée par les conserveries dans un délai très court garantissant une réponse rapide à la demande en maximisant la fraîcheur (et la qualité) du poisson. Aussi, quand ils ciblent la sardine, ils ont intérêts à privilégier les zones poissonneuses les moins éloignées, même si les abondances sont élevées au large. Pour cette raison, les données d'observations de sardines, en particulier, celles des senneurs (PS_SPF), se concentrent dans des zones très précises peu éloignées des zones de débarquements et de traitements dans les conserveries :

- Bolincheurs : Nord de la Bretagne et Bassin de Parentis – Douarnenez ; Lorient ; Quiberon ; Saint-Jean-de-Luz –
- Chaluts en bœufs pélagiques : Baie de Villaine jusqu'au Plateau de Rochebonne et Bassin de Parentis – La Turballe ; Saint-Gilles-Croix-de-Vie ; La Cotinière ; Saint-Jean-de-Luz –

Par ailleurs, les inférences sont également restreintes par la saisonnalité de l'activité de pêche, marquée par une faible activité commerciale pendant les mois d'hiver. Cela est à mettre en lien avec les exigences des conserveries sur la condition de la sardine. Afin de garantir les qualités organoleptiques du produit transformé, les conserveries exigent chez la sardine un taux de matière grasse supérieur à 8-10%. Or ce taux varie fortement au cours de l'année, avec des minimums observés en mars-avril, qui coïncident avec la période de fraie et à partir de laquelle ce taux augmente jusqu'à atteindre des valeurs maximales en Septembre-Octobre (Bandarra *et al.*, 1997). En exigeant un taux de matière grasse élevé, les conserveries entretiennent un effet filière sur l'exploitation de la sardine qui va directement influencer sur la période d'exploitation de la sardine. La saison d'activité commence ainsi vers mai, lorsque le taux de graisse de la sardine atteint un seuil suffisamment élevé pour les conserveurs, et s'achève au début de l'hiver (Doray, *com. pers.*). C'est cet effet filière qui explique la faible quantité de données commerciales en hiver.

Enfin, il est intéressant de relever que les observations issues des données commerciales renseignent sur le poisson commercialisable. Au-delà de la restriction de la réglementation sur

la taille minimale, les exigences du marché portent également sur la taille de la sardine et sa bonne condition. En conséquence, certains navires préfèrent parfois ne garder que les individus les plus gros et rejettent ainsi ceux de petites tailles (Cornou *et al.*, 2021). En effet, on observe des rejets importants dans l'activité des chaluts pélagiques : la sardine y représente 80% des rejets (soit 20% des captures totales) avec près de 50% d'individus rejetés alors qu'ils ont une taille au-delà de la limite réglementaire de 11 cm. Ainsi, chez les chaluts pélagiques, 90 % des sardines d'une taille inférieure à 13.5 cm sont rejetées.

4.4.2. Limites de la modélisation de l'échantillonnage préférentiel

Le modèle détecte un échantillonnage préférentiel pour 2 des 3 flottilles, ce qui est cohérent avec le fonctionnement des pêcheries d'espèces pélagiques qui les ciblent particulièrement (Vermard *et al.*, 2008). En effet, les sardines du Golfe de Gascogne représentent respectivement 90% et 35% des captures et des débarquements des chaluts pélagiques et des senneurs en 2019, avec des probabilités d'occurrence de 90% et 50% (Guerineau *et al.*, 2010; Cornou *et al.*, 2021).

Toutefois, le fait de le prendre en compte dans l'inférence n'a que peu d'influence sur les patrons de distributions de présence de la sardine, ce qui est consistant avec les résultats d'Alglave et al (under review). La prise en compte de l'échantillonnage préférentiel influence légèrement les inférences sur la distribution de la sardine en révélant des zones de forte probabilité de présence dans les aires où l'activité de pêche est élevée.

Cependant, l'intégration de l'échantillonnage préférentiel tel que modélisé dans ce travail reste très discutable et n'a finalement que peu d'intérêt dans l'objectif de valoriser la donnée commerciale pour inférer la distribution de la sardine pendant les mois où aucune donnée scientifique n'est disponible. En effet, la paramétrisation de l'EP proposé dans ce travail suppose directement que la fréquentation spatiale d'une zone par les pêcheurs dépend de la probabilité de présence de la sardine dans cette zone. Ainsi, l'EP interprète une zone non fréquentée comme une zone de faible probabilité de présence. Ce résultat apparaît dans les cartographies issues de la modélisation avec échantillonnage préférentiel pendant le mois de mai, la probabilité de présence est légèrement réduite dans les zones avec une probabilité de présence moyenne (aux alentours de 0.5) sur l'ensemble du Golfe de Gascogne. En effet, les aires où aucun pêcheur ne se rend, vont être considérée comme moins poissonneuses, ce qui affecte les inférences. Or, dans le cas de cette pêcherie, les contraintes de fonctionnement de la filière précédemment citées sont telles que les zones éloignées de principaux points de débarquement ne sont pas fréquentées non pas en raison de l'absence de sardine, mais pour d'autres raisons liées à la nécessité de fournir un produit frais dans des délais brefs. Ainsi, pour les mois sans données scientifiques, la prise en compte de l'EP tel que paramétré dans ce travail conduirait à la conclusion (fausse) que les sardines sont uniquement cantonnées aux zones de pêche.

Une perspective claire pour ce travail serait de développer la modélisation de l'échantillonnage préférentiel de façon à prendre en compte l'ensemble des facteurs qui expliquent la répartition des coups de pêche, comme la distance à la côte et/ou aux criées et conserveries par exemple.

4.4.3. Estimation de l'effet des covariables environnementales

L'ajout des covariables dans le modèle n'a pas permis de mettre en évidence d'effets importants des covariables sur les probabilités de présences de sardine dans le GdG. Ces résultats semblent contradictoires avec les connaissances disponibles sur les petits pélagiques. En effet, leur

distribution spatiale est supposée être très dépendante des conditions environnementales (Schickele *et al.*, 2020). Cependant, Planque *et al.* (2007) observent que l'influence des facteurs hydrographiques sur la distribution de la sardine en mai semble moins prononcée que pour d'autres espèces pélagiques, notamment l'anchois. Ils remarquent ainsi que tous les facteurs hydrographiques semblent avoir un degré d'influence similaire sur la distribution de la sardine, ce qui suggère une plus grande tolérance aux conditions environnementales. Ce résultat est cohérent avec le comportement de la sardine qui est connue pour nager sur de grandes distances (Parrish *et al.*, 1981; Doston and Griffith, 1996), avoir une distribution spatiale plus fragmentée (Barange and Hampton, 1997) et être généralement plus flexible sur le plan environnemental que d'autres petits pélagiques comme l'anchois (Bakun and Broad, 2003). Ces caractéristiques peuvent ainsi mener à de grandes variations dans sa distribution spatiale d'une année à l'autre, comme observées par Doray *et al.* (2018a) qui rapportent que des sardines ont été trouvées dans la plupart des zones du Golfe de Gascogne sans que sa distribution ne soit restreinte à un habitat spécifique. De même l'ICES (2010) décrit que *S. pilchardus* présente une grande extension spatiale dans les eaux de l'Atlantique Nord-Est, où les adultes effectuent des migrations à grande échelle englobant une large gamme de conditions hydrologiques. Ces résultats suggèrent que les distributions de *S. pilchardus* pourraient être moins influencées par les conditions hydrologiques locales du GdG, et plus probablement déterminées par des processus à plus grande échelle (Stratoudakis *et al.*, 2004). Aussi, il est probable que les conditions environnementales ne soient pas assez contrastées sur le domaine d'étude pour exercer un contrôle clair et détectable sur la distribution de la sardine.

De même, les gammes de températures de surface décrites comme optimales par d'Arbault et Lacroix (1977) et Planque et al (2007) correspondent aux valeurs de SST testées dans notre cas d'étude (12-15°C), ce qui peut expliquer les faibles effets de la SST, la gamme de températures testées étant trop restreintes.

Enfin il est important de rappeler que le modèle se base sur des données de présence-absence (des '0' et des '1') qui sont moins informatives que des données d'abondances ou de biomasses, ce qui pourrait également expliquer le faible effet des covariables.

4.4.4. Biais liés aux données.

Une première limite concerne les données commerciales et la réallocation des déclarations de pêche aux pings VMS. Pour obtenir la donnée de déclaration de pêche à la résolution des pings VMS, les données de pêche sont réallouées de façon uniformes sur les pings VMS correspondants. Dans le cas d'un modèle de présence/absence, dès qu'une déclaration de pêche est positive, l'ensemble des pings associés seront considérés comme des présences ce qui pourrait conduire à la surestimation de la probabilité de présence.

Une seconde limite concerne la campagne d'échantillonnage JUVENA. D'une part, le jeu de données obtenu ne permettait pas de distinguer les classes de tailles et donc de filtrer les individus d'une taille inférieure à la réglementation en vigueur (11 cm). D'autres part, son échantillonnage adaptatif n'est pas identique dans le temps et ne couvre pas exactement l'aire d'étude ici analysée (correspondant à l'aire d'étude de PELGAS, qui est identique chaque année).

5. Conclusion

L'approche développée a permis d'étudier l'application d'un modèle intégrée combinant données de campagnes acoustiques et données 'VMS x logbooks' pour inférer des cartes de probabilités de présence/absence de la sardine du Golfe de Gascogne.

Pour les mois de mai (PELGAS) et septembre (JUVENA) pour lesquels des campagnes scientifiques acoustiques sont disponibles, les cartes reflètent essentiellement l'information portée par les données scientifiques. Elles permettent d'identifier des patrons de distribution cohérents avec la littérature. Sur les pas de temps non couverts par les données scientifiques, les données commerciales apportent de l'information dans le rayon d'action des flottilles restreint à une bande côtière. Mais une grande partie de la zone d'étude n'est pas recouverte par les flottilles commerciales et les cartes obtenues fournissent une image partielle de la distribution de la sardine.

Pour autant, l'approche a permis de produire des cartographies sur un pas de temps mensuel entre 2009 et 2018. Cela représente un apport significatif par rapport aux dernières cartes qui recouvrent l'ensemble de l'année et qui datent des années 70. De plus, les patrons de distributions côtiers semblent présenter des similitudes entre les mois de campagnes (mai et septembre) et sur la période printemps-automne ce qui laisse penser que les zones de fortes probabilités de présences de la sardine identifiées par les études précédentes (uniquement sur le mois de mai) perdurent sur les mois d'été.

L'approche de modélisation intégrée entreprise dans ce travail présente un certain nombre d'intérêt, mais est aussi marquée par de nombreuses limites qui sont résumées dans le tableau 1.

Tableau 1 : Intérêts et limites de l'approche intégrée

	Intérêts	Limites
Données commerciales 'VMS x logbooks'	Données disponibles en dehors des périodes de campagne	Activités commerciales restreinte dans l'espace (concentrée dans des zones côtières) et dans le temps (peu d'activité en hiver) Captures règlementées et structurée par la demande Réallocations des débarquements aux données VMS par carré statistique Rejets des petits individus Hyperstabilité des captures
Données scientifiques	Observations scientifiques nombreuses, échantillonnage dense et recouvrant l'ensemble du domaine d'étude	Uniquement pendant les mois de mai (PELGAS) et septembre (JUVENA)
Qualité d'ajustement du modèle	AUC élevée	Modèle de présence-absence

	Bon comportement en validation croisée	Pas de possibilité de dériver des indices d'abondance continus (pas de CPUE)
Estimations	Comportement des flottilles consistants avec la connaissance disponible pour les senneurs et les chaluts ciblant ($b > 0$ – ciblage élevé)	Faible effet des covariables environnementales sur les inférences, pas de relation espèce-habitat
	Distributions inférées cohérentes avec la littérature (en Mai).	Cartes de distribution biaisées en dehors des périodes de campagne.

Il est alors intéressant de comparer l'approche méthodologique mise en œuvre pour cartographier l'abondance d'espèces benthodémersale dans Alglave *et al.* (under review). Dans cette approche sur la sardine, les caractéristiques de la pêche commerciale sont finalement peu informatives de la distribution de la sardine. Les données scientifiques restent donc la source d'information à privilégier.

A l'inverse, dans le cas des espèces benthodémersales traitées dans Alglave *et al.* (under review), les données commerciales sélectionnées pour cartographier la distribution sont plus informatives : seuls les chalutiers de fonds sont utilisés – OTB, OTT et ceux-ci ont une couverture spatiale plus large et un comportement de ciblage opportuniste. A l'inverse du cas de la sardine, l'information portée par les données commerciales domine celle portée par les campagnes scientifiques. En effet, la taille des échantillons issus des campagnes de chalutage scientifique est beaucoup plus réduite. Le tableau 2 résume les différences entre les 2 types de cas d'étude.

Tableau 2 : Différences entre cas d'étude pélagique et benthodémersaux

Caractéristique	Espèces benthodémersales (Sole, Merlu et calamars)	Espèces pélagiques
Flottilles commerciales	Ciblage faible sur un nombre élevé d'espèces (pêcheries plurispécifiques)	Ciblage fort sur quelques espèces (pêche monospécifique pour les chaluts pélagiques)
	Plusieurs centaines de navires	Quelques dizaines de navires
	Rayon d'action large	Rayon d'action restreint dans l'espace
	Comportement opportuniste	Saisonnalité très marquée
Flottilles scientifiques	Faible taille d'échantillon (~150 traits de chalut au maximum)	Forte densité d'échantillonnage (~1500 observations acoustiques associées à des traits de chaluts)

Références

- Abbott JK, Haynie AC, Reimer MN** (2015) Hidden Flexibility: Institutions, Incentives, and the Margins of Selectivity in Fishing. *Land Econ* **91**: 169–195
- Alglave B, Rivot E, Etienne MP, Woillez M, Thor J, Vermard Y** (in press) Integrated framework accounting for preferential sampling to infer fish spatial distribution. *ICES J. Mar. Sci.*
- Arbault S, Lacroix N** (1970) Quatre ans de mesures volumétriques de plancton total dans le Golfe de Gascogne (1964-1967). *Rev Trav L'Institut Pêch Marit* **34**: 59–68
- Arbault S, Lacroix N** (1977) Œufs et larves de Clupéidés et Engraulidés dans le golfe de Gascogne (1969 - 1973). Distribution des frayères. Relations entre les facteurs du milieu et de la reproduction. *Rev Trav L'Institut Pêch Marit* **41**: 227–254
- Archambault B, Rivot E, Savina M, Le Pape O** (2018) Using a spatially structured life cycle model to assess the influence of multiple stressors on an exploited coastal-nursery-dependent population. *Estuar Coast Shelf Sci* **201**: 95–104
- August T, Harvey M, Lightfoot P, Kilbey D, Papadopoulos T, Jepson P** (2015) Emerging technologies for biological recording. *Biol J Linn Soc* **115**: 731–749
- Bakun A** (1996) *Patterns in the Ocean: Ocean Processes and Marine Population Dynamics*, San Diego: University of California Sea Grant , USA, in co-operation with Centro de Investigaciones Biológicas de Noroeste. La Paz, Baja California Sur, Mexico. 323 pp.
- Bakun A, Broad K** (2003) Environmental ‘loopholes’ and fish population dynamics: comparative pattern recognition with focus on El Niño effects in the Pacific: *Environmental loopholes and fish population dynamics*. *Fish Oceanogr* **12**: 458–473
- Bálint M, Pfenninger M, Grossart H-P, Taberlet P, Vellend M, Leibold MA, Englund G, Bowler D** (2018) Environmental DNA Time Series in Ecology. *Trends Ecol Evol* **33**: 945–957
- Bandarra NM, Batista I, Nunes ML, Empis JM, Christie WW** (1997) Seasonal Changes in Lipid Composition of Sardine (*Sardina pilchardus*). *J Food Sci* **62**: 40–42
- Barange M, Coetzee JC, Twatwa NM** (2005) Strategies of space occupation by anchovy and sardine in the southern Benguela: the role of stock size and intra-species competition. *ICES J Mar Sci* **62**: 645–654
- Barange M, Hampton I** (1997) Spatial structure of co-occurring anchovy and sardine populations from acoustic data: implications for survey design. *Fish Oceanogr* **6**: 94–108
- Bellier E, Planque B, Petitgas P** (2007) Historical fluctuations in spawning location of anchovy (*Engraulis encrasicolus*) and sardine (*Sardina pilchardus*) in the Bay of Biscay during 1967–73 and 2000–2004. *Fish Oceanogr* **16**: 1–15

- Bez N, Walker E, Gaertner D, Rivoirard J, Gaspar P** (2011) Fishing activity of tuna purse seiners estimated from vessel monitoring system (VMS) data. *Can J Fish Aquat Sci* **68**: 1998–2010
- Blangiardo M, Cameletti M, Baio G, Rue H** (2013) Spatial and spatio-temporal models with R-INLA. *Spat Spatiotemporal Epidemiol* **4** : 33 - 49
- Booth A** (2000) Incorporating the spatial component of fisheries data into stock assessment models. *ICES J Mar Sci* **57**: 858–865
- Boyra G, Rico I, Udane Martínez** (2020) Acoustic surveying of anchovy Juveniles in the Bay of Biscay: JUVENA 2020 Survey Report. doi: 10.13140/RG.2.2.36115.09768
- Branch TA, Hilborn R, Haynie AC, Fay G, Flynn L, Griffiths J, Marshall KN, Randall JK, Scheuerell JM, Ward EJ, et al** (2006) Fleet dynamics and fishermen behavior: lessons for fisheries managers. *Can J Fish Aquat Sci* **63**: 1647–1668
- Burgess JW, Shaw E** (1979) Development and ecology of fish schooling. *Oceanus* **27**: 11–17
- Conn PB, Thorson JT, Johnson DS** (2017) Confronting preferential sampling when analysing population distributions: diagnosis and model-based triage. *Methods Ecol Evol* **8**: 1535–1546
- Coombs SH, Smyth TJ, Conway DVP, Halliday NC, Bernal M, Stratoudakis Y, Alvarez P** (2006) Spawning season and temperature relationships for sardine (*Sardina pilchardus*) in the eastern North Atlantic. *J Mar Biol Assoc U K* **86**: 1245–1252
- Cornou A-S, Quinio-Scavinner M, Sagan J, Cloatre T, Dubroca L, Billet N, Roy EL, Chassanite A, Boiron-Leroy A, Martin-Baillet V** (2021) Captures et rejets des métiers de pêche français. Résultats des observations à bord des navires de pêche professionnelle en 2019. *Obsmer*. 544
- Cury P, Bakun A, Crawford RJM, Jarre A, Quinones RA, Shannon LJ, Verheye HM** (2000) Small pelagics in upwelling systems: patterns of interaction and structural changes in “wasp-waist” ecosystems. *ICES J Mar Sci* **57**: 603–618
- Delage N, Le Pape O** (2016) Inventaire des zones fonctionnelles pour les ressources halieutiques dans les eaux sous souveraineté française. Première partie: définitions, critères d’importance et méthode pour déterminer des zones d’importance à protéger en priorité. Pôle halieutique AGROCAMPUS OUEST, 2016. 36
- Diggle PJ, Menezes R, Su T** (2010) Geostatistical inference under preferential sampling. *J R Stat Soc Ser C Appl Stat* **59**: 191–232
- Doray M, Hervy C, Huret M, Petitgas P** (2018a) Spring habitats of small pelagic fish communities in the Bay of Biscay. *Prog Oceanogr* **166**: 88–108
- Doray M, Masse J, Petitgas P** (2010) Pelagic fish stock assessment by acoustic methods at Ifremer. R.INT. DOP/DCN/EMH 10- 02. 18
- Doray M, Petitgas P, Romagnan JB, Huret M, Duhamel E, Dupuy C, Spitz J, Authier M, Sanchez F, Berger L, et al** (2018b) The PELGAS survey: Ship-based integrated

- monitoring of the Bay of Biscay pelagic ecosystem. *Prog Oceanogr* **166**: 15–29
- Doston RC, Griffith DA** (1996) A high-speed midwater trawl for collecting coastal pelagic fishes. *CalCOFI Rep* **37**: 134–139
- Ettahiri O, Berraho A, Vidy G, Ramdani M, Do chi T** (2003) Observation on the spawning of *Sardina* and *Sardinella* off the south Moroccan Atlantic coast (21–26°N). *Fish Res* **60**: 207–222
- Fréon P, Misund OA** (1999) Dynamics of pelagic fish distribution and behaviour: effects on fisheries and stock assessment. Fishing News Books, Oxford
- Furnestin J, Furnestin M-L** (1959) La reproduction de la sardine et de l'anchois des côtes atlantiques du Maroc (saisons et aires de ponte). *Rev Trav Inst Pêch Marit* **23**: 79–104
- Gaston KJ** (2000) Global patterns in biodiversity. *Nature* **405**: 220–227
- Gatti P, Cominassi L, Duhamel E, Grellier P, Le Delliou H, Le Mestre S, Petitgas P, Rabiller M, Spitz J, Huret M** (2018) Bioenergetic condition of anchovy and sardine in the Bay of Biscay and English Channel. *Prog Oceanogr* **166**: 129–138
- Georgakarakos S, Trygonis V, Haralabous J** (2011) Accuracy of Acoustic Methods in Fish Stock Assessment Surveys. *Sonar Syst.* doi: 10.5772/18631
- Gerritsen H, Lordan C** (2011) Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES J Mar Sci* **68**: 245–252
- Gibb R, Browning E, Glover-Kapfer P, Jones KE** (2019) Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol Evol* **10**: 169–185
- Girardin R, Hamon KG, Pinnegar J, Poos JJ, Thébaud O, Tidd A, Vermard Y, Marchal P** (2017) Thirty years of fleet dynamics modelling using discrete-choice models: What have we learned? *Fish Fish* **18**: 638–655
- Gómez-Rubio, V** (2020) Bayesian Inference with INLA. Chapman & Hall/CRC Press. Boca Raton, FL.
- Gonzalez GM, Wiff R, Marshal CT, Cornullier T** (2021) Estimating spatio-temporal distribution of fish and gear selectivity functions from pooled scientific survey and commercial fishing data. *Fish Res* **243**: 11
- Guerineau L, Rochet M-J, Peronnet I** (2010) Panorama des rejets dans les pêcheries françaises. 49
- Hilborn R, Walters CJ** (1992) Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty., Springer US.
- Hintzen NT, Bastardie F, Beare D, Piet GJ, Ulrich C, Deporte N, Egekvist J, Degel H** (2012) VMStools: Open-source software for the processing, analysis and visualisation of fisheries logbook and VMS data. *Fish Res* **115–116**: 31–43

