



Caractérisation de l'utilisation de l'arc Atlantique-Nord-Est
par les migrants terrestres et l'avifaune marine
à l'aide de méthodes complémentaires

1^{er} rapport méthodologique

LOT 6

Développement de méthodes de traitement et d'analyse des
données

Avril 2024





1^{er} rapport méthodologique

Livrable 6.1

LOT 6

Développement de méthodes de traitement et d'analyse des données

Titulaires du lot

Centre d'Etudes Fonctionnelles et Evolutives (CNRS, Univ Montpellier, EPHE-PSL University, IRD)

Auteurs

ROQUES Sébastien¹, QUEROUE Maud¹, BESNARD Aurélien¹

¹CEFE, Univ Montpellier, CNRS, EPHE-PSL University, IRD, Montpellier, France

Partenaires scientifiques et techniques du lot



CENTRE D'ÉCOLOGIE
FONCTIONNELLE
& EVOLUTIVE



Citation du document

ROQUES, S., QUEROUE, M. & BESNARD, A. 2024. MIGRATLANE - Caractérisation de l'utilisation de l'arc Atlantique Nord-Est par les migrants terrestres et l'avifaune marine à l'aide de méthodes complémentaires : 1^{er} rapport méthodologique. Lot 6 – Développement de méthodes de traitement et d'analyse des données. *Rapport pour l'OFB*. 28 PP.

SUIVI DU DOCUMENT

Coordinateur du projet	Nina Cudennec, Florian Le Bail	Office français de la biodiversité (OFB)
Coordinateur scientifique et technique du projet	Yann Planque Antoine Chabrolle	France Énergies Marines (FEM) Muséum National d'Histoire Naturelle (MNHN)
Auteur.e	Sébastien Roques Maud Quérroué Aurélien Besnard	Centre d'Ecologie fonctionnelle et Evolutive (CEFE)
Titre du document	Premier rapport méthodologique du lot 6 du programme MIGRATLANE	
Statut du document	Version finale	
Date de sauvegarde	10/07/2024	
Date de diffusion	10/07/2024	
Version du document	Version 3	
Nom du fichier	20240415_MIGRATLANE_Lot6_L.6.1_Rapport-méthodologique_(1er)_VF	
Nombre de pages	28	
Niveau de diffusion	Consortium / Gouvernance/ Public	

HISTORIQUE DES CHANGEMENTS

Version	Date	Modifié par	Modifications / Sections
1	12/04/24	Sébastien Roques, Maud Quérroué, Aurélien Besnard	Document initial
2	30/04/2024	Karen Bourgeois, Aurélie Blanck (OFB), Jehanne Rivet (FEM)	Relecture et modifications mineures dans l'ensemble du document
3	07/05/2024	Sébastien Roques, Maud Quérroué, Aurélien Besnard	Modifications selon les commentaires des relecteurs de la version initiale

APPROBATION

Version	Date	Approuvé par	Titre
Finale	10/07/2024	Yann Planque, FEM ; Antoine Chabrolle, MNHN/CESCO	Coordinateurs scientifiques

Signature :  

Finale	10/07/2024	Nina Cudennec, OFB ; Florian Le Bail, OFB	Pilotage du projet
--------	------------	---	--------------------

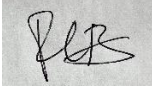
Signature :  

Table des matières

Table des matières	5
Table des figures.....	6
Rappels sur le programme MIGRATLANE.....	7
Objectif du rapport méthodologique	8
Lot 6 : Développement de méthodes de traitement et d'analyse des données	9
1. Rappels sur les objectifs du lot	9
1.1. Consortium du lot 6	9
1.2. Rappel des objectifs généraux du lot 6.....	9
2. L'intégration de données.....	10
3. Oiseaux marins nicheurs et hivernants	11
3.1. Intégration des comptages provenant des différentes campagnes aériennes en mer	11
3.1.1. Données de survols aériens	11
3.1.2. Modélisation intégrée des différentes données de comptages par survols aériens	12
3.1.2.1. Processus ponctuel de Poisson inhomogène	12
3.1.2.2. Modèles d'Occupancy	13
3.1.2.3. Sélection de covariables environnementales.....	14
3.1.3. Simulations à partir des transects existants	14
3.2. Intégration de la télémétrie et campagnes aériennes en mer	16
3.2.1. Modélisation des données de télémétrie dans le cadre d'une modélisation intégrée	17
3.2.2. Présentation de la méthodologie intégrée développée dans MIGRALION	17
4. Spatialisation des flux d'oiseaux migrants	18
4.1. Intégration des données de radar météorologiques et de radar à la côte via un IPP.	18
4.1.1. Exemple d'application de l'IPP radar sur des données simulées	19
4.2. Spatialisation des flux à partir des données de télémétrie	21
4.2.1. Exemple d'application d'une RSF sur des localisations simulées	22
4.3. Autres méthodes disponibles dans la littérature.....	23
5. Altitudes et directions de vol.....	24
5.1. Oiseaux marins et altitude de vol	24
5.2. Migrants terrestres et altitude de vol	24
5.3. Migrants terrestres et directions de vol	25
Références bibliographiques.....	25

Table des figures

Figure 1. Transects aériens effectués lors de campagnes de comptages hors MIGRATLANE identifiées par le lot 5.	12
Figure 2. Simulation d'une distribution d'oiseaux marins en fonction de covariables environnementales (a) la distance à la côte, (b) la pente sous-marine, (c) la production primaire nette. Des relations linéaires ont été simulées entre la densité d'individus et les covariables environnementales (d) relation négative avec la distance à la côte, (e) relation positive avec la pente sous-marine, (f) relation positive avec la production primaire nette (en log). (g) simulation d'une distribution : à partir des trois relations avec les covariables environnementales, une distribution est simulée sur l'ensemble de la zone. (h) simulation d'individus : à partir de la densité d'individus simulée, des localisations d'individus (points noirs) ont été simulées à l'aide d'un processus de Poisson inhomogène.	15
Figure 3. Simulation de jeux de données de transects aériens. (a) Simulation d'observations. Les points roses représentent les individus qui auraient été observés en suivant les transects aériens (traits noirs), en considérant une détection parfaite et uniforme dans les 200m de chaque côté de l'avion. (b) Efforts cumulés (transects effectués en mètres) dans chacun des sites de l'aire d'étude.	16
Figure 4. Simulation d'une intensité migratoire sur la zone d'étude de MIGRATLANE	19
Figure 5. Simulation de récolte de données par les différents dispositifs dans la zone d'étude (Birdscan, acoustique et radars météorologiques).....	20
Figure 6. Prédiction de l'intensité migratoire sur toute la zone d'étude avec (a) seulement les radars météorologiques (b) les radars météorologiques ainsi que les radars à la côte et les enregistreurs acoustiques.	21
Figure 7. Données de localisation simulées sur un axe migratoire fictif similaire à celui de l'IPP radar.	22
Figure 8. Résultat de l'utilisation de l'espace estimé par le modèle RSF à partir des données simulées.	22
Figure 9. Schéma représentant le cadre de l'analyse intégrée présentée dans Meehan et al. (2022)..	23

Rappels sur le programme MIGRATLANE

La mise en place du programme MIGRATLANE par l'OFB en 2022 repose sur l'identification de très fortes lacunes de connaissances sur l'utilisation de l'espace en mer par l'avifaune et les chiroptères sur l'Arc Atlantique Nord-Est français, c'est-à-dire au niveau des sous-régions marines Manche-mer du Nord, mers Celtiques et golfe de Gascogne. Ces lacunes de connaissances concernent à la fois l'avifaune terrestre migratrice (*e.g.* passages en mer notamment pendant les périodes de migration pré-nuptiale puis post-nuptiale), l'avifaune marine (*e.g.* zones fonctionnelles en mer) et les chiroptères (*e.g.* utilisation de l'espace marin).