- Hosmer DW, Lemeshow S, Sturdivant RX** (2013) Applied logistic regression, Third edition. Wiley, Hoboken, New Jersey
- Hunter E, Berry F, Buckley AA, Stewart C, Metcalfe JD** (2006) Seasonal migration of thornback rays and implications for closure management: Ray migration and closure management. *J Appl Ecol* **43**: 710–720
- Hunter, JR, Alheit J** (1995) International GLOBEC Small Pelagic Fishes and Climate Change program. Report of the First Planning Meeting, La Paz, Mexico, June 20-24, 1994.. GLOBEC Rep. 8, 72 pp.
- Huret M, Bourriau P, Doray M, Gohin F, Petitgas P** (2018) Survey timing vs. ecosystem scheduling: Degree-days to underpin observed interannual variability in marine ecosystems. *Prog Oceanogr* **166**: 30–40
- ICES** (2005) Report of the Workshop on Survey Design and Data Analysis (WKSAD), Sète, France. ICES CM 2005/B:07. 170
- ICES** (2016) Report of the Working Group on Southern Horse Mackerel, Anchovy and Sardine (WGHANSA), 24–29 June 2016, Lorient, France. ICES CM 2016/ACOM:17. 588 pp.
- ICES** (2020) Sardine (*Sardina pilchardus*) in divisions 8.a-b and 8.d (Bay of Biscay). Rep ICES Advis Comm 2020 ICES Advice 2020 Pil278abd. doi: 10.17895/ICES.ADVICE.5906
- Isaac NJB, Jarzyna MA, Keil P, Dambly LI, Boersch-Supan PH, Browning E, Freeman SN, Golding N, Guillera-Arroita G, Henrys PA, et al** (2020) Data Integration for Large-Scale Models of Species Distributions. *Trends Ecol Evol* **35**: 56–67
- Jiménez-Valverde A, Lobo JM, Hortal J** (2008) Not as good as they seem: the importance of concepts in species distribution modelling. *Divers Distrib* **14**: 885–890
- Krainski ET, Gomez-Rubio V, Bakka H, Lenzi A, Castro-Camilo D, Simpson D, Lindgren F, Rue H** (2019) Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. Chapman & Hall/CRC Press. Boca Raton, FL.
- Lamoreux JF, Morrison JC, Ricketts TH, Olson DM, Dinerstein E, McKnight MW, Shugart HH** (2006) Global tests of biodiversity concordance and the importance of endemism. *Nature* **440**: 212–214
- Lavialle G, Duhamel E, Véron M** (2019) Preliminary results on the comparison of the sardine growth between the channel part (>48°N; Douarnenez Bay) and the south part of the bay of Biscay Sardine.
- Lee J, South AB, Jennings S** (2010) Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. *ICES J Mar Sci* **67**: 1260–1271
- L’Herrou R** (1967) Répartition des oeufs et larves de sardine dans le Golfe de Gascogne et sur le plateau celtique (mai 1966 ; février et mai 1967). *CM 1967 J* 14
- Lindgren F, Rue H, Lindström J** (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach:

- Link between Gaussian Fields and Gaussian Markov Random Fields. *J R Stat Soc Ser B Stat Methodol* **73**: 423–498
- Lobo JM, Jiménez-Valverde A, Real R** (2008) AUC: a misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr* **17**: 145–151
- Mackinson S** (1999) Cross-scale observations on distribution and behavioural dynamics of ocean feeding Norwegian spring-spawning herring (*Clupea harengus* L.). *ICES J Mar Sci* **56**: 613–626
- Maclennan DN, Fernandes PG, Dalen J** (2002) A consistent approach to definitions and symbols in fisheries acoustics. *ICES J Mar Sci* **59**: 365–369
- Marrs SJ, Tuck ID, Atkinson RJA, Stevenson TDI, Hall C** (2002) Position data loggers and logbooks as tools in fisheries research: results of a pilot study and some recommendations. *Fish Res* **58**: 109–117
- Marshall CE, Glegg GA, Howell KL** (2014) Species distribution modelling to support marine conservation planning: The next steps. *Mar Policy* **45**: 330–332
- Marzuki MI** (2017) VMS data analyses and modeling for the monitoring and surveillance of Indonesian fisheries. *Computer Vision and Pattern Recognition [cs.CV]*. Ecole nationale supérieure Mines-Télécom Atlantique
- Maunder MN, Punt AE** (2013) A review of integrated analysis in fisheries stock assessment. *Fish Res* **142**: 61–74
- Meyer C, Holland K, Papastamatiou Y** (2007) Seasonal and diel movements of giant trevally *Caranx ignobilis* at remote Hawaiian atolls: implications for the design of Marine Protected Areas. *Mar Ecol Prog Ser* **333**: 13–25
- Moriarty M, Sethi SA, Pedreschi D, Smeltz TS, McGonigle C, Harris BP, Wolf N, Greenstreet SPR** (2020) Combining fisheries surveys to inform marine species distribution modelling. *ICES J Mar Sci* **77**: 539–552
- Murray LG, Hinz H, Hold N, Kaiser MJ** (2013) The effectiveness of using CPUE data derived from Vessel Monitoring Systems and fisheries logbooks to estimate scallop biomass. *ICES J Mar Sci* **70**: 1330–1340
- Navigs robert** (2005) *Fishing Vessel Monitoring Systems: Past, Present and Future*. The High Seas Task Force OECD. Paris.
- Nielsen JR** (2015) Methods for integrated use of fisheries research survey information in understanding marine fish population ecology and better management advice: improving methods for evaluation of research survey information under consideration of survey fish detection and catch efficiency.
- Parrish RH, Nelson CS, Bakun A** (1981) Transport Mechanisms and Reproductive Success of Fishes in the California Current. *Biol Oceanogr* **1**: 175–203
- Parrish RH, Serra R, Grant WS** (1989) The Monotypic Sardines, *Sardina* and *Sardinops* : Their Taxonomy, Distribution, Stock Structure, and Zoogeography. *Can J Fish Aquat*

- Pennino MG, Paradinas I, Illian JB, Muñoz F, Bellido JM, López-Quílez A, Conesa D** (2019) Accounting for preferential sampling in species distribution models. *Ecol Evol* **9**: 653–663
- Petitgas P, Masse J, Bourriau Paul, Beillois P, Delmas D, Herbland A, Koueta N, Froidefond JM, Santos M** (2006) Hydro-plankton characteristics and their relationship with sardine and anchovy distributions on the French shelf of the Bay of Biscay. *Sci Mar* **70**: 161–172
- Petitgas P, Renard D, Desassis N, Huret M, Romagnan J-B, Doray M, Woillez M, Rivoirard J** (2020) Analysing Temporal Variability in Spatial Distributions Using Min–Max Autocorrelation Factors: Sardine Eggs in the Bay of Biscay. *Math Geosci* **52**: 337–354
- Pitcher TJ** (1980) Some ecological consequences of fish school volumes. *Freshw Biol* **10**: 539–544
- Pitcher TJ** (1995) The impact of pelagic fish behavior on fisheries. *Sci Mar* **59**: 295–306
- Planque B, Bellier E, Lazure P** (2007) Modelling potential spawning habitat of sardine (*Sardina pilchardus*) and anchovy (*Engraulis encrasicolus*) in the Bay of Biscay. *Fish Oceanogr* **16**: 16–30
- Punt AE, Dunn A, Elvarsson BP, Hampton J, Hoyle SD, Maunder MN, Methot RD, Nielsen A** (2020) Essential features of the next-generation integrated fisheries stock assessment package: A perspective. *Fish Res* **229**: 105617
- van Putten IE, Kulmala S, Thébaud O, Dowling N, Hamon KG, Hutton T, Pascoe S** (2012) Theories and behavioural drivers underlying fleet dynamics models: Theories and behavioural drivers. *Fish Fish* **13**: 216–235
- Quero J-C, Dardignac J, Vayne J-J** (1989) Les poissons du golfe de Gascogne.
- Robinson CJ** (2004) Responses of the northern anchovy to the dynamics of the pelagic environment: identification of fish behaviours that may leave the population under risk of overexploitation. *J Fish Biol* **64**: 1072–1087
- Rochette S, Le Pape O, Vigneau J, Rivot E** (2013) A hierarchical Bayesian model for embedding larval drift and habitat models in integrated life cycles for exploited fish. *Ecol Appl* **23**: 1659–1676
- Rue H, Martino S, Chopin N** (2009) Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations. *J R Stat Soc Ser B* **71**: 319–392
- Rufener M-C** (2020) Rufener, M.-C. 2020. Integrating commercial fisheries and scientific survey data: Advances, new tools and applications to model the fish and fishery dynamics. National Institute of Aquatic Resources, Denmark, DTU Aqua. 209 pp.
- Salas S, Gaertner D** (2004) The behavioural dynamics of fishers: management implications.

- Schaub M, Gimenez O, Sierro A, Arlettaz R** (2007) Use of Integrated Modeling to Enhance Estimates of Population Dynamics Obtained from Limited Data. *Conserv Biol* **21**: 945–955
- Schickele A, Leroy B, Beaugrand G, Goberville E, Hattab T, Francour P, Raybaud V** (2020) Modelling European small pelagic fish distribution: Methodological insights. *Ecol Model* **416**: 108902
- Silva A, Garrido S, Ibaibarriaga L, Pawlowski L, Riveiro I, Marques V, Ramos F, Duhamel E, Iglesias M, Bryère P, et al** (2019) Adult-mediated connectivity and spatial population structure of sardine in the Bay of Biscay and Iberian coast. *Deep Sea Res Part II Top Stud Oceanogr* **159**: 62–74
- Silva A, Skagen DW, Uriarte A, Masse J, Santos MB, Marques V, Carrera P, Beillois P, Pestana G, Porteiro C, et al** (2009) Geographic variability of sardine dynamics in the Iberian Biscay region. *ICES J Mar Sci J Cons* **66**: 495–508
- Spitz J, Ridoux V, Trites AW, Laran S, Authier M** (2018) Prey consumption by cetaceans reveals the importance of energy-rich food webs in the Bay of Biscay. *Prog Oceanogr* **166**: 148–158
- Stephenson F, Mill AC, Scott CL, Stewart GB, Grainger MJ, Polunin NVC, Fitzsimmons C** (2018) Socio-economic, technological and environmental drivers of spatio-temporal changes in fishing pressure. *Mar Policy* **88**: 189–203
- Stratoudakis Y, Coombs S, de Lanzós AL, Halliday N, Costas G, Caneco B, Franco C, Conway D, Santos MB, Silva A, et al** (2007) Sardine (*Sardina pilchardus*) spawning seasonality in European waters of the northeast Atlantic. *Mar Biol* **13**
- Tosello-Bancal F** (1994) L' évolution de la pêche de la sardine sur le littoral français. Thèse de doctorat : Géographie. Paris 4, Paris
- Vermard Y, Marchal P, Mahévas S, Thébaud O** (2008) A dynamic model of the Bay of Biscay pelagic fleet simulating fishing trip choice: the response to the closure of the European anchovy (*Engraulis encrasicolus*) fishery in 2005. *Can J Fish Aquat Sci* **65**: 2444–2453
- Whitehead P** (1985) Clupeoid fishes of the world. An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, anchovies and wolfherrings. Part 1 - Chirocentridae, Clupeidae and Pristigasteridaec., Rome. Rome
- Whittaker RH, Levin SA, Root RB** (1973) Niche, Habitat, and Ecotope. *Am Nat* **107**: 321–338
- Worm B, Hilborn R, Baum JK, Branch TA, Collie JS, Costello C, Fogarty MJ, Fulton EA, Hutchings JA, Jennings S, et al** (2009) Rebuilding Global Fisheries. *Science* **325**: 578–585
- Wright AJ, Kyhn LA** (2015) Practical management of cumulative anthropogenic impacts with working marine examples: Practical Cumulative Impact Management. *Conserv Biol* **29**:

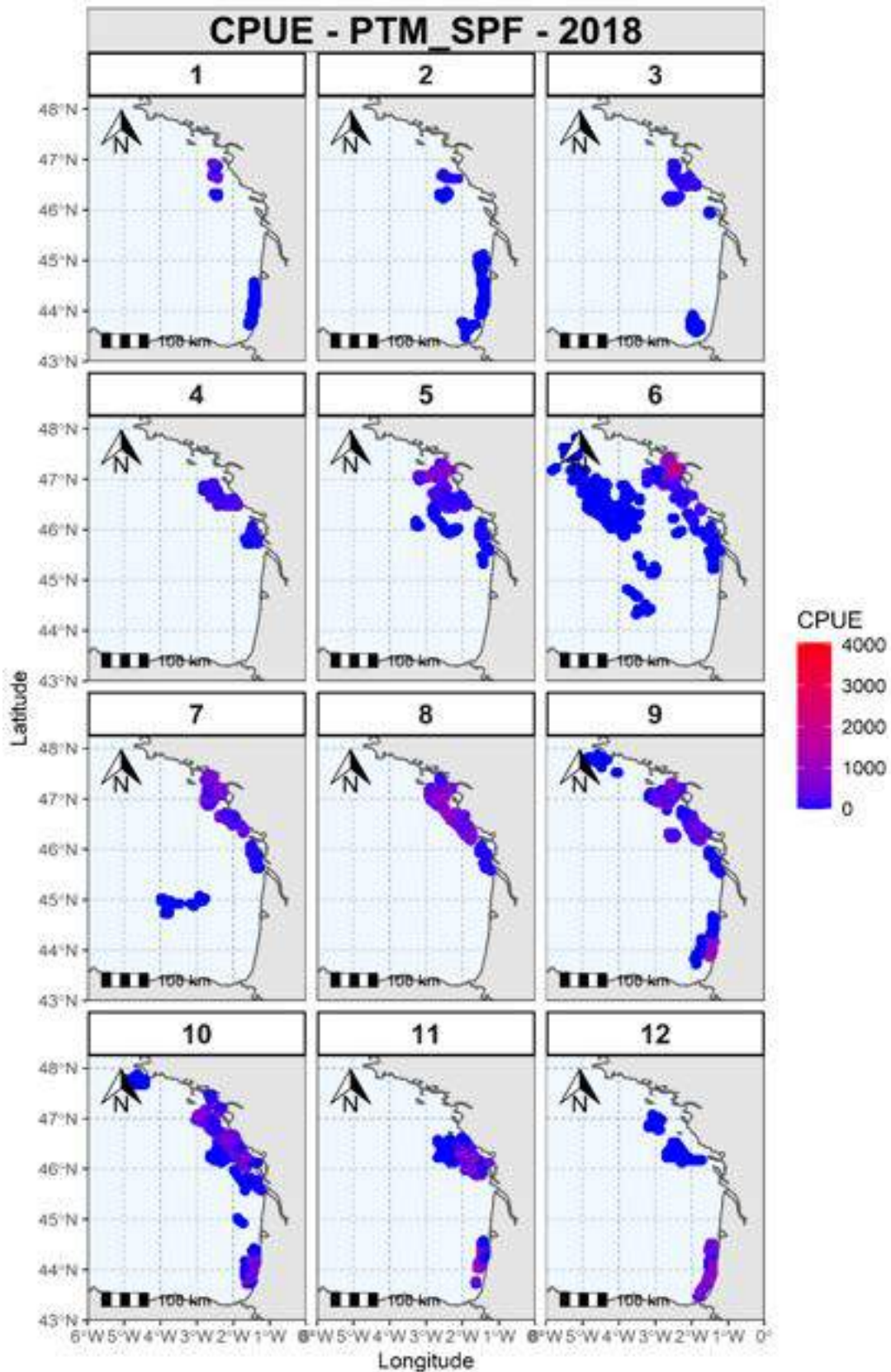
333–340

Zipkin EF, Inouye BD, Beissinger SR (2019) Innovations in data integration for modeling populations. *Ecology* e02713

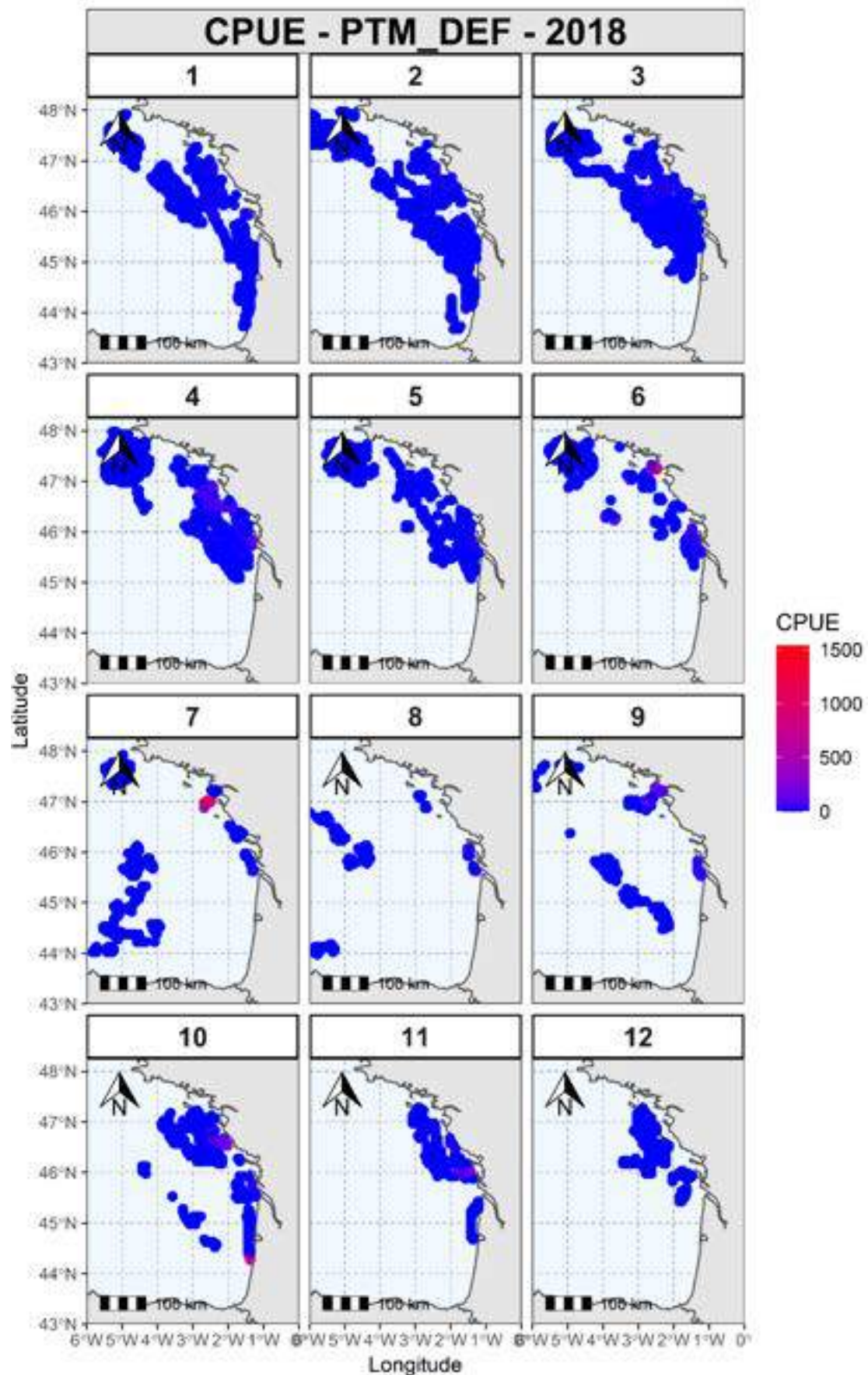
Zwolinski JP, Oliveira PB, Quintino V, Stratoudakis Y (2010) Sardine potential habitat and environmental forcing off western Portugal. *ICES J Mar Sci* **67**: 1553–1564

Annexes

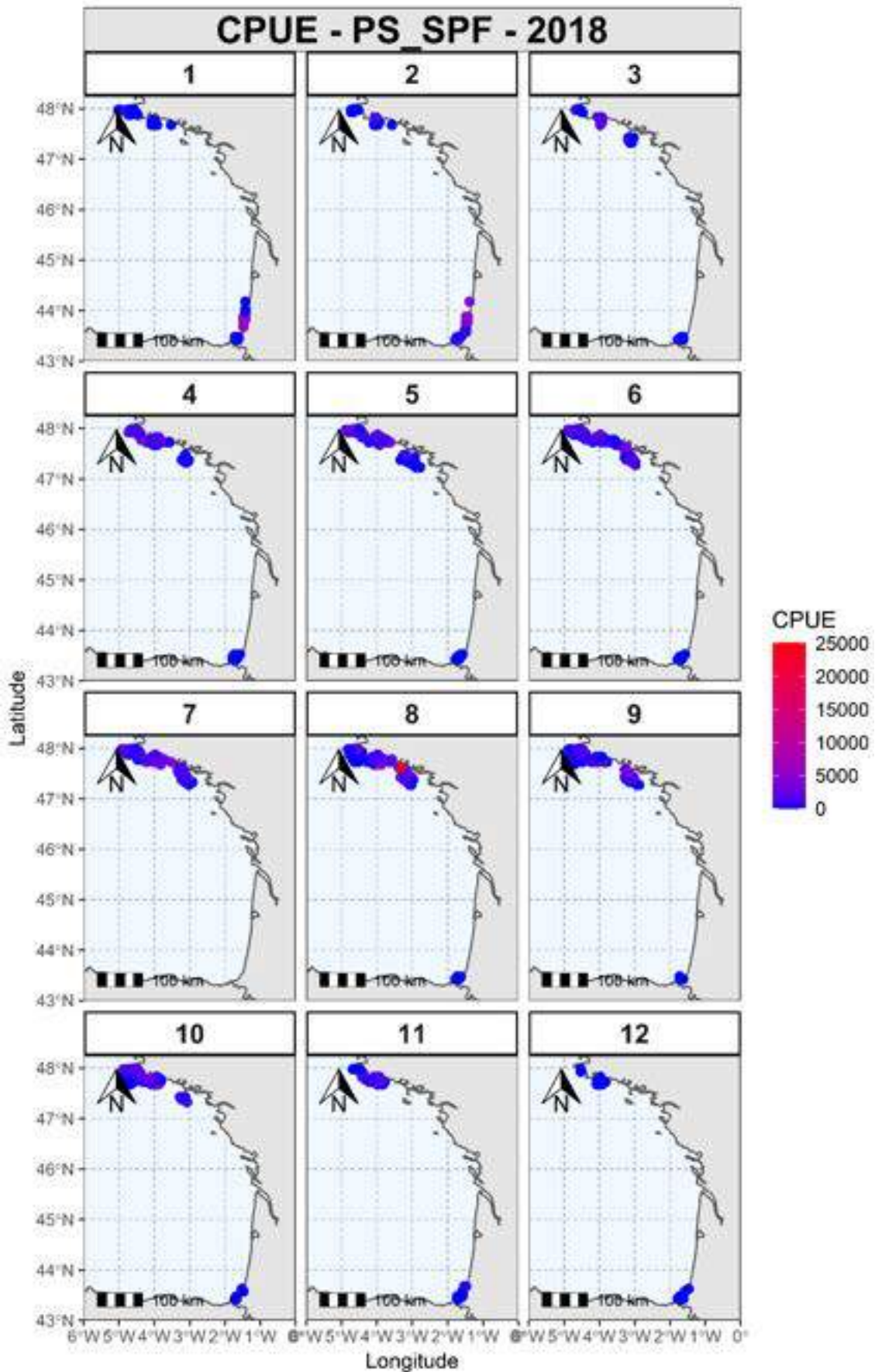
Annexe 1 : Analyse exploratoire des données commerciales : Spatialisation des CPUE en Mai 2018 de la flottille PTM_SPF.



Annexe 2 : Analyse exploratoire des données commerciales : Spatialisation des CPUE en Mai 2018 de la flottille PTM_DEF.



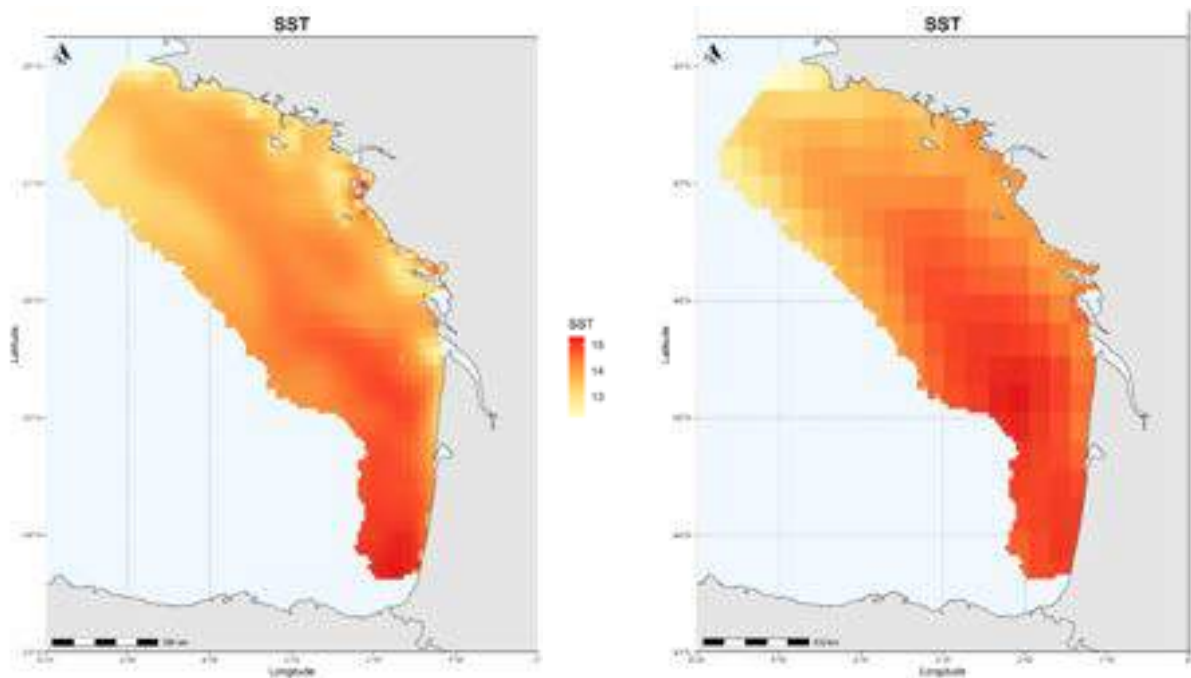
Annexe 3 : Analyse exploratoire des données commerciales : Spatialisation des CPUE en Mai 2018 de la flottille PS_SPF.



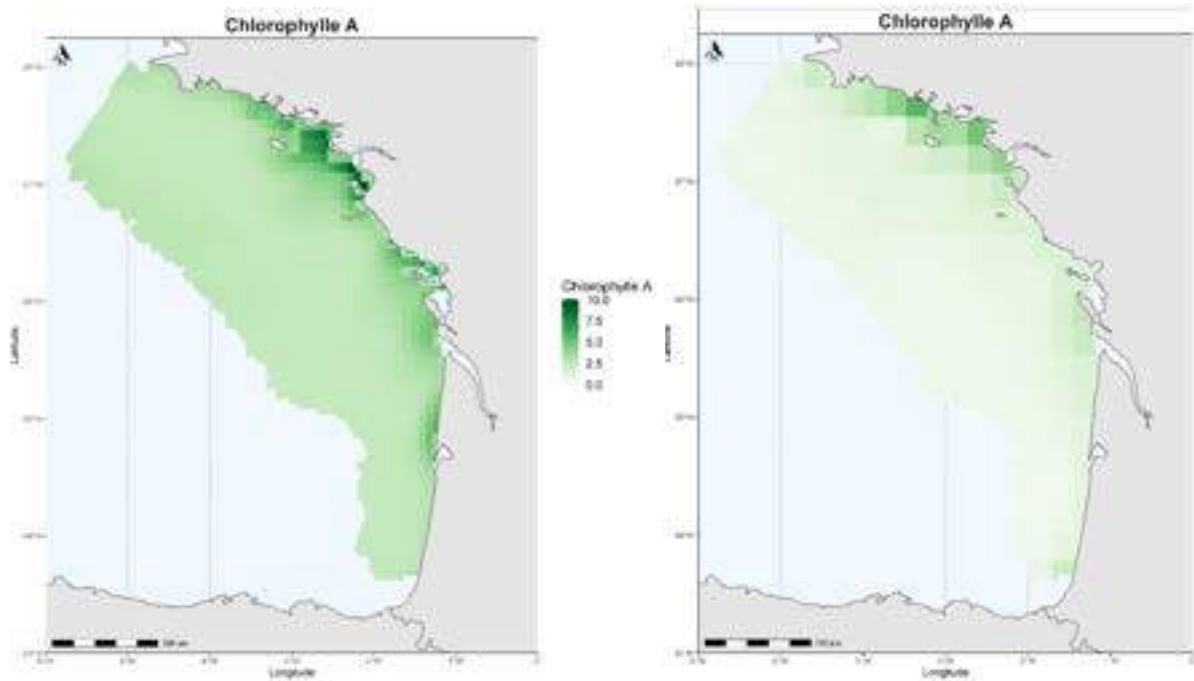
Annexe 4 : Séries temporelles des profils de débarquements entre 2008-2018 à un pas de temps mensuel pour les 3 flottilles commerciales.



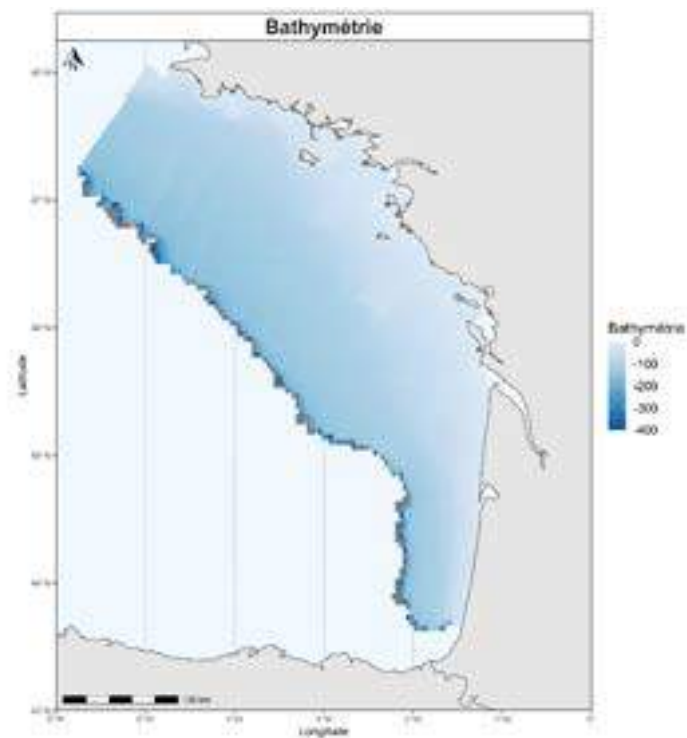
Annexe 5 : Cartographie de la SST en Mai 2018 (données issues du modèle biogéochimique POLCOM-ERSEM à gauche et satellitaires à droite).



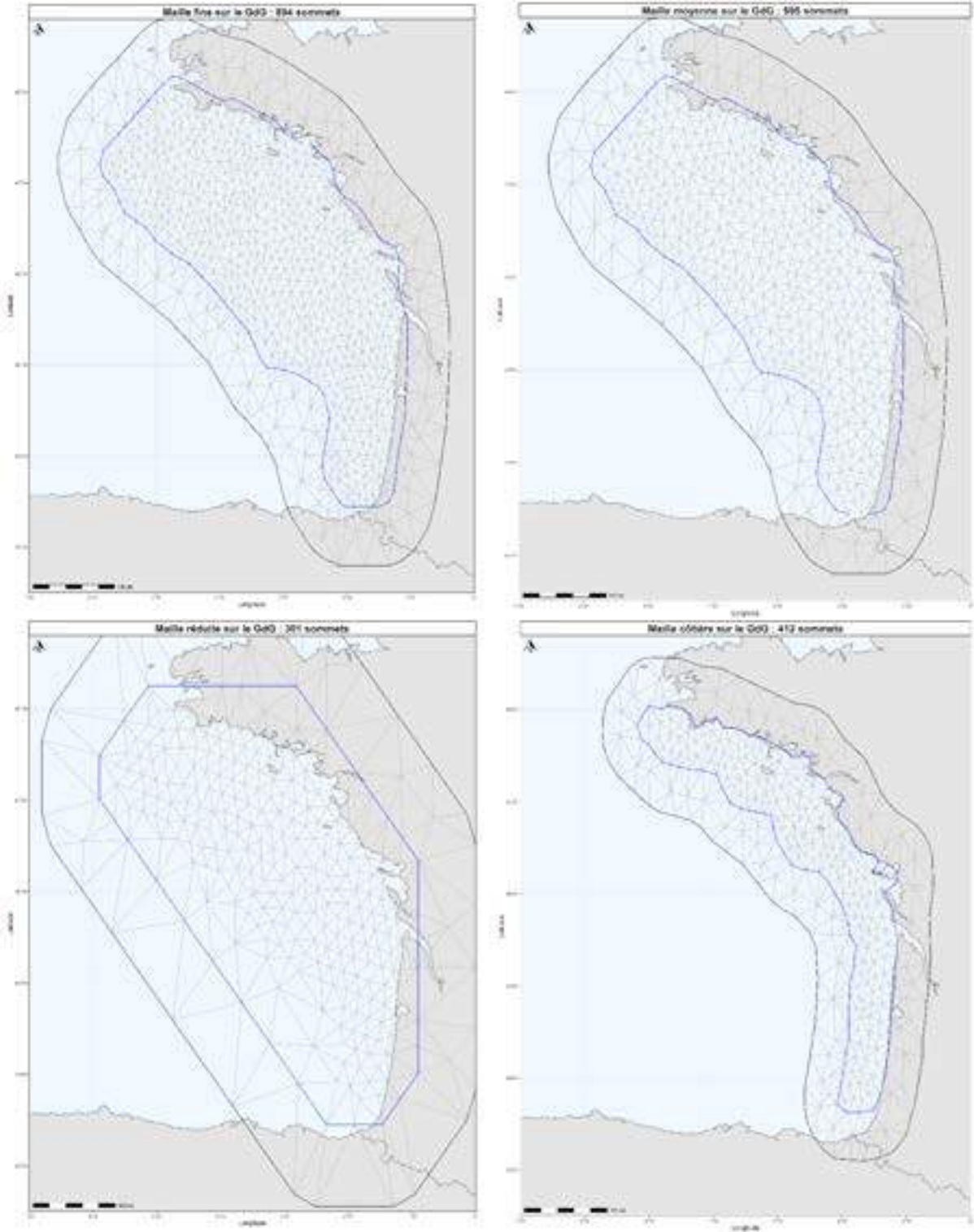
Annexe 6 : Cartographie de la chlorophylle A en Mai 2018 (données issues du modèle biogéochimique POLCOM-ERSEM à gauche et satellitaires à droite).



Annexe 7 : Cartographie de la bathymétrie en Mai 2018.



Annexe 8 : Maillages utilisés



Annexe 9 : Fonctionnement des PC priors et récapitulatif des paramétrisations testées

En suivant l'approche de Simpson *et al.* (2017), nous définissons les SPDE avec des PC priors (penalized complexity priors). La distribution des PC priors est unique pour chaque paramètre. Ils correspondent à la probabilité de s'écarter d'un modèle moins complexe. Pour le paramètre de la variance, le modèle moins complexe correspond à un $\sigma^2 = 0$, et pour le paramètre de distance celui-ci a une valeur infinie (soit dépendance parfaite i.e. moins de variabilité). La paramétrisation suit l'idée suivante : Soit τ un paramètre, nous définissons (u, α) tel que :

$$Prob(\tau > u) = \alpha, u > 0, 0 < \alpha < 1$$

La première représente la valeur du paramètre, la seconde est le quantile à partir duquel on dépasse la valeur choisie. Par exemple, nous posons pour la distance de corrélation qu'elle a 10% de chance d'être supérieure à 0.01°. Nous avons testé différentes paramétrisations afin d'étudier leur influence sur l'inférence du champ latent et la relation espèce habitat :

Tableau 1 : Paramétrisations testées pour l'effet aléatoire.

Hyperparamètre	Valeurs	Quantile
Distance de corrélation (r)	0.01	0.1
	2	0.1
Ecart-type (σ^2)	100	0.1
	1	0.1

Tableau 2 : Paramétrisations testées pour la relation espèce habitat.

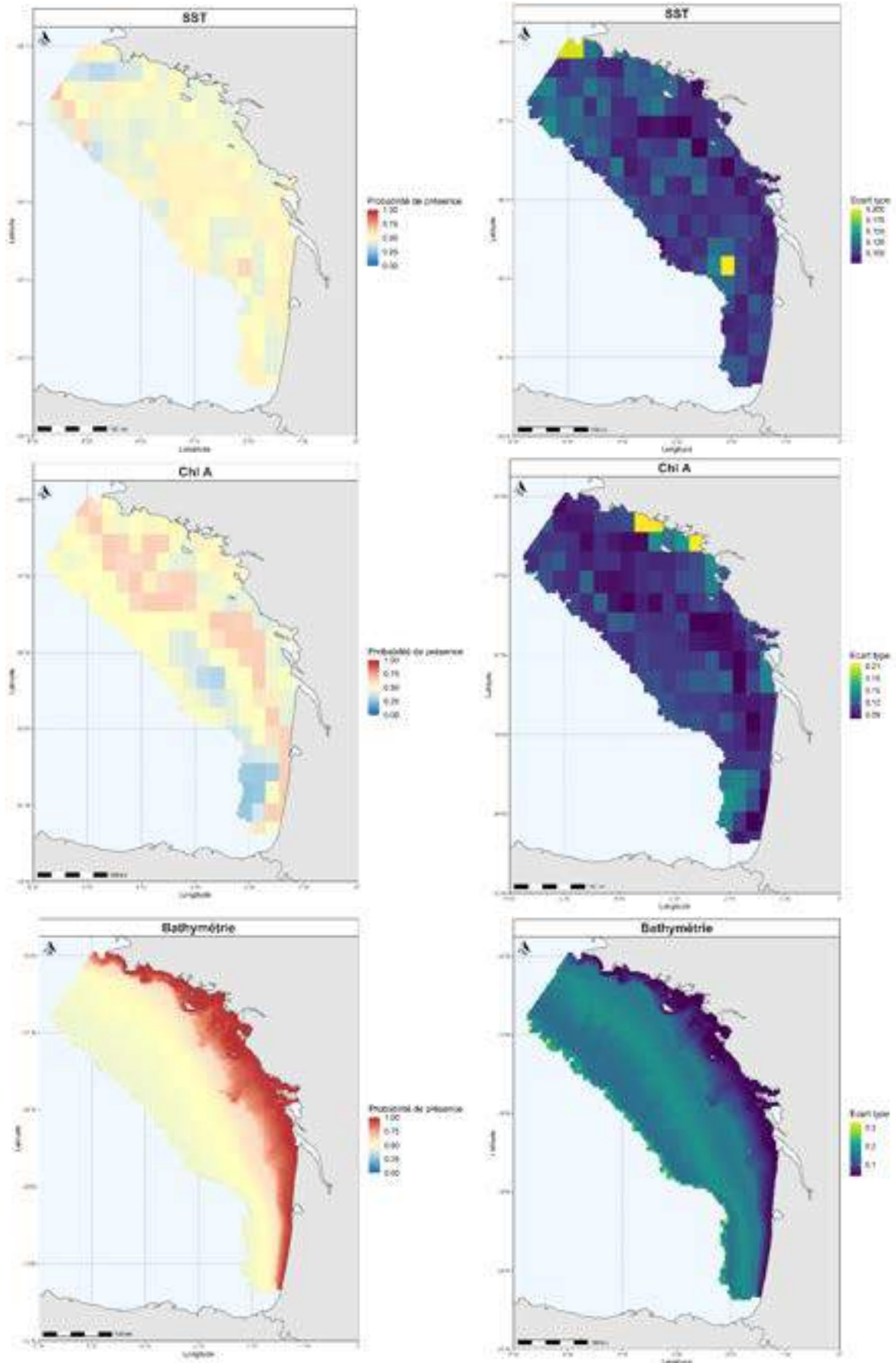
Hyperparamètre	Valeurs	Quantile
Distance de corrélation (r)	0.01	0.1
	2	0.1
Ecart-type (σ^2)	100	0.1
	2	0.1

En parallèle, les priors des effets fixes sont par défaut des logGamma de paramétrisations (1, 5^{e-5}).