Le secteur maritime de l'Arc Atlantique Nord-Est est marqué par de multiples projets d'usage, dont des projets de développement de parcs éoliens en mer. Ainsi, l'état français prévoit, dans le cadre de la programmation pluriannuelle de l'énergie (PPE), d'arriver à une puissance installée d'éoliennes en mer (posé et flottant) d'environ 5 GW en 2028 et d'environ 45 GW à l'horizon 2050. Il est ainsi apparu essentiel d'acquérir des connaissances permettant de caractériser les flux migratoires et les fonctionnalités des zones en mer pour l'avifaune et les chiroptères, à l'échelle de l'Arc Atlantique Nord-Est, que ce soit pour des espèces résidentes, de passage, présentes au cours d'une partie du cycle biologique annuel ou ponctuellement présentes afin de mieux anticiper les zones d'impacts potentiels de ces nouveaux usages et de les éviter le cas échéant.

L'OFB a ainsi lancé en 2022 le programme MIGRATLANE qui vise à acquérir des connaissances de base sur les migrations, les comportements de vol et les zones fonctionnelles en mer de l'avifaune terrestre migratrice, l'avifaune marine et les chiroptères, en considérant leurs interactions potentielles avec les projets de parcs éoliens. Une hiérarchisation des enjeux a été réalisée par un groupe d'experts rassemblés par l'OFB, afin d'identifier les questions scientifiques, les moyens et approches à prioriser pour atteindre cet objectif d'acquisition de connaissances. Ces enjeux de connaissance sont :

- Enjeux de connaissance prioritaires : les voies de migration, le flux des migrants dans le golfe de Gascogne et la Manche, les zones fonctionnelles des espèces marines, les trajets non migratoires des espèces marines.
- Enjeux de connaissance importants : la phénologie des migrations en mer, les caractéristiques de vol des oiseaux terrestres et marins en mer, la distribution spatiale des espèces marines.
- Enjeux de connaissance secondaires : les trajets non migratoires des espèces terrestres, les caractéristiques de vol des chiroptères en mer.

Le programme MIGRATLANE se déroule de 2023 à 2027, et l'acquisition des données est réalisée de 2023 à 2026. Il apportera des éléments de connaissance sur la faune volante terrestre et marine avec des données inédites sur la spatio-temporalité des déplacements locaux et les migrations des espèces au niveau de l'Arc Atlantique Nord-Est. L'acquisition de ces connaissances vise alors à éclairer la mise en œuvre des politiques publiques de préservation de ces espèces et de leurs habitats naturels, et ainsi à mieux évaluer et réguler les interactions de ces espèces avec les activités anthropiques.

Le programme s'articule autour de six lots, mettant en œuvre un ensemble de techniques et de méthodes complémentaires afin d'acquérir des connaissances sur l'avifaune et les chiroptères :

- **Lot n°1** : Coordination, synthèse et dissémination des résultats du programme.
- **Lot n°2** : Télémétrie sur les migrants terrestres et oiseaux marins.
- **Lot n°3** : Suivis acoustiques ornithologiques et à ultrasons, observations visuelles (sur des stations à la côte ou des points fixes en mer, sur des navires d'opportunité).
- **Lot n°4** : Radars installés à terre et à la côte (radars ornithologiques dédiés ou exploitation de radars météorologiques).
- **Lot n°5** : Campagnes aériennes en mer (par observations visuelles et/ou digitales).
- **Lot n°6** : Application de méthodes d'analyse multimodale des différentes données produites dans le cadre du programme et issues d'autres projets.

Le programme a officiellement commencé le 15 novembre 2022, suite à la notification des lots du marché de MIGRATLANE, et la date de début de la « mise en œuvre des actions du programme » a été actée au 15 janvier 2023 (cf. section « *Calendrier du programme* » ci-après). Le lot 6 a quant à lui commencé en décalé par rapport aux autres lots du programme (le 15 juillet 2023). La clôture du projet est prévue pour avril 2027.

Objectif du rapport méthodologique

Le présent rapport présente les méthodologies envisagées dans le cadre des analyses intégrées des données récoltées dans le programme MIGRATLANE. Une première partie résume les connaissances sur la modélisation intégrée, les parties suivantes s'articulent autour (1) de l'utilisation de l'arc Atlantique Nord-Est par les oiseaux marins, au travers de comptages et/ou de suivi GPS, (2) de l'analyse des flux migratoires se basant sur les données radars, télémétriques et acoustiques et (3) des approches envisagées pour l'analyse intégrée des hauteurs et directions de vol, pour les oiseaux marins et les migrateurs terrestres.