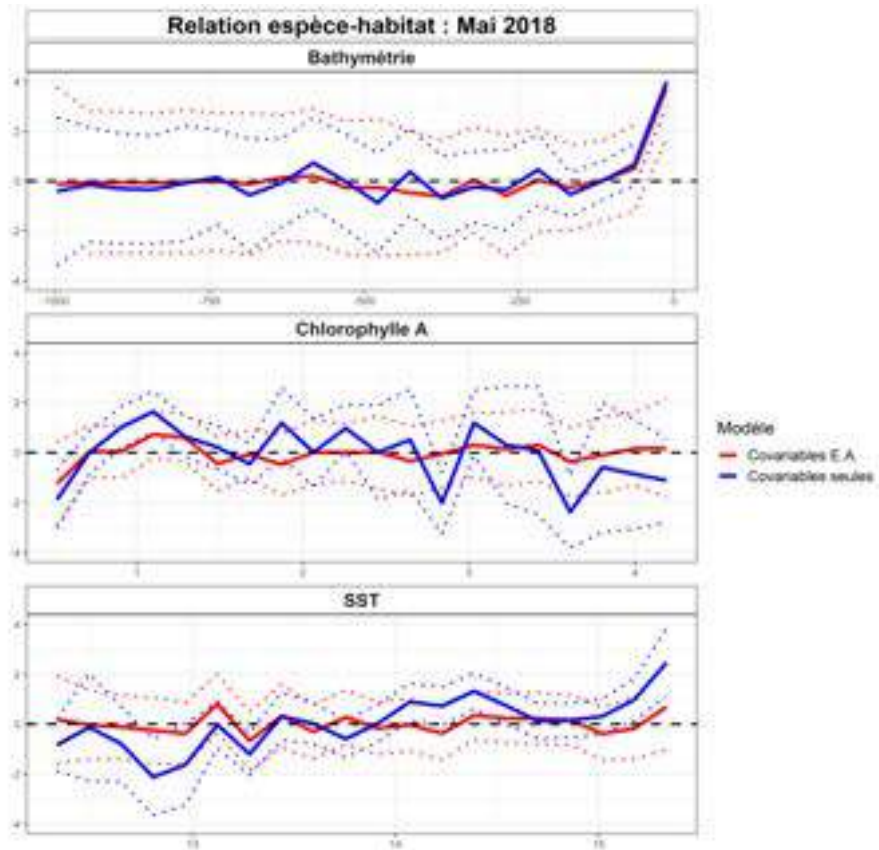
Annexe 10 : Tableau récapitulatif des différentes modélisations testées (toutes les combinaisons n'ont pas été testées).

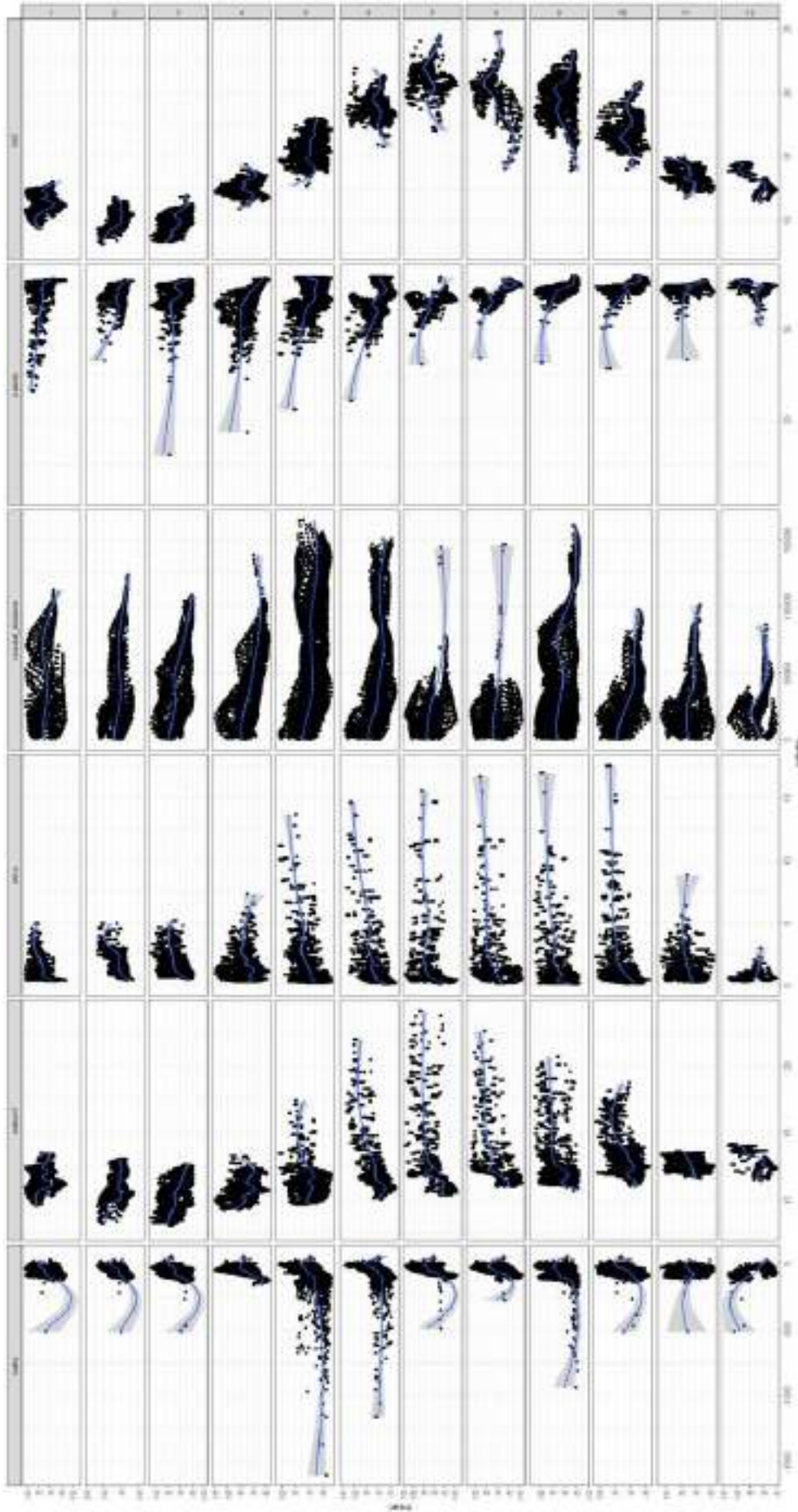
Modèle	Zone d'étude	Maille	Sources de données
Donnée seule	GdG	Réduite	2 survey (PELGAS + JUVENA) + 2 flottilles commerciales (PS + PTM)
Intégré	Frange côtière (50km)	Moyenne	2 survey + 3 flottilles commerciales (PS_SPF + PTM_SPF + PTM_DEF)
Ech. Préférentiel		Fine	
Co-variables		Côtière	
Intégré annuel			

Annexe 11 : Inférences obtenues pour chaque variables environnementale (SST, chlorophylle A, bathymétrie).



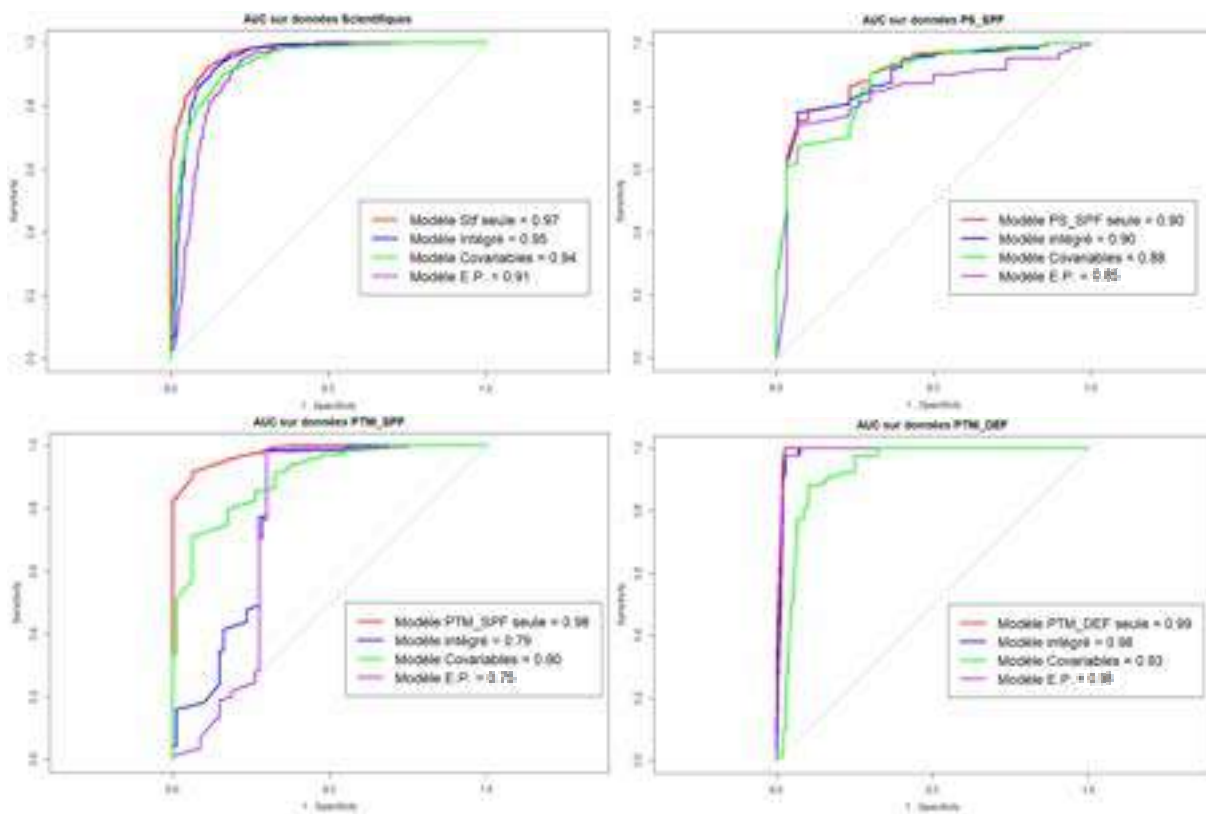
Annexe 12 : Comparaison effet des conditions environnementaux avec ou non une prise en compte de l'effet aléatoire (données satellitaires).



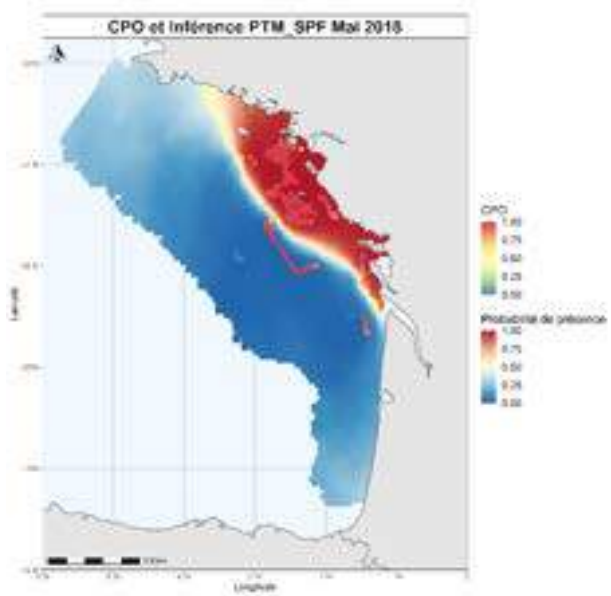
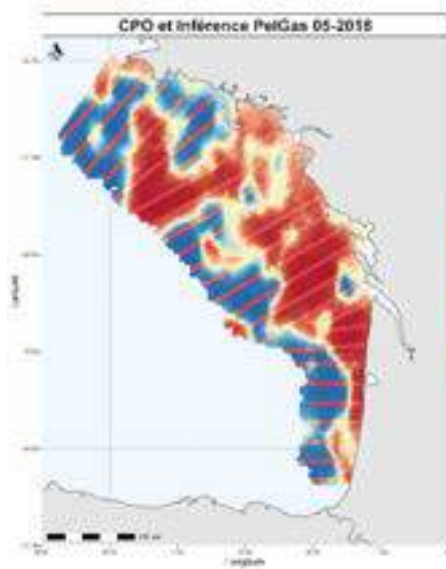


Annexe 13 : Estimation du champ aléatoire de présence-absence en fonction des co-variables environnementales par mois en 2018 (bathymétrie, température de fond, distance à la côte, chlorophylle a, température de surface)

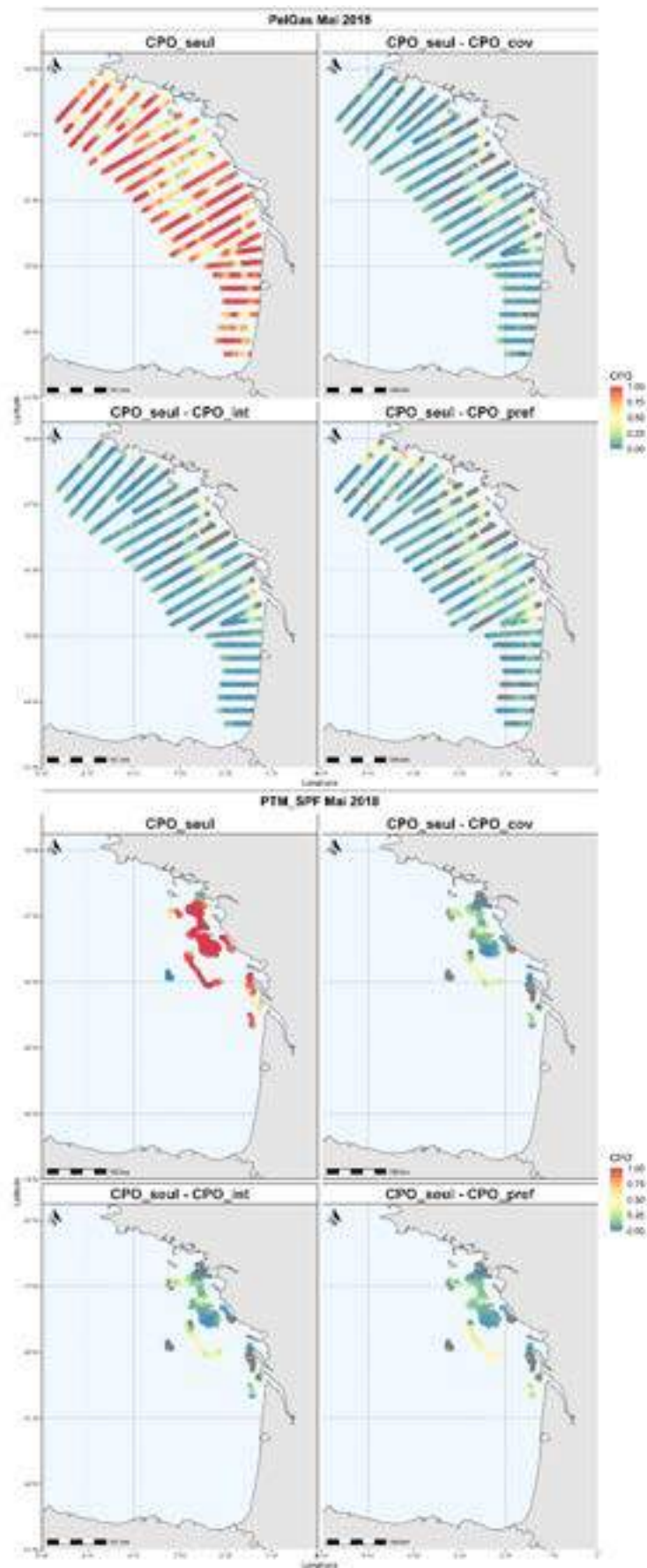
Annexe 14 : Courbes ROC et AUC pour chaque source de données et modèles.

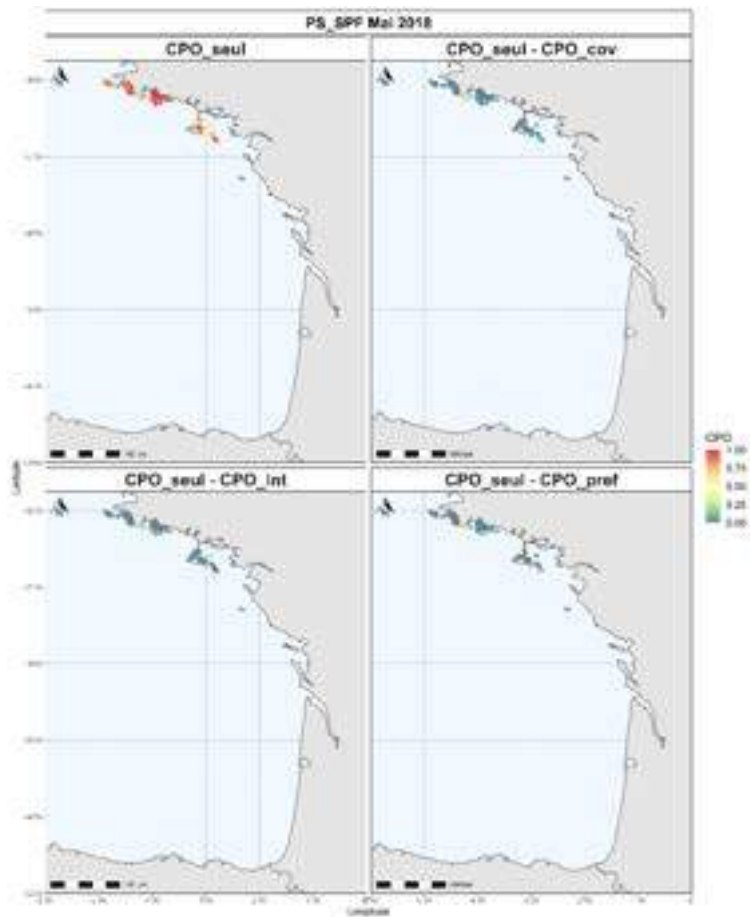
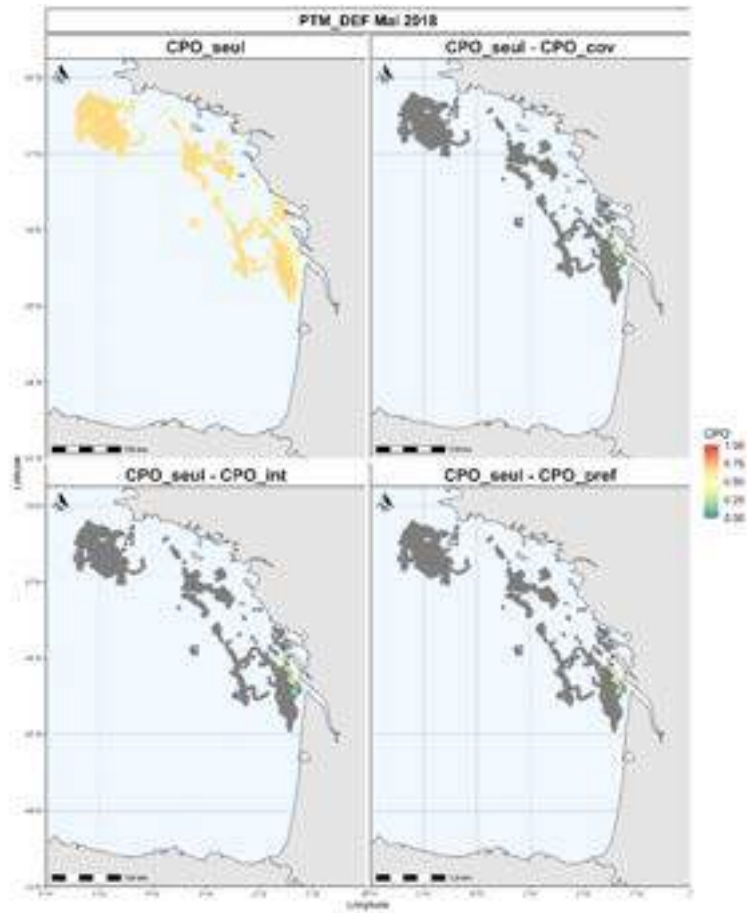


Annexe 15 : Cartographies des CPO pour chaque source de donnée issue du modèle « donnée seule ». Les CPO sont représentées sur les cartes d'inférences respectives.

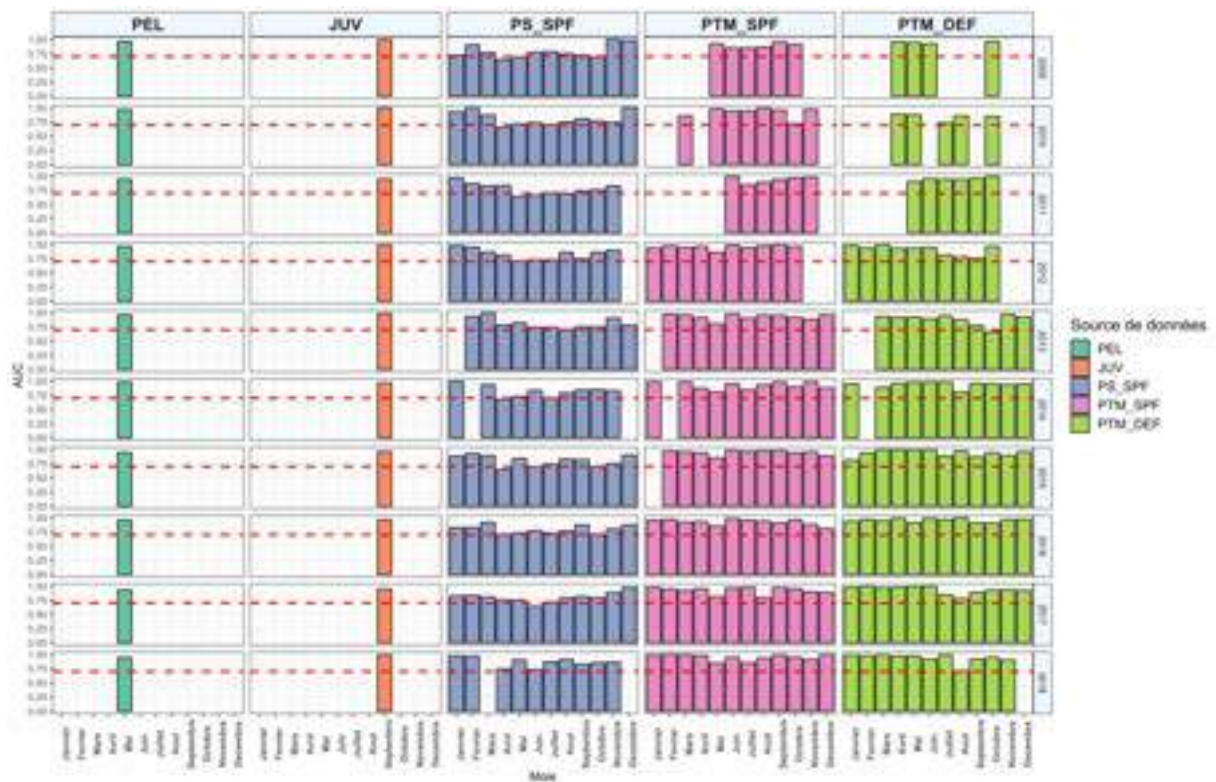


Annexe 16 : Cartographies des CPO pour chaque source de donnée et issue du modèle « donnée seule », et différence avec la CPO de ce modèle.