Lot 6 : Développement de méthodes de traitement et d'analyse des données

1. Rappels sur les objectifs du lot

1.1. Consortium du lot 6

Ce document s'inscrit dans le cadre du lot n°6 du programme MIGRATLANE dont le Centre d'Ecologie Fonctionnelle et Evolutive de Montpellier (CEFE-CNRS) est titulaire.

Le Centre d'Ecologie Fonctionnelle et Evolutive de Montpellier (UMR5175 CEFE) est un des plus importants laboratoires d'Écologie en Europe. Le projet du CEFE vise à comprendre la dynamique, le fonctionnement et l'évolution du vivant. Il s'appuie sur trois ambitions : (1) comprendre le monde vivant pour anticiper ce que sera demain, (2) conduire à des innovations et répondre aux attentes de la société ; (3) pratiquer une science « rassembleuse » et diverse dans ses approches disciplinaires.

1.2. Rappel des objectifs généraux du lot 6

L'objectif du lot 6 est de développer et de mettre en œuvre des méthodologies de traitement et d'analyses statistiques des données acquises par l'ensemble des moyens déployés dans les lots 2 à 5 de manière à caractériser de manière fiable et robuste l'utilisation de l'espace par la faune volante en mer sur l'ensemble de la zone d'étude. Il s'agit en particulier de :

- Veiller à la cohérence des différents plans d'échantillonnage spatio-temporels prévus dans les lots 2 à 5 pour s'assurer que les données obtenues pourront être comparées et agrégées dans des analyses combinées
- Proposer une analyse critique des jeux de données acquis et notamment de leur représentativité spatio-temporelle
- Identifier, sélectionner et développer, le cas échéant, des méthodes d'analyses des données permettant leur combinaison pour répondre aux objectifs du programme MIGRATLANE.
- Mobiliser les méthodes identifiées et/ou développées pour modéliser l'utilisation de l'espace par la faune volante de la zone d'étude à partir des données des lots 2 à 5 mais aussi de données issues d'autres projets déployés dans la zone (études d'impacts de parcs, suivis de colonies d'oiseaux marins par des associations, programmes de télémétrie antérieurs, suivis des oiseaux côtiers hivernants en Manche-mer du Nord, suivis de la mégafaune marine au sein des Aires Marines Protégées, etc.).

Concernant les deux derniers points, le cadre d'analyse sera nécessairement différent suivant la manière dont la faune volante utilise la zone : oiseaux marins nicheurs vs oiseaux migrateurs par exemple. La caractérisation de l'utilisation de l'espace par les oiseaux marins nicheurs aura plus attrait à des zones utilisées pour l'alimentation et le nourrissage des jeunes quand l'utilisation de l'espace par les migrateurs aura plus attrait à la caractérisation de l'organisation des flux migratoires.

Dans les deux situations, l'analyse combinée des données produites par les différents lots à travers de la modélisation intégrée implique à la fois une forte technicité autour de la modélisation statistique mais aussi du développement méthodologique.

Le rapport ci-joint est organisé de manière à présenter une introduction à l'intégration de données (Partie 2), puis de séparer les types d'analyses envisagées selon qu'elles traitent de la répartition spatiale des oiseaux marins nicheurs et hivernants (Partie 3), de la répartition spatiale des flux migratoires (Partie 4) ou des altitudes et directions de vols de l'avifaune sur la zone d'étude (Partie 5).

2. L'intégration de données

Dans de nombreuses situations d'étude de systèmes écologiques, les écologues disposent de plusieurs types d'observations ou d'ensembles de données décrivant le même système (Zipkin et al. 2019, Kéry and Royle 2020). En effet, un même système écologique peut être étudié à travers plusieurs programmes d'acquisition de connaissances (Zipkin and Saunders 2018). Dans de nombreuses situations, en particulier lors d'études à large échelle spatiale, il est ainsi fréquent de disposer de plusieurs jeux de données de nature différentes. Cela peut être par exemple des données de comptages et de données de détection/non-détection. Ces différentes sources de données contiennent des informations communes sur la distribution de l'espèce cible. Dans ce cas, il semble pertinent de combiner leur analyse pour exploiter toutes les informations que les données contiennent (Besbeas et al. 2002, Schaub and Abadi 2011, Zipkin and Saunders 2018, Miller et al. 2019). Il existe plusieurs formes de combinaison de données dans la littérature allant de l'analyse séparée de jeux de données dont les résultats sont simplement comparés jusqu'à l'analyse à l'aide de modèles spécifiquement construits pour combiner simultanément plusieurs jeux de données dans un cadre unifié, modèles dits « intégrés » (Kéry and Royle 2020). Les deux principaux avantages des modèles intégrés par rapport à des approches plus simples sont une amélioration de la précision des estimations des paramètres écologiques d'intérêts, et parfois, la possibilité d'estimer des paramètres écologiques supplémentaires qui ne pourraient pas être estimés à partir de l'analyse de chaque source de données séparément (Kéry and Royle 2020, Schaub and Kéry 2021).

Le principe de la plupart des modèles intégrés est que plusieurs jeux de données informent un même processus écologique sous-jacent, par exemple la probabilité d'utilisation de l'espace par une espèce. Le modèle statistique est construit en spécifiant une vraisemblance jointe, dans laquelle au moins un paramètre de ce processus écologique sous-jacent est partagé entre les différents jeux de données à analyser (Fletcher et al. 2019, Kéry and Royle 2020). Les différents jeux de données provenant souvent de protocoles et de méthodes d'acquisition différents, leur modélisation est ensuite basée sur l'identification/estimation de paramètres spécifiques décrivant le processus d'observation gérant les spécificités de chaque protocole de collecte (Fletcher et al. 2019, Zipkin et al. 2019). Ainsi, ces modèles gèrent un processus d'observation différent pour chaque jeu de données.

Les développements méthodologiques des dernières années autour des questions d'analyse intégrée des données ont rendu possible la combinaison de nombreux types de données différents. Il est notamment possible de combiner des données de comptages avec des données issues de suivis individuels de type capture-recapture (Besbeas et al. 2002, Jiménez et al. 2022), ou encore de combiner l'analyse de données de présence-absence collectées de manières protocolées avec des données de science participatives non-protocolées (Sun et al. 2019, cf. Kéry and Royle 2020 pour une revue). En raison de leurs possibilités d'application dans de nombreuses situations, l'utilisation des modèles intégrés en écologie s'est largement accrue ces dernières années.

Dans le cadre de MIGRATLANE, nous nous inspirons de méthodes d'analyse déjà publiées dans la littérature scientifique tout en les adaptant aux spécificités de notre cas d'étude. Nous nous appuyons aussi fortement sur les développements méthodologiques réalisés dans MIGRALION, notamment en ce qui concerne la combinaison des données de télémétrie et de comptage. Par ailleurs, la combinaison de certaines données nécessite des développements méthodologiques qui n'ont jamais été publiés, *e.g.* combiner télémétrie et données radar.

Les premières réflexions sur les données disponibles et sur les plans d'échantillonnages actuels de MIGRATLANE conduisent à proposer une séparation des analyses de l'utilisation de l'espace par les oiseaux marins d'un côté, et par les oiseaux migrateurs de l'autre. En effet, les oiseaux marins utilisent la zone d'étude notamment pour s'y nourrir ou se déplacer entre les colonies de reproduction et les zones d'alimentation ou comme aire de repos pendant l'hivernage. Il y a donc une utilisation continue de la zone d'étude bien que pouvant subir des variations saisonnières (Navarro et al. 2009, Markones et al. 2010). Pour ces espèces marines, nous chercherons ainsi à estimer une utilisation de l'espace qui englobe l'ensemble des comportements de (reproduction, alimentation, déplacements). A l'opposé, les oiseaux migrateurs survolent l'Arc Atlantique Nord-Est seulement deux fois par an pour relier leur zone d'hivernage à leur zone de reproduction. Les périodes migratoires s'étalent de février à juin pour la migration pré-nuptiale, et de juillet à novembre pour la migration post-

nuptiale. Caractériser l'utilisation de l'arc Atlantique Nord-Est par les migrateurs restreint les périodes de collecte et les jeux de données disponibles mais implique aussi de modéliser essentiellement des déplacements/des flux plutôt que l'utilisation de l'espace sensu stricto comme pour les oiseaux marins.