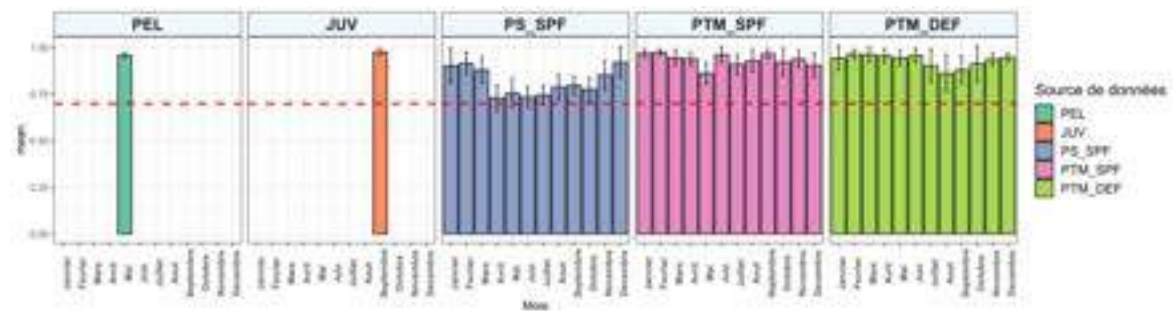




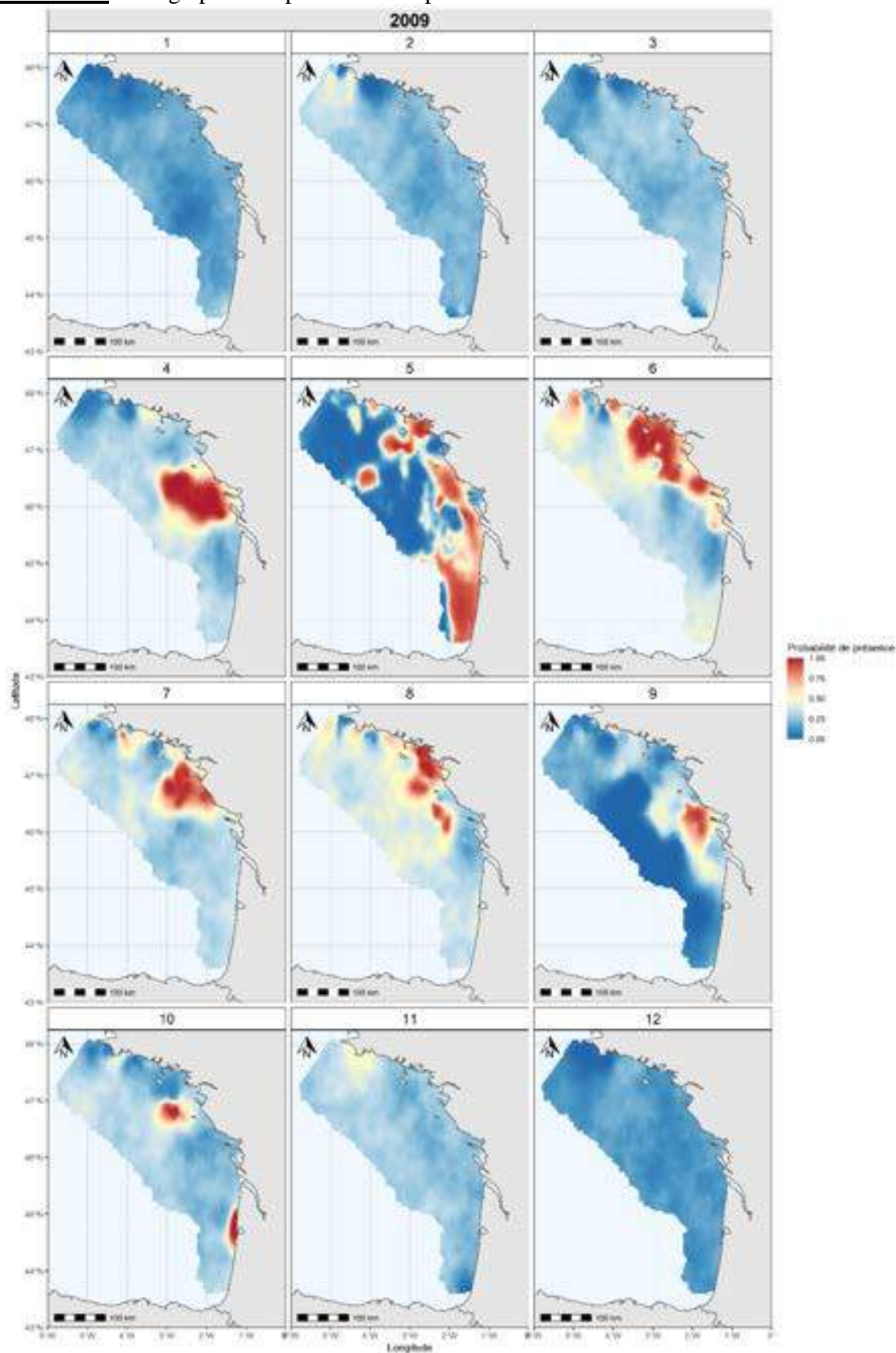
Annexe 17 : Scores AUC des modèles spatio-temporels par mois et source de donnée entre 2009 et 2018. Le trait rouge correspond à une AUC de 0.70



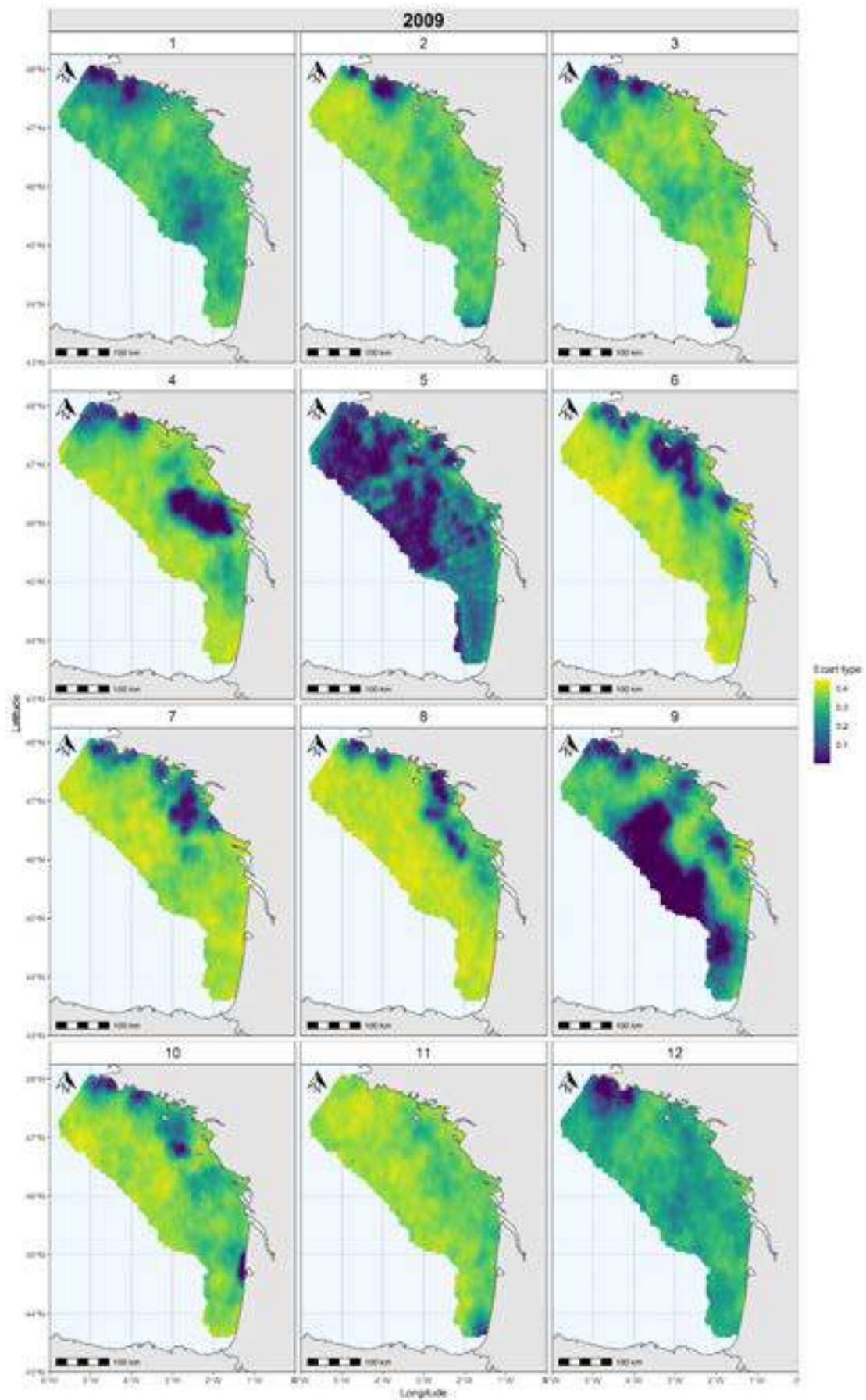
Annexe 18 : AUC moyenné par mois et source de donnée (2009-2018) et écarts types. Le trait rouge correspond à une AUC de 0.70.



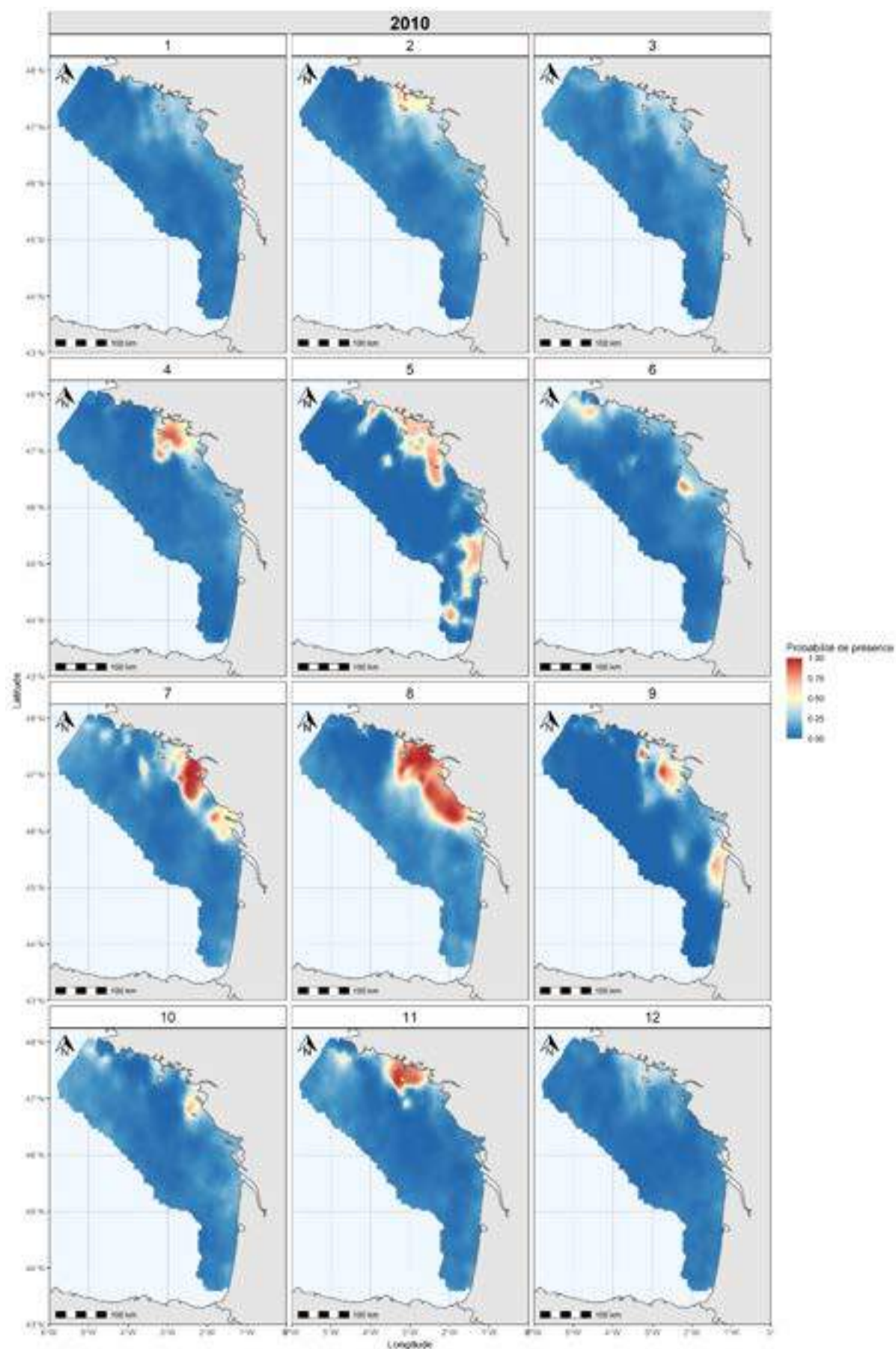
Annexe 19 : Cartographies de probabilité de présence de la sardine dans le GdG en 2009.



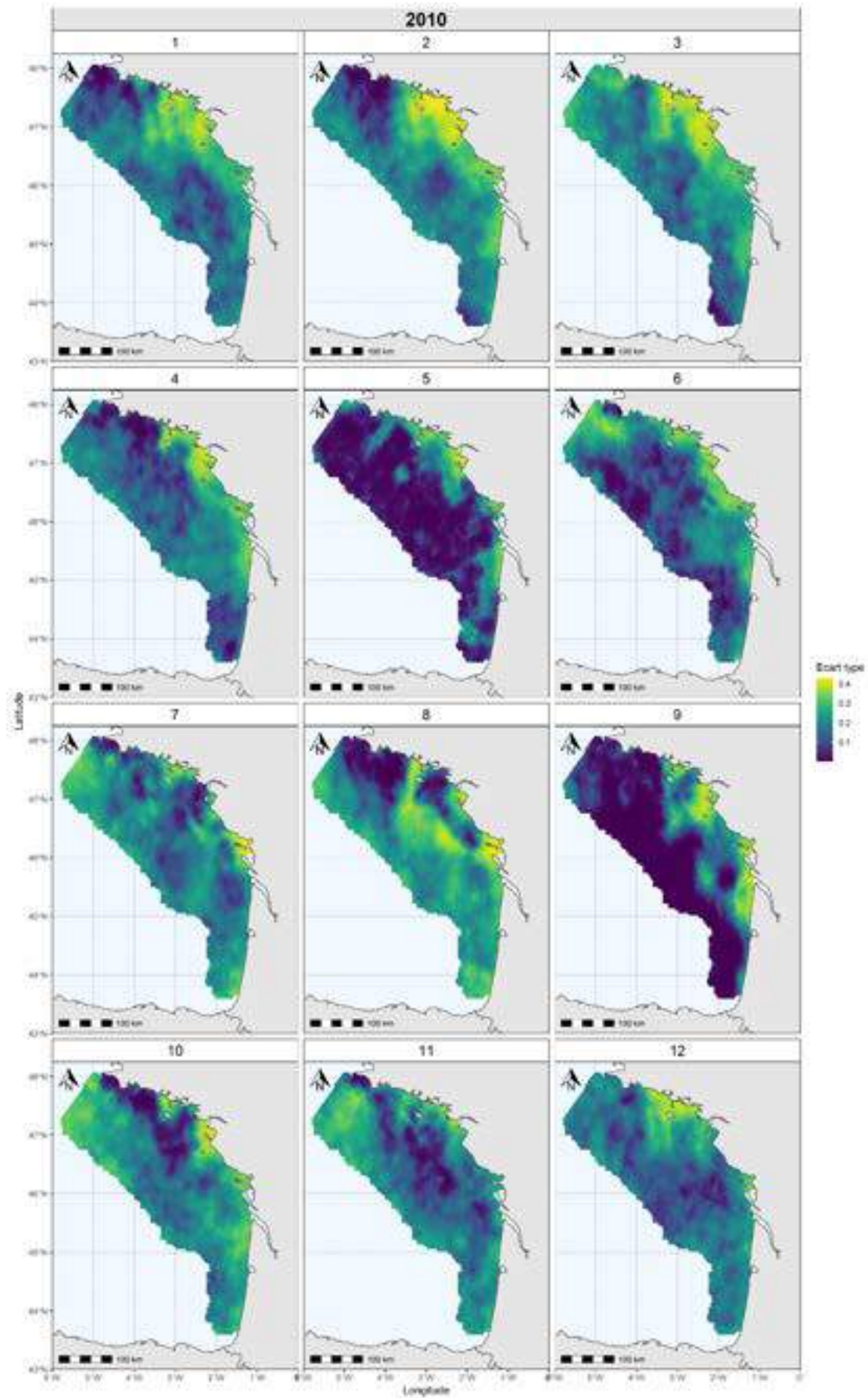
Annexe 20 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2009.



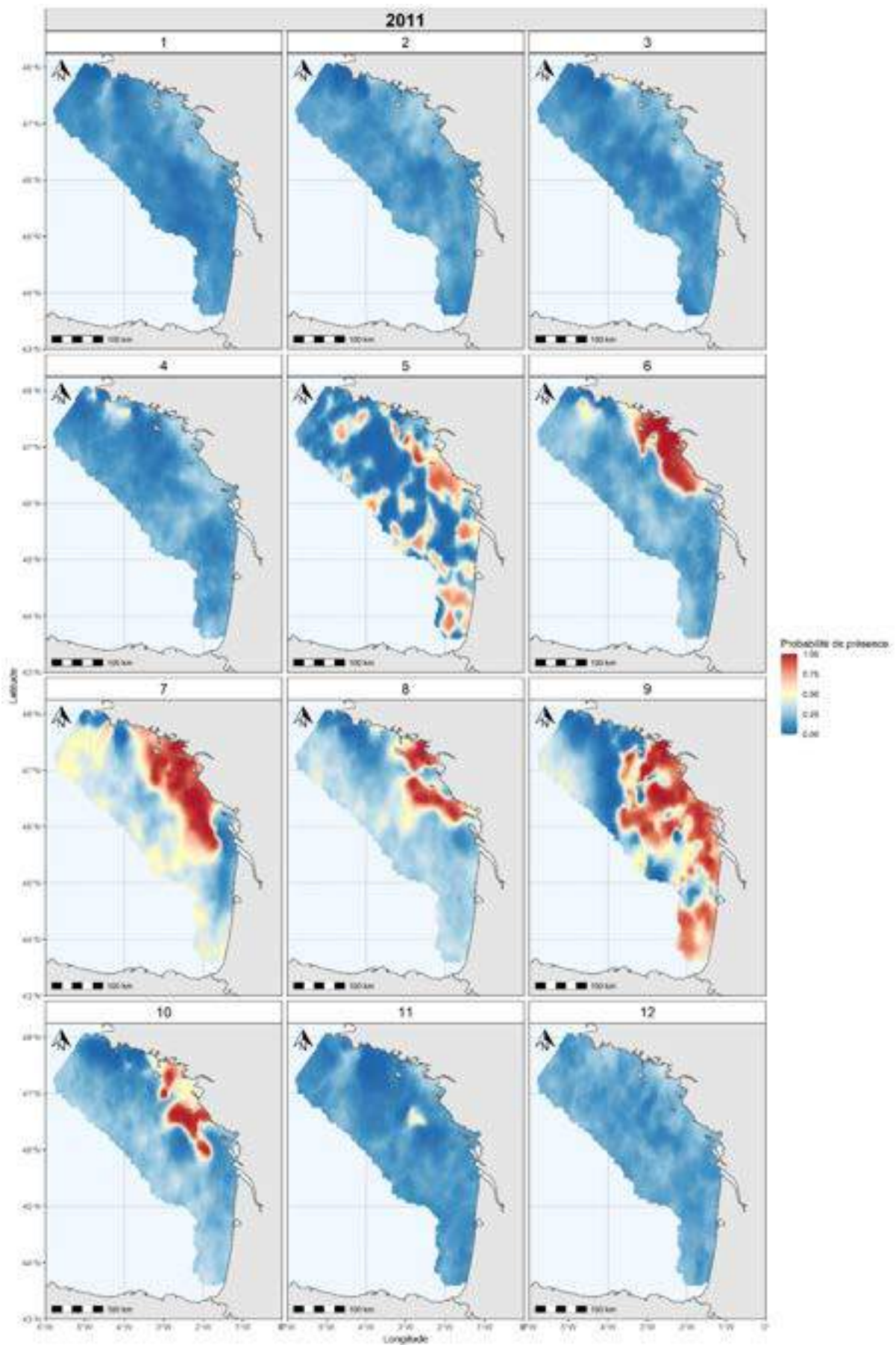
Annexe 21 : Cartographies de probabilité de présence de la sardine dans le GdG en 2010.