Bien que la question de l'utilisation de l'arc Atlantique Nord-Est reste identique pour ces deux groupes d'espèces dans le contexte de développement éolien, les écologies fondamentalement différentes entre oiseaux marins et migrateurs exigent des approches d'analyses statistiques différentes. Dans les sections suivantes, nous détaillons les approches et développements méthodologiques envisagés pour répondre à ces questions à partir de la combinaison des différents jeux de données.

3. Oiseaux marins nicheurs et hivernants

3.1. Intégration des comptages provenant des différentes campagnes aériennes en mer

Des données de comptage en mer peuvent être utilisées pour caractériser l'usage de l'espace par les oiseaux marins dans l'arc Atlantique Nord-Est. A partir de ces données, collectées par des observateurs et/ou par un système digital d'acquisition d'images en avion parcourant des transects, la distribution des espèces peut être prédite à l'aide de différentes méthodes de modélisation statistique généralement regroupées sous le terme générique de *Species Distribution Models* (Elith and Leathwick 2009, Miller 2010).

3.1.1. Données de survols aériens

Le lot 5 s'est chargé du recensement des données de campagnes aériennes récentes (depuis 2018) sur l'ensemble de la zone géographique de l'arc Atlantique Nord-Est (cf. Livrable MIGRALANE 5.1). Il en ressort que 25 différentes campagnes aériennes pourraient être exploitées dans le cadre de MIGRATLANE (**Figure 1**). De plus, quatre zones supplémentaires vont être échantillonnées par le lot 5 dans le cadre de MIGRATLANE. Par ailleurs, si d'autres campagnes aériennes plus récentes n'ayant peut-être pas encore été recensées nous sont rendues disponibles, alors elles pourront être ajoutées aux analyses.

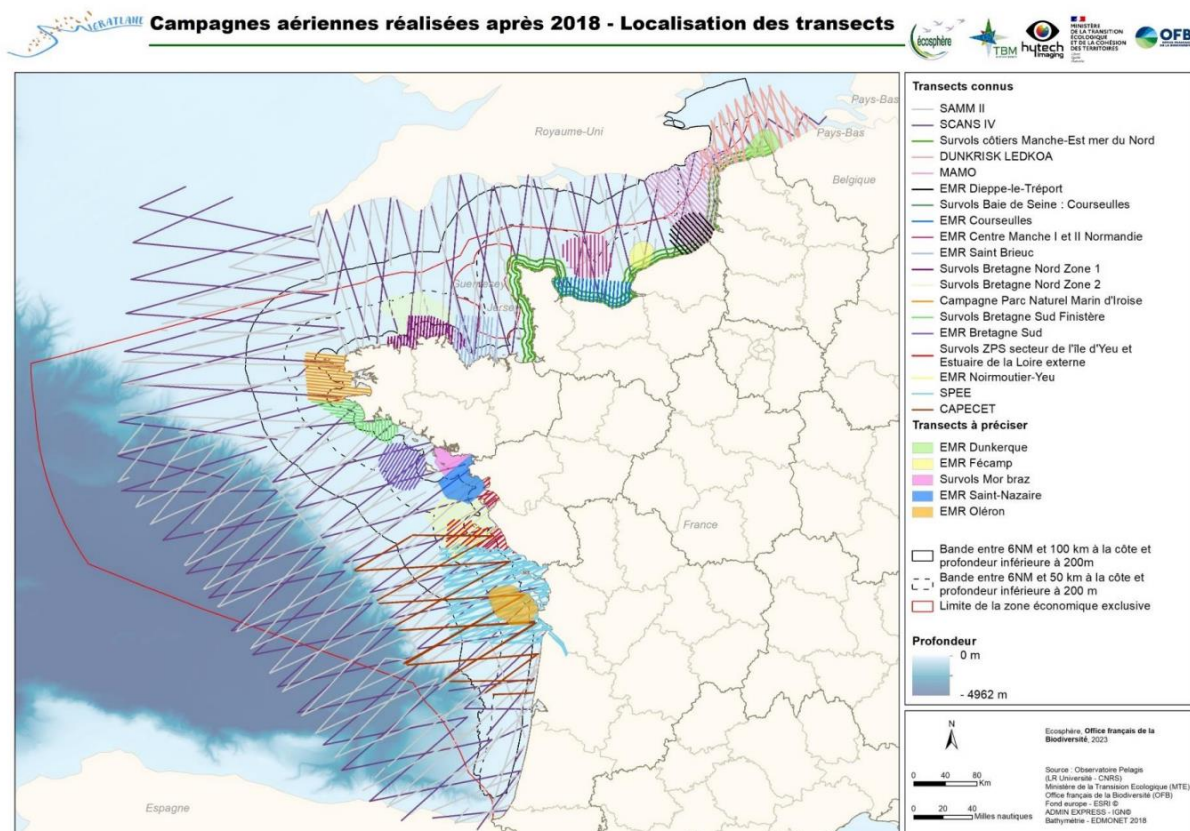


Figure 1. Transects aériens effectués lors de campagnes de comptages hors MIGRATLANE identifiées par le lot 5.

3.1.2. Modélisation intégrée des différentes données de comptages par survols aériens

Mobiliser l'ensemble de ces données disponibles doit, en théorie, permettre de produire des prédictions plus fiables et plus précises que ce qui serait obtenu d'une analyse reposant sur un seul de ces jeux de données (Isaac et al. 2020). Néanmoins, chaque jeu de données est associé à des sources de biais intrinsèques comme le processus d'observation associé (visuel, digital, visuel complété par du digital), l'effort d'échantillonnage spatial (zones d'étude de taille très hétérogènes, nombre de passage hétérogène, couvertures hétérogènes) ainsi que les conditions environnementales dans lesquelles chacun de vols a été réalisé. Ceci limite la pertinence de l'utilisation d'un modèle unique pour toutes les sources de données (Miller et al. 2019).

Afin de dépasser ces contraintes, il est préférable d'utiliser des modèles dits "intégrés" (Fletcher et al. 2019). Il s'agit de modèles statistiques qui peuvent combiner plusieurs sources de données dans un unique cadre de modélisation hiérarchique (Kéry and Royle 2020). Le principe de ces modèles hiérarchiques est de séparer la modélisation du processus latent (*i.e.* la distribution des individus dans l'espace), commune à tous les jeux de données, du processus de détection, spécifique à chaque jeu de données. Dans les paragraphes suivants, nous détaillons deux méthodes de modèles intégrés qui peuvent être utilisés pour analyser les données de transects aériens.

3.1.2.1. Processus ponctuel de Poisson inhomogène

Un processus ponctuel de Poisson inhomogène (IPP) décrit la distribution stochastique de points aléatoires (pouvant être interprétés ici comme des oiseaux marins) dans l'espace où la densité attendue des points est corrélée à des variables spatiales (*e.g.* des covariables environnementales). Il estime une fonction d'intensité qui renvoie λ_s , le nombre attendu de points à chaque emplacement infinitésimal s dans l'espace S . L'intensité attendue λ peut s'écrire sous la forme d'une fonction log-linéaire des prédictors environnementaux X :

$$\log(\lambda_s) = \beta_0 + \sum \beta X_s$$

où β_0 est l'ordonnée à l'origine de la densité et β'_s est le coefficient de régression des prédicteurs environnementaux X .