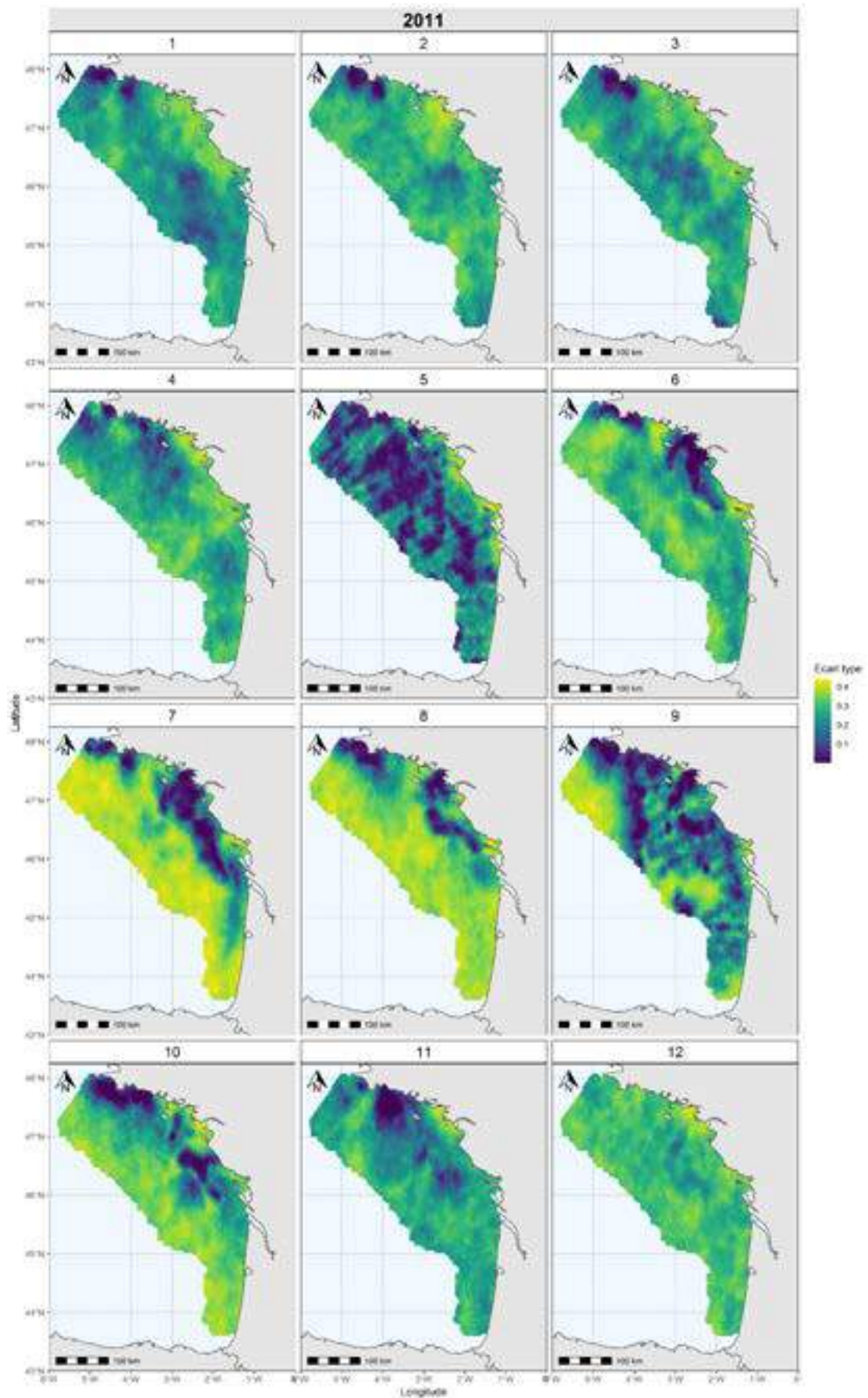
Annexe 22 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2010.



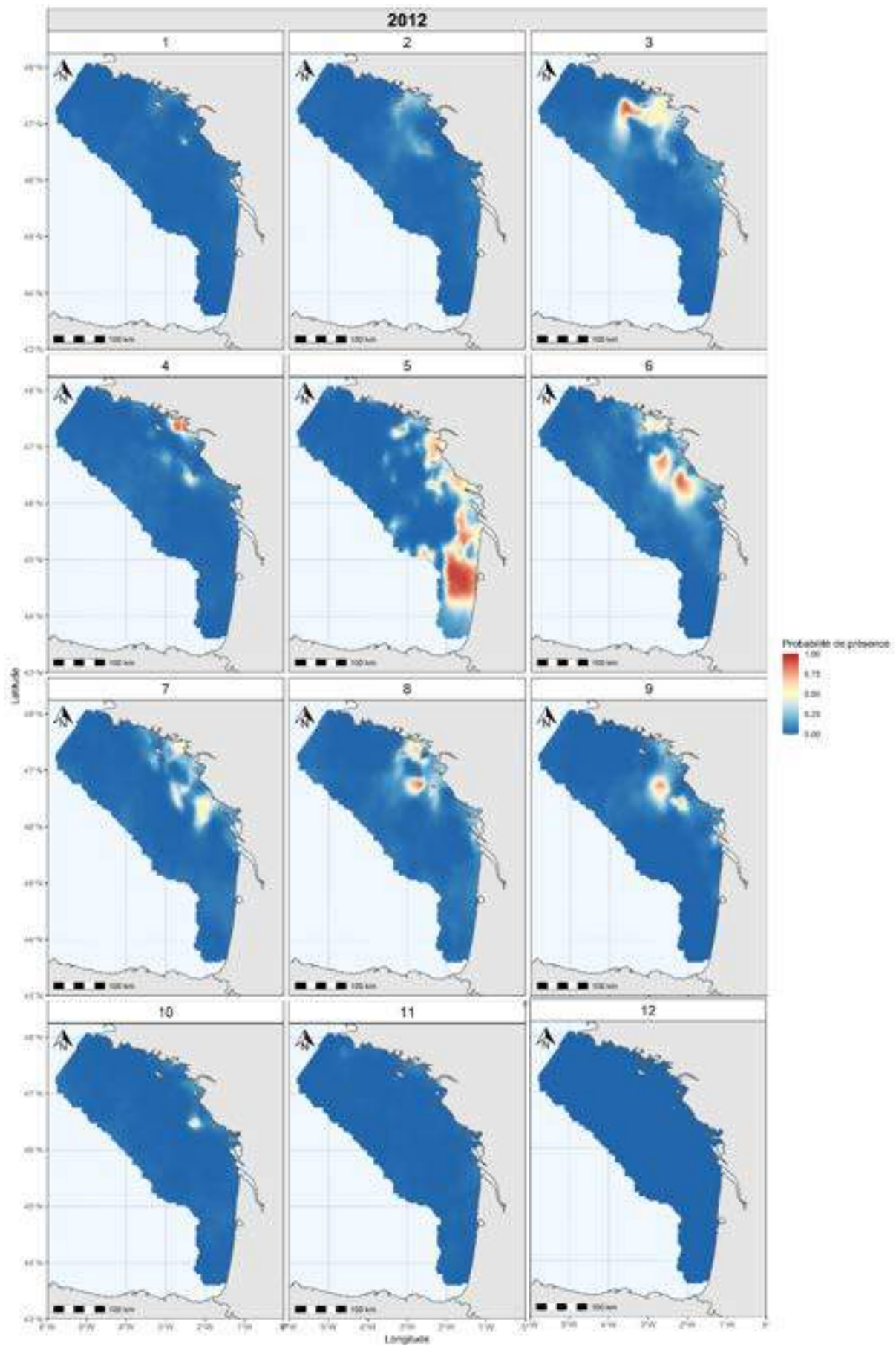
Annexe 23 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2011.



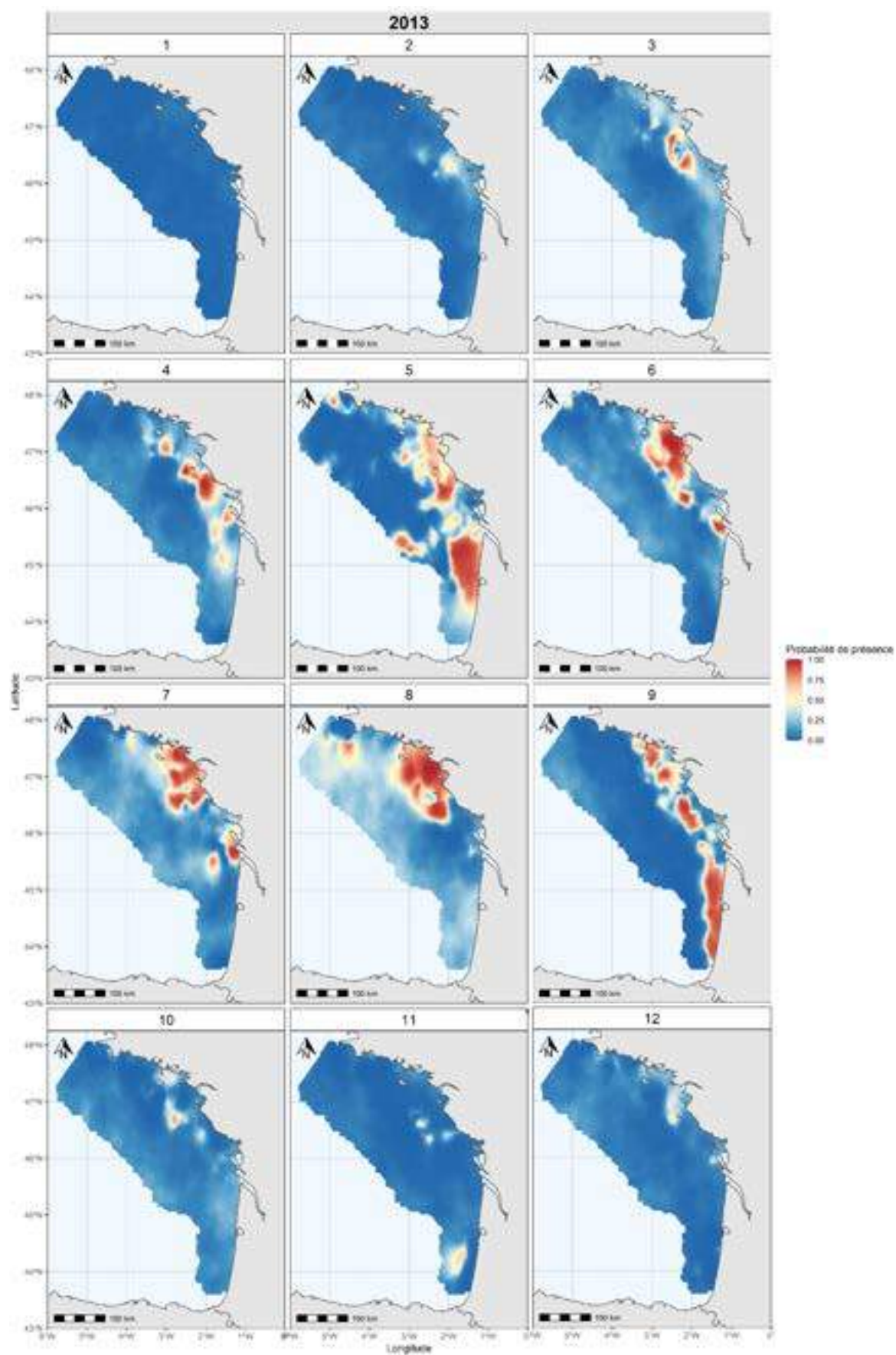
Annexe 24 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2011.



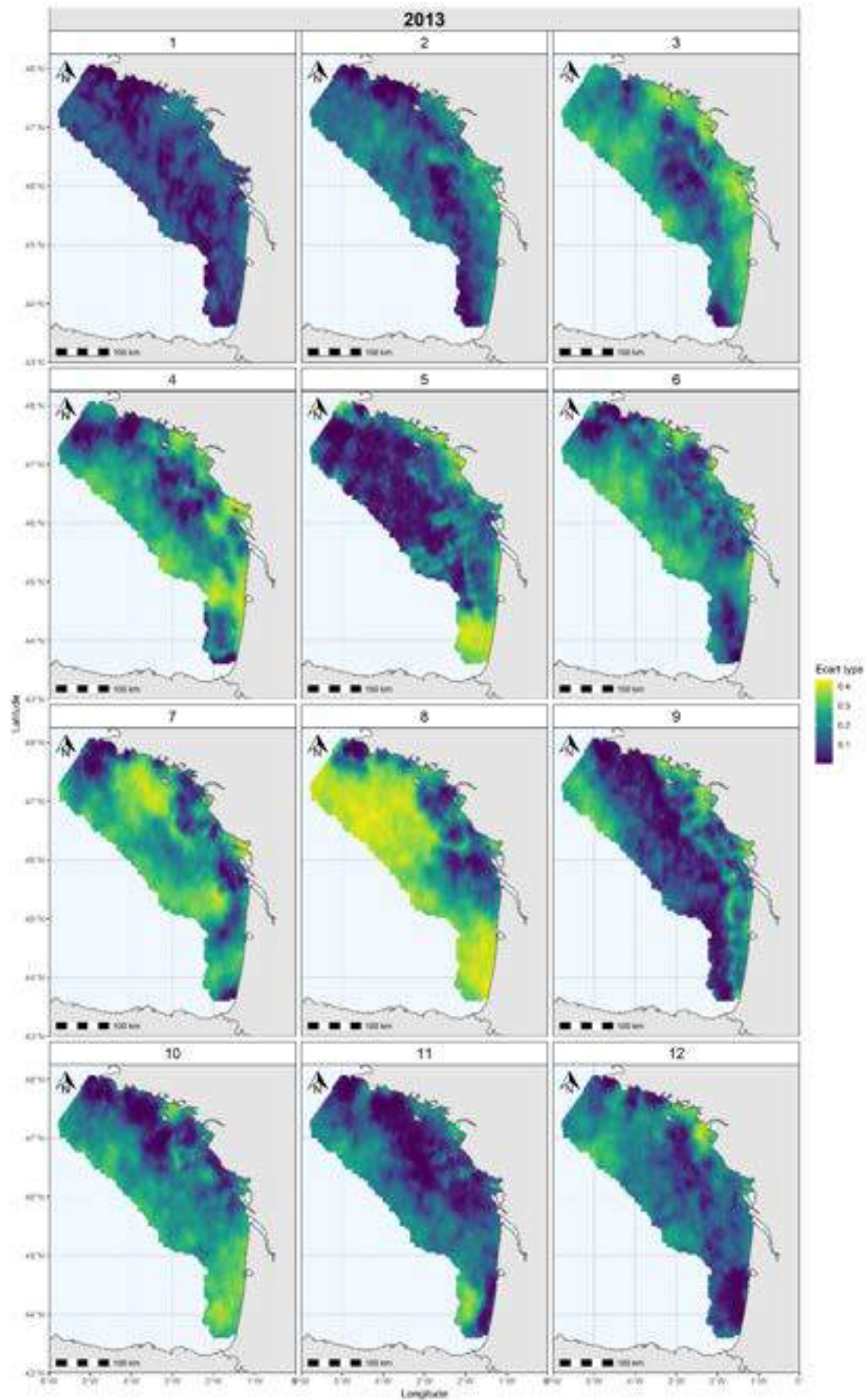
Annexe 25 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2012.



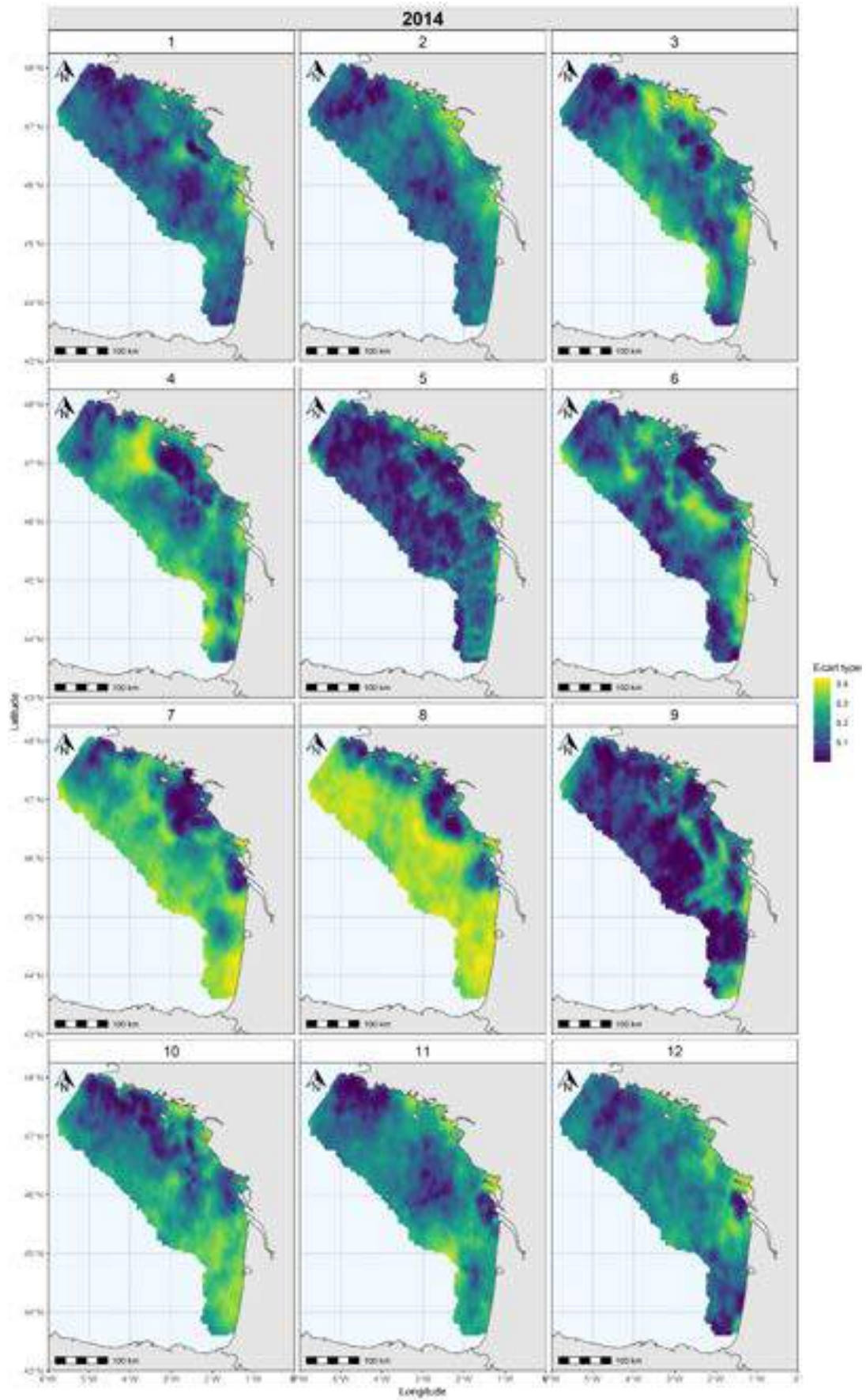
Annexe 27 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2013.



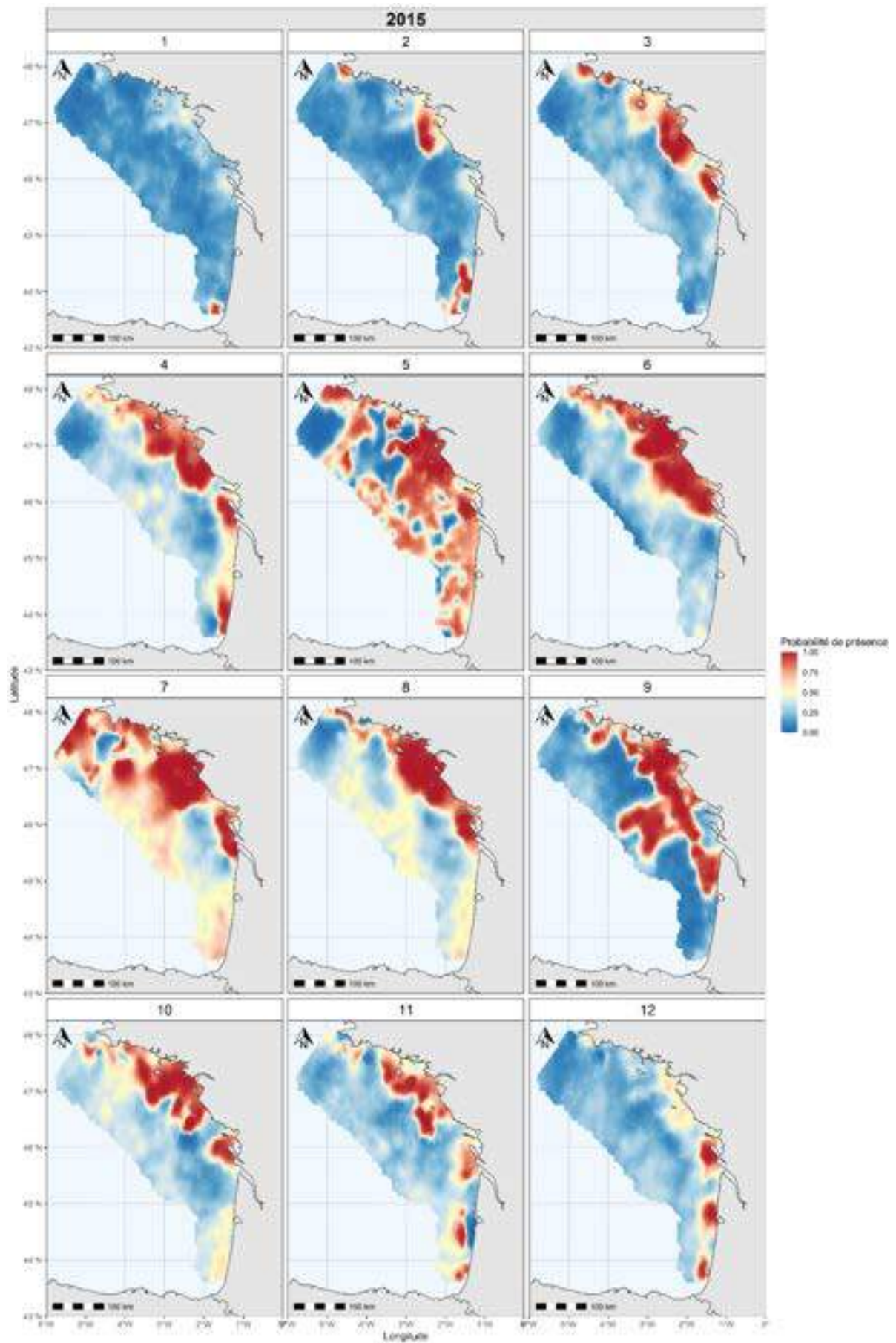
Annexe 28 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2013.



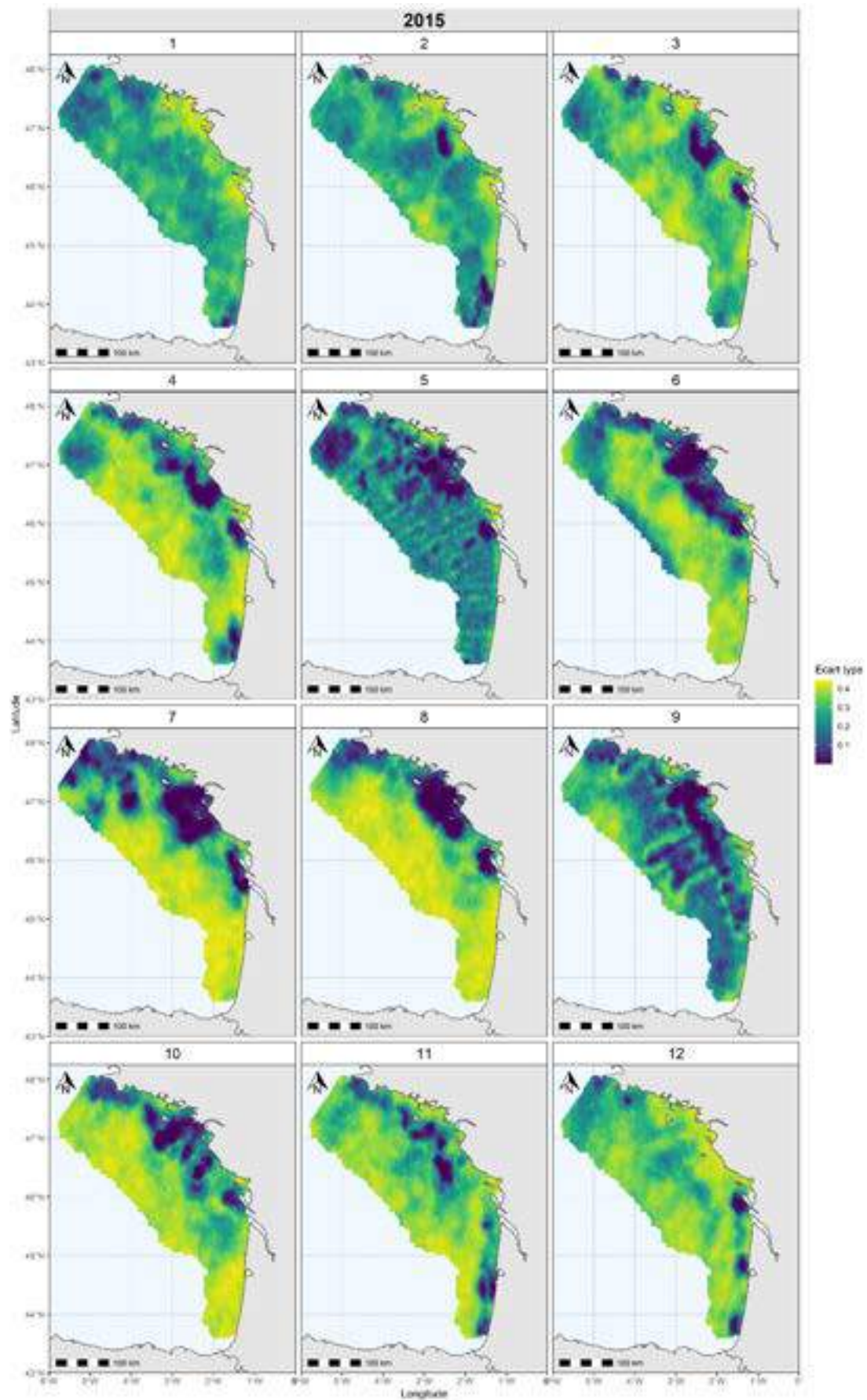
Annexe 30 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2014.



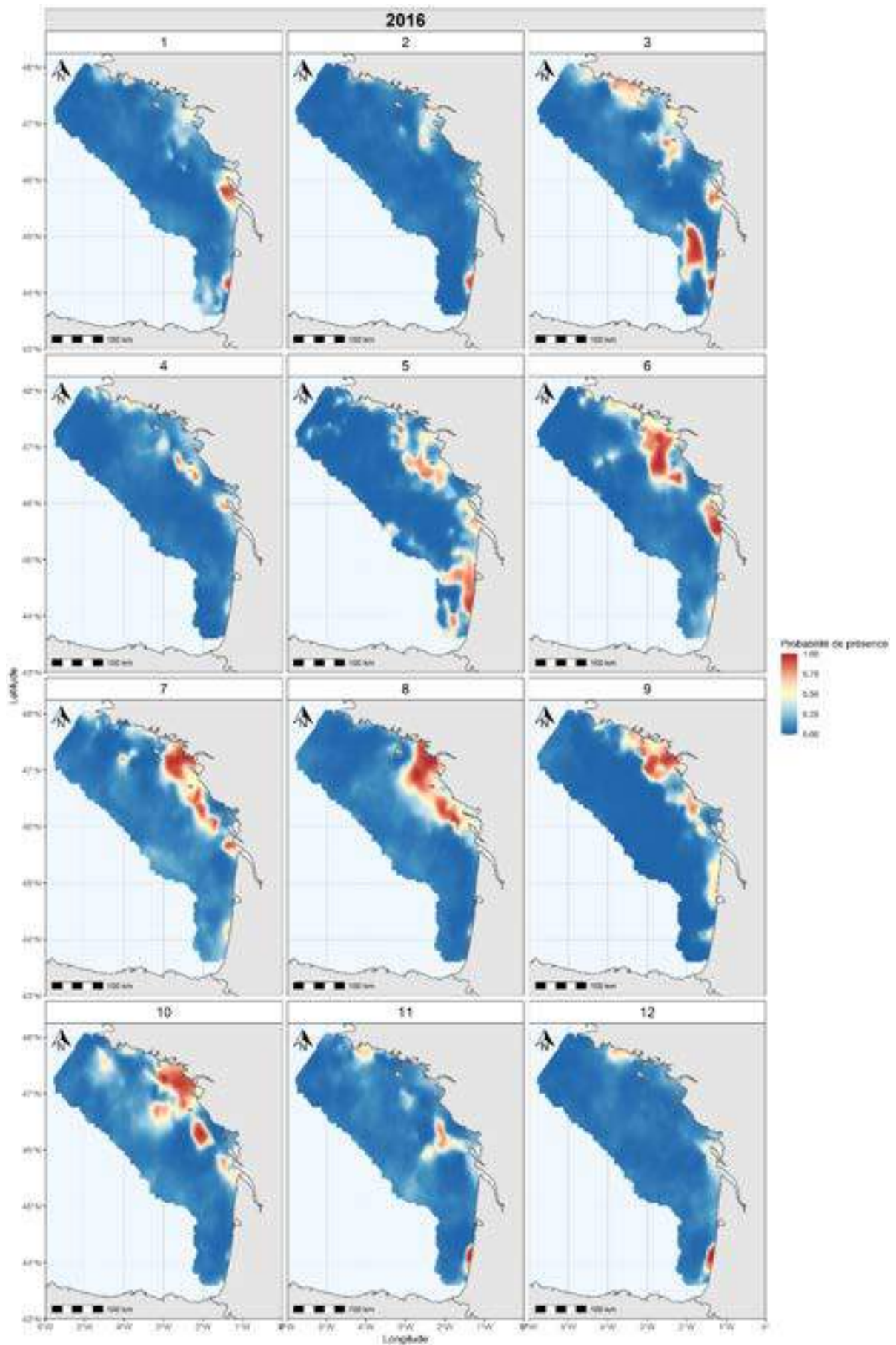
Annexe 31 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2015.



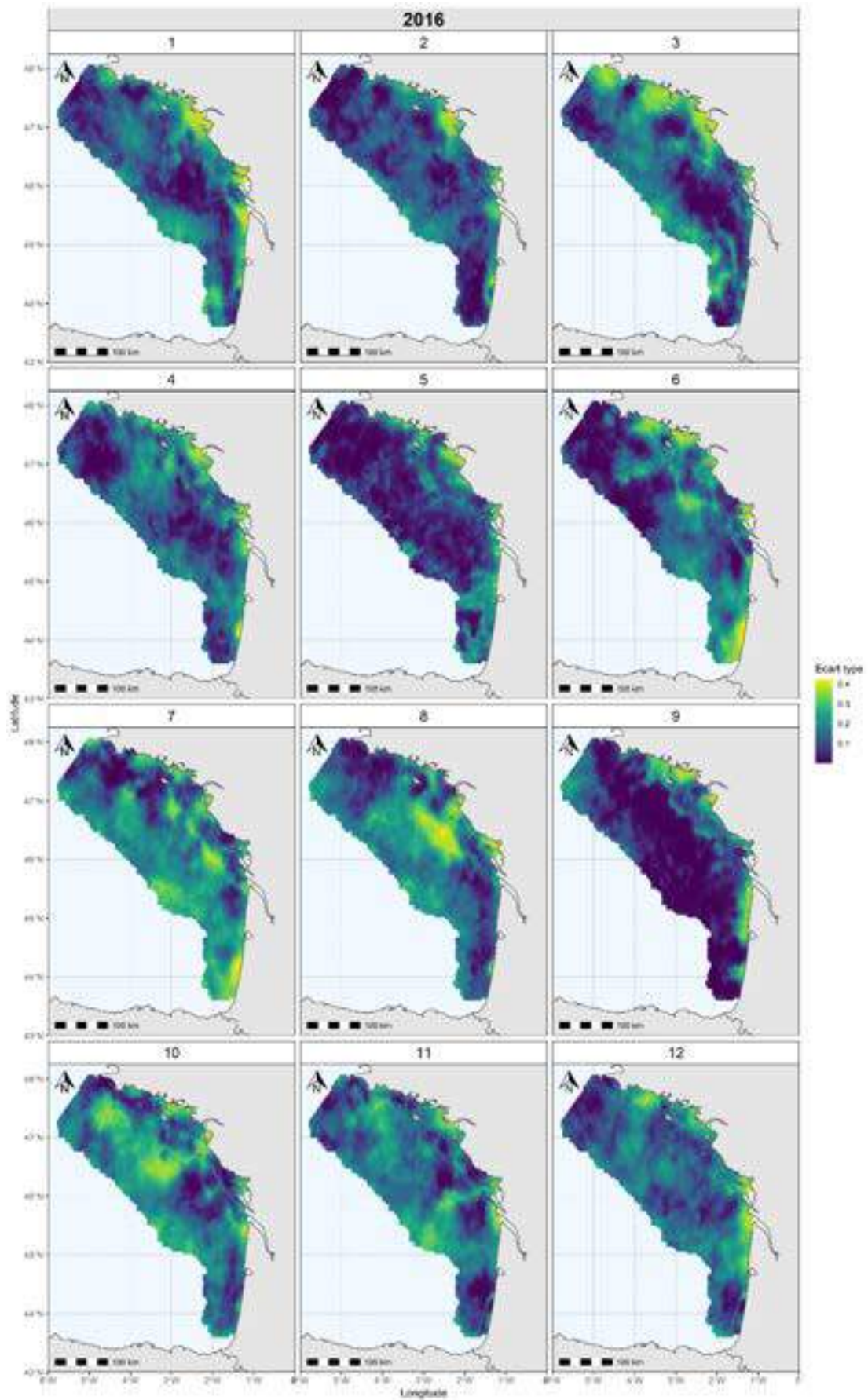
Annexe 32 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2015.



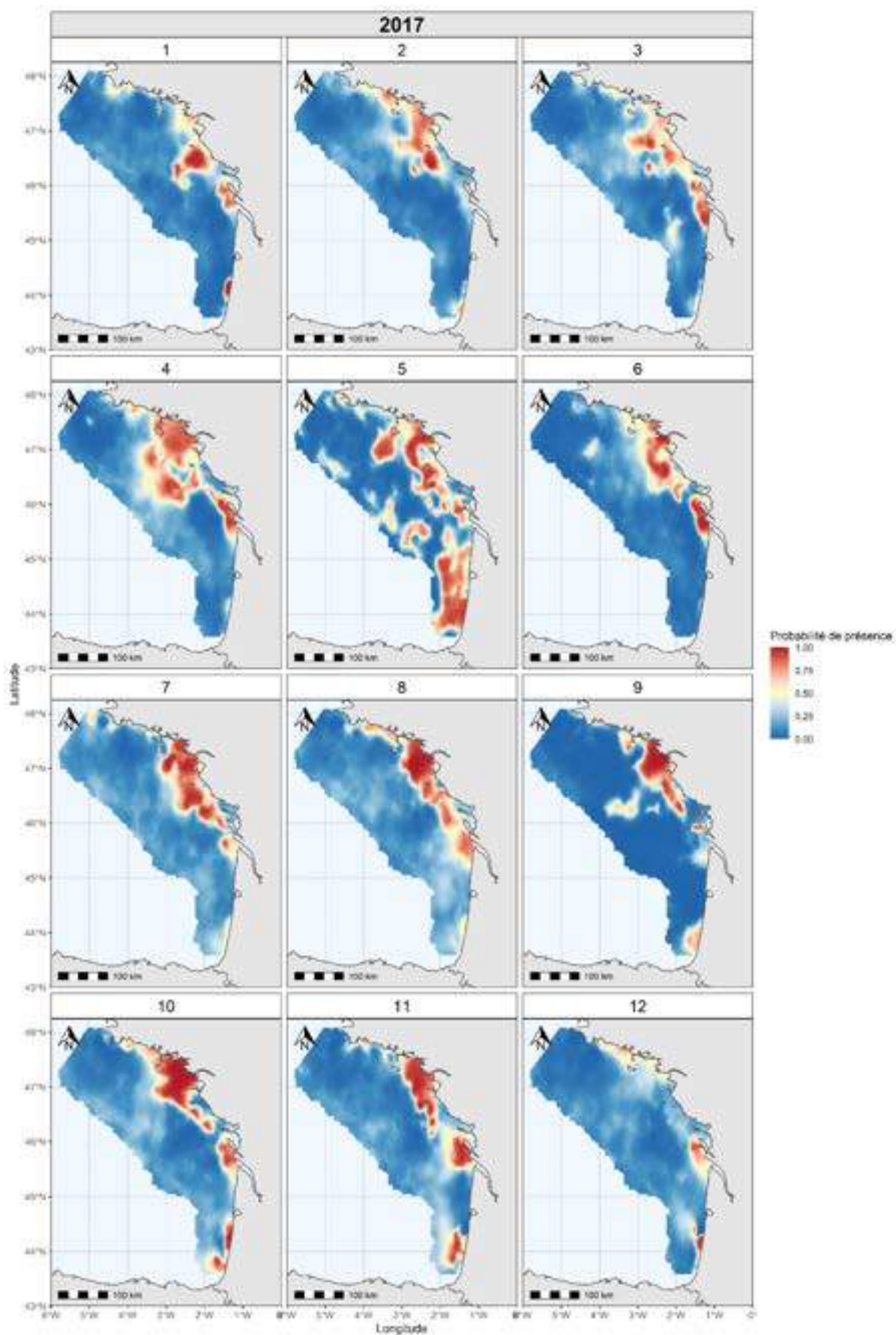
Annexe 33 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2016.



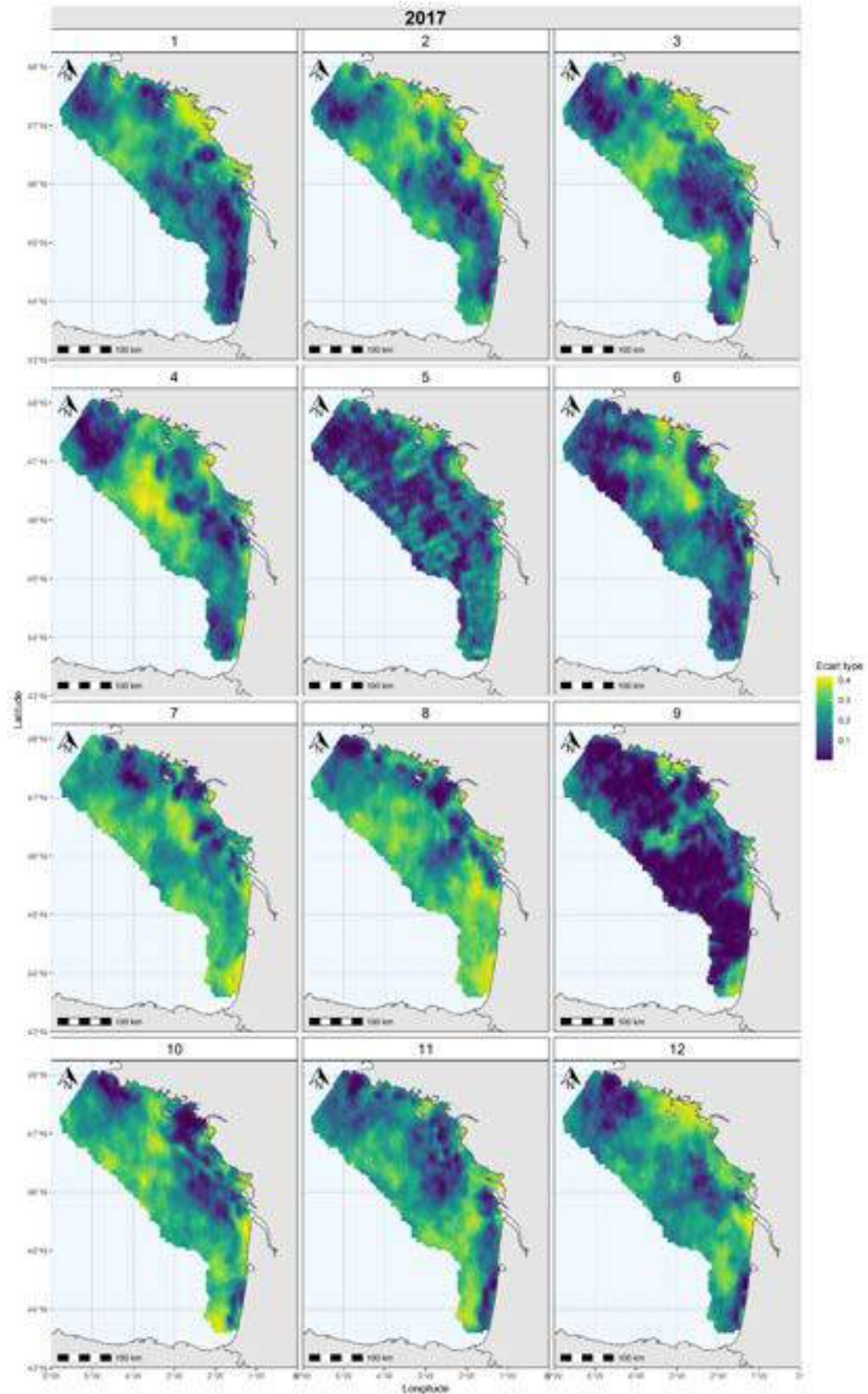
Annexe 34 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2016.



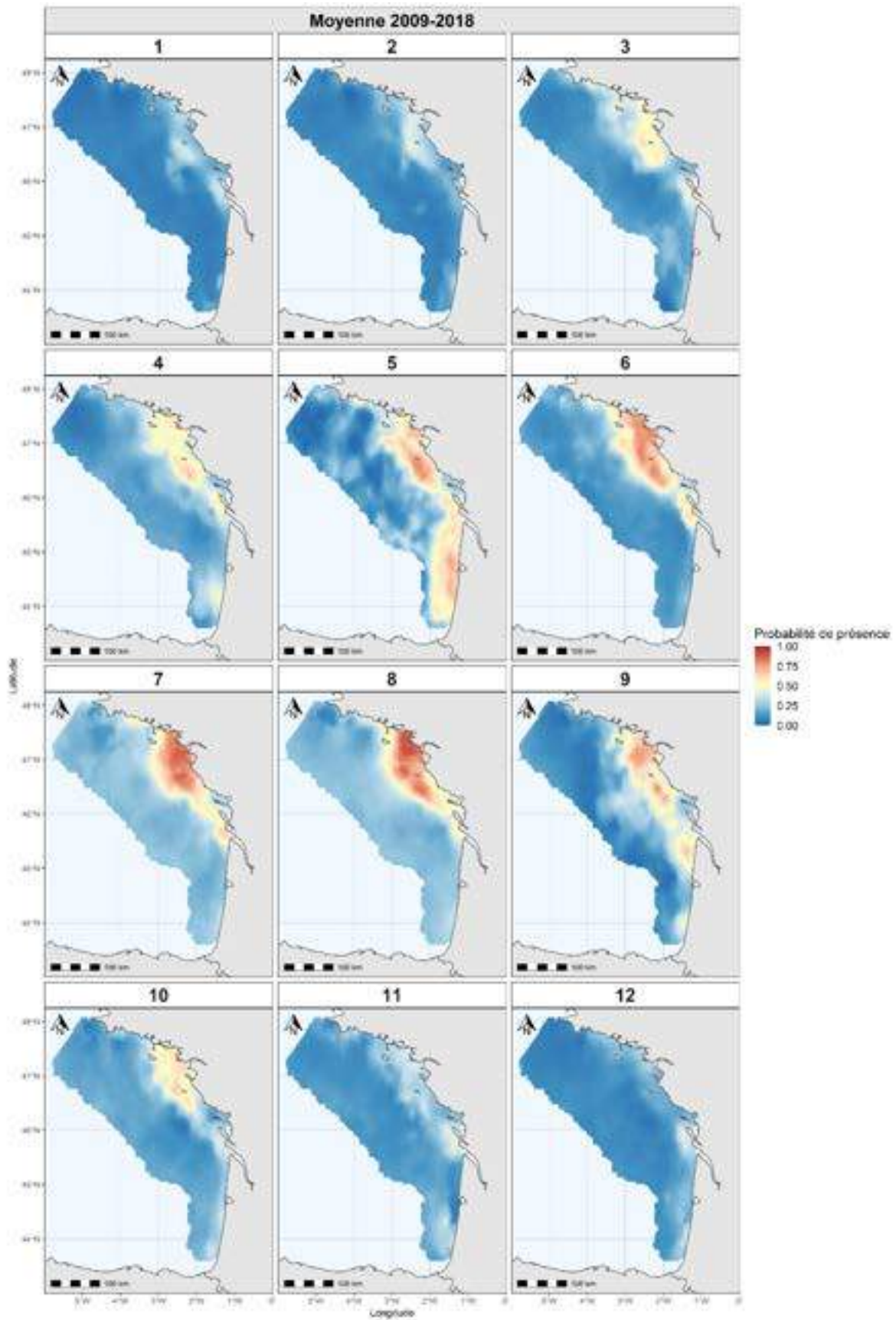
Annexe 35 : Cartographies de la probabilité de présence de la sardine dans le GdG en 2017.



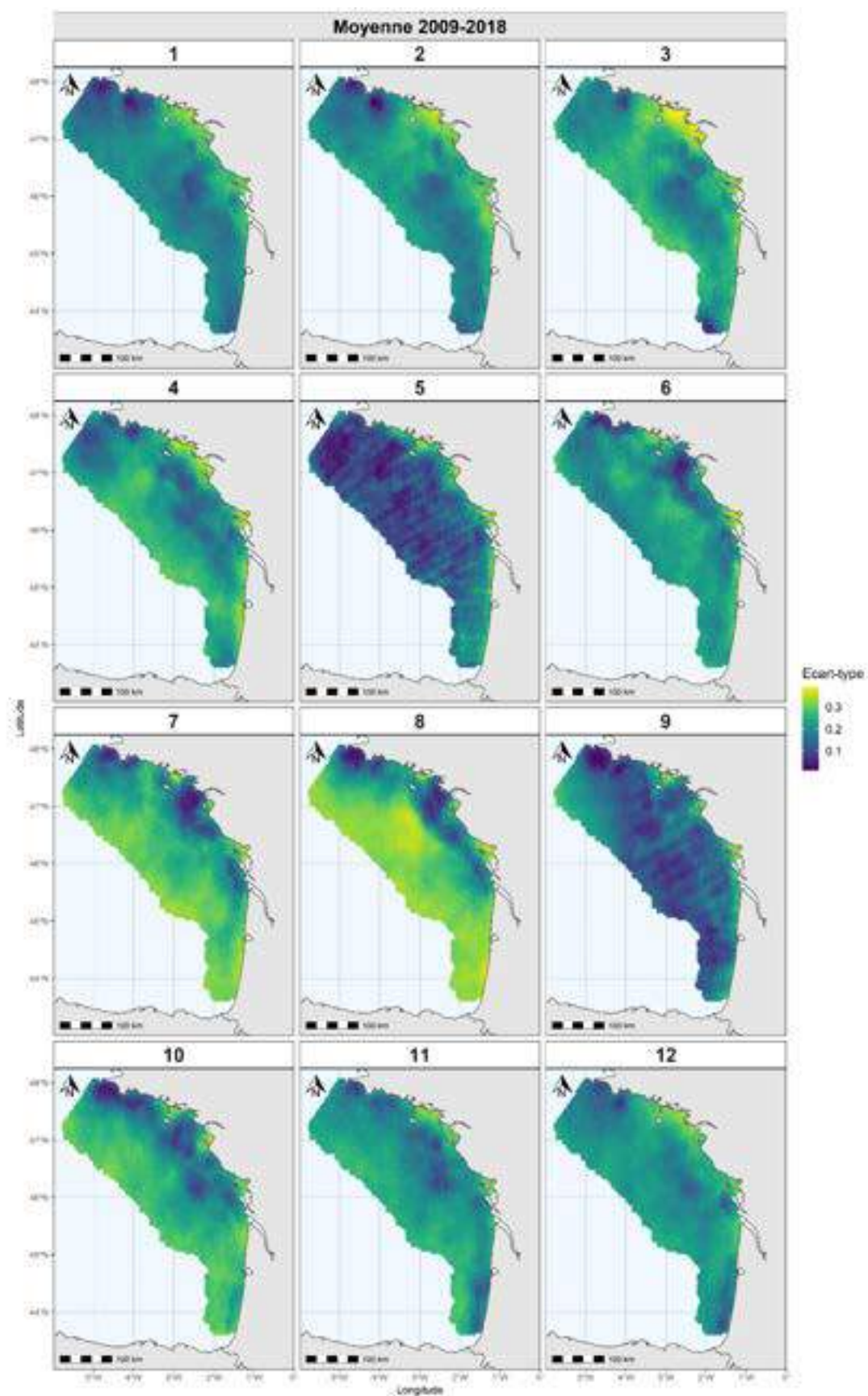
Annexe 36 : Cartographies des écarts types de la probabilité de présence de la sardine dans le GdG en 2017.



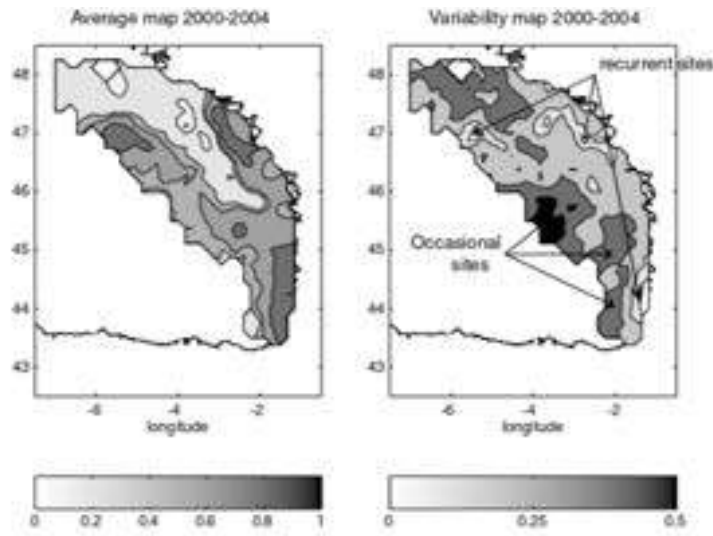
Annexe 37 : Cartographies moyennes de la probabilité de présence de la sardine dans le GdG entre 2009-2018



Annexe 38 : Cartographies moyennes des écarts types de la probabilité de présence de la sardine dans le GdG entre 2009-2018.

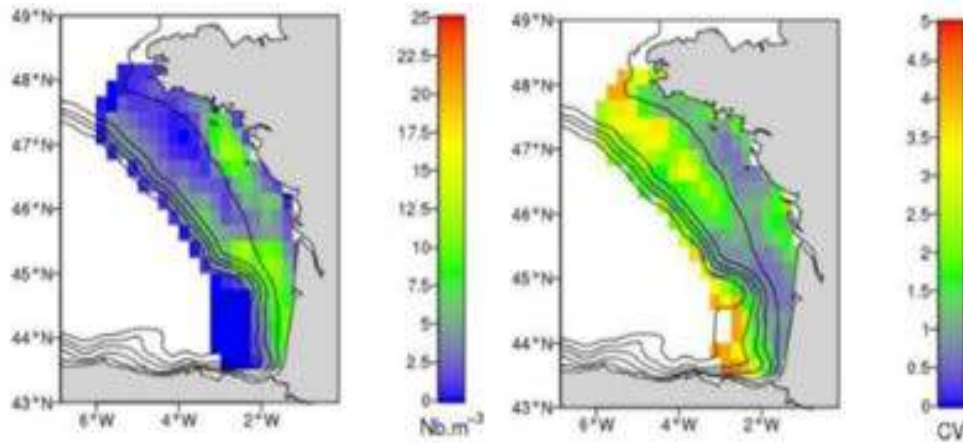


Annexe 39 : Cartographies de probabilités de présence d’œufs de sardines entre 2000 et 2004 (Bellier et al, 2007), à gauche cartes de présence moyenne et coefficients de variations à droite.

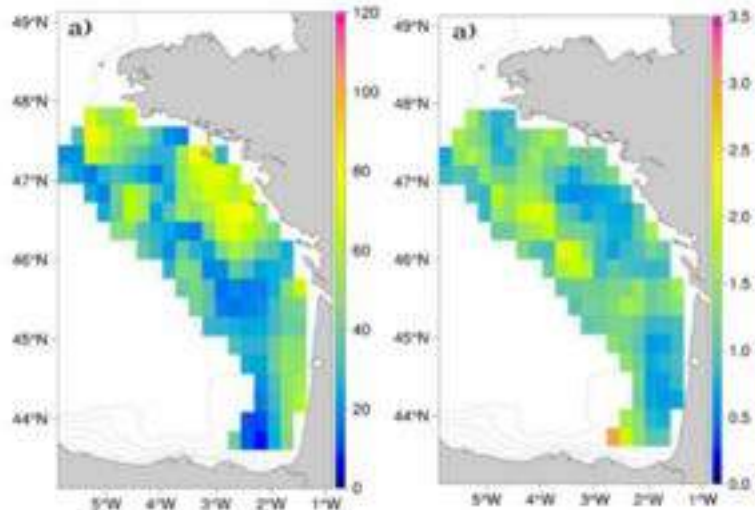


Probability of eggs presence:

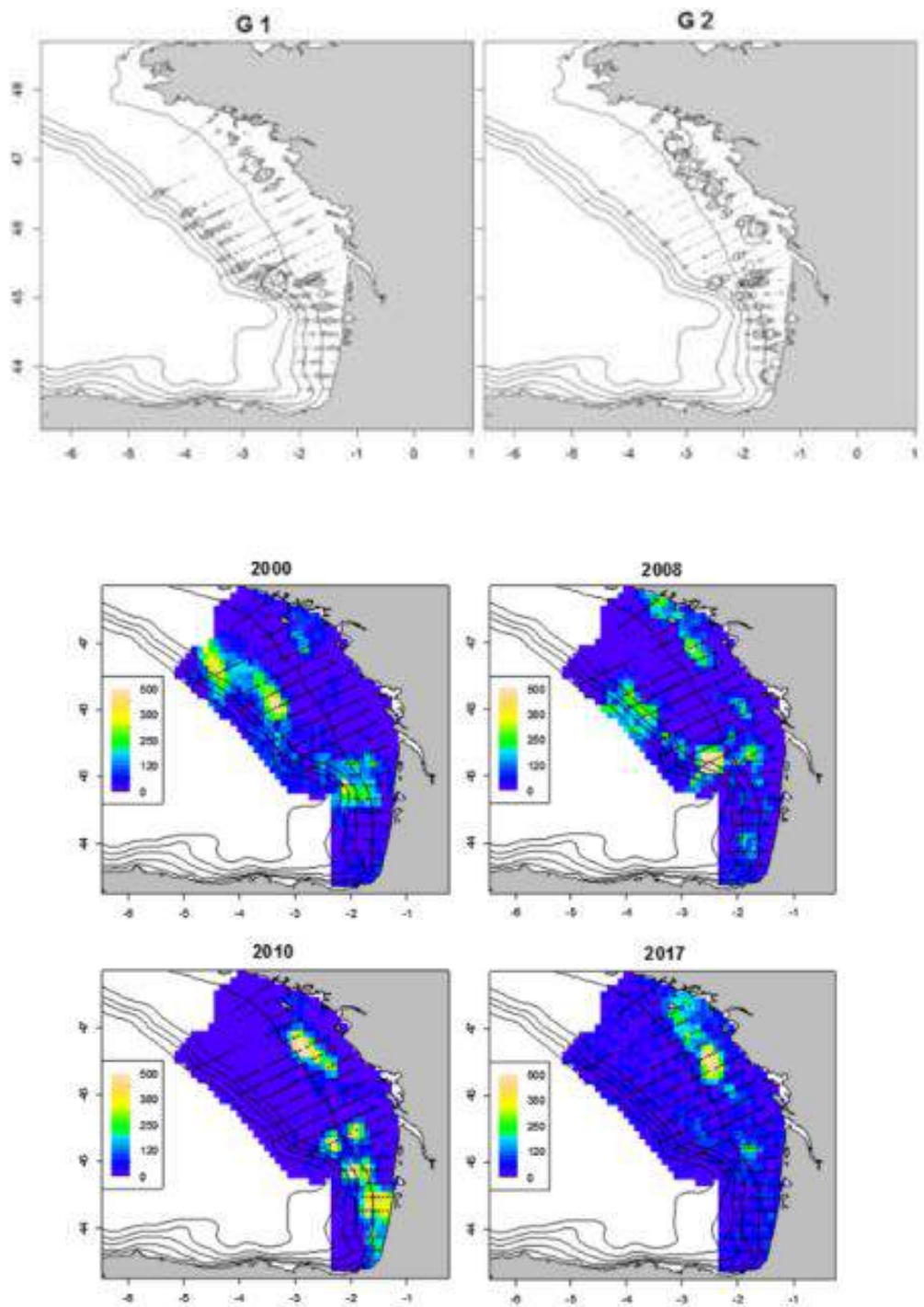
Annexe 40 : Cartographies d’abondance d’œufs de sardines (2000-2016) (Huret et al, 2018), à gauche cartes de d’abondances moyennes et coefficients de variations à droite.



Annexe 41 : Cartographies de biomasse de sardines (2000-2015) (Doray al, 2017a), à gauche cartes de présence moyenne et coefficients de variations à droite.



Annexe 42 : Distribution moyenne des abondances d'oeufs de sardines réparties en 2 groupes à partir d'un cluster sur 13 MAFS et exemples de 4 cas représentatifs des 2 groupes (G1 au-dessus : 2000 et 2008 ; G2 en-dessous : 2010 et 2017) après krigeage à partir du modèle MAF (Données PELGAS 2000- 2017) (Petitgas *et al.*, 2020).



	Diplôme : Ingénieur Agronome Spécialité : Sciences Halieutiques et Aquacoles (SHA) Spécialisation / option : Ressources et Ecosystèmes Aquatiques (REA) Enseignant référent : Olivier Le Pape
Auteur(s) : Florian Quemper Date de naissance* : 11/10/1996	Organisme d'accueil : Institut Agro – Agrocampus Ouest Rennes
Nb pages : 34 Annexe(s) : 40	Adresse : 65 rue de st Briec 35042 Rennes
Année de soutenance : 2021	Maître de stage : Baptiste Alglave
<u>Titre français :</u> Modélisation de la distribution spatiale de la sardine du Golfe de Gascogne (<i>Sardina pilchardus</i>) par intégration de données commerciales et scientifiques : enjeux et limites.	
<u>Titre anglais :</u> Modelling the spatial distribution of sardine in the Bay of Biscay (<i>Sardina pilchardus</i>) by combining commercial and scientific data: issues and limits.	
<u>Résumé :</u> En vue d'assurer une gestion durable des ressources marines, une connaissance fine de la dynamique spatio-temporelle des espèces exploitées est nécessaire. Les campagnes scientifiques fournissent des données standardisées afin de nourrir les modèles de distribution d'espèce. Néanmoins, l'information apportée par ces données reste limitée puisque l'échantillonnage est restreint dans le temps et dans l'espace. Les données de débarquements (logbooks) couplées aux données de position des navires de pêche (VMS) constituent une source d'information additionnelle qui peut être mobilisée pour cartographier la distribution des espèces à une résolution spatio-temporelle fine. Le développement d'un modèle spatio-temporel combinant les deux sources de données et prenant en compte le comportement de ciblage des pêcheurs a déjà permis d'inférer la distribution de plusieurs espèces benthodémersales du golfe de Gascogne (GdG - e.g. la sole, la baudroie, le merlan). Cette étude vise à évaluer l'intérêt de l'approche intégrée dans le cas d'une espèce pélagique (la sardine du GdG), dont l'écologie, la dynamique de flottille et la nature des données de campagne diffèrent des cas étudiés jusqu'à maintenant. Nous présentons les défis méthodologiques propres à l'application de l'approche intégrée au cas de la sardine du GdG - flottilles côtières, comportement de ciblage important, saisonnalité des pêcheries -, et le modèle utilisé pour combiner les différentes sources de données et cartographier la sardine. Sur les mois de campagne, les cartes sont essentiellement drivées par les données scientifiques. En dehors de ces mois, les données commerciales apportent de l'information dans le rayon d'action des flottilles. Toutefois, une grande partie de la zone d'étude n'est pas recouverte par les flottilles commerciales et les cartes obtenues fournissent une image partielle de la distribution de la sardine.	
<u>Abstract :</u> In order to ensure sustainable management of marine resources, detailed knowledge of the spatio-temporal dynamics of exploited species is necessary. Scientific surveys provide standardized data to feed species distribution models. However, the information provided by these data remains limited since sampling is restricted in time and space. Landings data (logbooks) coupled with vessel monitoring position data (VMS) provide an additional source of information that can be mobilized to map species distribution at a fine spatio-temporal resolution. The development of a spatio-temporal model combining the two data sources and taking into account the targeting behaviour of fishermen has already made it possible to infer the distribution of several benthodemersal species in the Bay of Biscay (BoB - e.g. sole, anglerfish, whiting). This study aims to evaluate the interest of the integrated approach in the case of a pelagic species (the sardines in the BoB), whose ecology, fleet dynamics and the nature of the survey data are very different from the cases studied so far. We present the methodological challenges of applying the integrated approach to the sardines in the BoB - coastal fleets, strong targeting behaviour, seasonality of fisheries - and the model used to combine the different data sources and map the sardine. When the surveys occurs (May and September), the maps are essentially driven by scientific data and allow the identification of distribution patterns consistent with the literature. Outside these months, commercial data provide information within the range of the fleets. However, the commercial fleets do not cover a large part of the study area and the maps obtained provide a partial representation of sardines distribution.	
<u>Mots-clés :</u> <i>Sardina pilchardus</i> , Golfe de Gascogne, données VMS, données logbooks, PELGAS, JUVENA, modélisation spatiale et spatio-temporelle, INLA.	
<u>Key Words :</u> <i>Sardina pilchardus</i> , Bay of Biscay, VMS, logbook, PELGAS, JUVENA, spatial and temporal modelling, INLA.	

* Élément qui permet d'enregistrer les notices auteurs dans le catalogue des bibliothèques universitaires