Pour estimer la sélection de l'habitat à partir des données de comptage, il est possible d'ajuster un modèle linéaire généralisé (GLM) de Poisson sur un espace géographique discret où le comptage attendu N dans la maille suit l'intensité λ d'un IPP (Aarts et al. 2012). Les données de comptage collectées constituent N , un ensemble de données de comptage. N_s représente le comptage attendu dans la maille s et est estimé comme un tirage dans une loi de Poisson de paramètre λ_s .

$$N_s \sim \text{Poisson}(\lambda_s)$$

Le GLM de Poisson peut être étendu pour tenir compte de la détection imparfaite par une formulation hiérarchique (Royle and Nichols 2003). Soit, Y_s le nombre d'individus observés dans la cellule s . Nous considérons N_s le nombre latent d'individus dans la maille s comme un tirage binomial avec une probabilité de détection p .

$$Y_s \sim \text{Binomial}(N_s, p)$$

p représentant alors la probabilité de détecter un individu sachant qu'il est présent dans la maille. p peut être spécifique à une cellule de la grille et estimé comme une régression logistique des covariables spatiale ou temporelle (*e.g.* l'effort d'observation).

3.1.2.2. Modèles d'Occupancy

Dans MIGRALION, d'autres types de modèles intégrés dits "d'occupancy" (MacKenzie et al. 2002) ont été utilisés pour estimer des distributions à partir de différentes données de transects (bateau et avion) disponibles. Ces modèles "dégradent" les données de comptages en données de détection / non-détection dans le but d'estimer une probabilité de présence. Les modèles sont implémentés dans le logiciel R (R Core Team 2024) via le package *spOccupancy* (Doser et al. 2022).

Comme précédemment énoncé, les modèles hiérarchiques permettent de séparer la modélisation du processus d'occurrence (latent), commune à tous les jeux de données, du processus de détection, spécifique à chaque jeu de données. Concernant le processus d'occurrence, si on note Z_j la présence ou l'absence d'une espèce sur le site j , alors cette variable d'occurrence résulte d'un processus de Bernoulli suivant :

$$Z_j \sim \text{Bernoulli}(\psi_j)$$

$$\text{logit}(\psi_j) = X_j^\top \beta$$

où ψ_j est la probabilité d'occurrence au site j , qui peut être décrite par une fonction des covariables X spécifiques au site et par un vecteur de coefficients de régression β .

Dans les modèles intégrés, chaque source de données est une représentation imparfaite du processus d'occurrence unique et partagé. Soit $Y_{r,a,k}$ la détection ou la non-détection d'une espèce dans le jeu de données r , sur le site a , au cours de la répétition k . Le modèle de détection peut alors être implémenté de manière distincte pour chaque source de données. Les données de détection-non-détection de r sont considérées comme résultantes d'un processus de Bernoulli conditionnel au processus latent d'occurrence :

$$Y_{r,a,k} \sim \text{Bernoulli}(p_{r,a,k} Z_{j[a]})$$

$$\text{logit}(p_{r,a,k}) = v_{r,a,k}^\top \alpha_r$$

où $p_{r,a,k}$ est la probabilité de détecter une espèce sur le site a pendant la $k_{\text{ème}}$ répétition (étant donné qu'elle est présente sur le site a) pour la source de données r , qui est une fonction des covariables v spécifiques au site, à la répétition et à la source de données, ainsi que d'un vecteur de coefficients de régression spécifiques à chaque source de données.

3.1.2.3. Sélection de covariables environnementales

Les covariables environnementales utilisées pour modéliser la distribution des individus fournissent des indications indirectes de l'abondance des ressources dans chaque site. Afin d'identifier les covariables les plus pertinentes pour prédire la distribution des différentes espèces d'oiseaux marins étudiées, il faut effectuer une sélection de modèles. Plusieurs modèles intégrés contenant différentes combinaisons de covariables environnementales sont évalués en fonction de leur WAIC (Critère d'Information Largement Applicable, Watanabe 2010) et de leur performance prédictive. Le modèle possédant le WAIC le plus faible est ensuite sélectionné nous permettant de conserver la combinaison de covariables environnementales la plus pertinente pour prédire la distribution des individus. Dans le cas où plusieurs combinaisons de variables différentes auraient des performances équivalentes, nous pourrions utiliser des méthodes de « model averaging » (Dormann et al. 2018).

3.1.3. Simulations à partir des transects existants

Les simulations permettent de vérifier si les méthodes envisagées sont adaptées aux particularités des jeux de données de transects aériens et ainsi vérifier la validité des méthodes et/ou de l'échantillonnage disponibles pour les futures analyses. L'objectif est de créer des jeux de données simulés se rapprochant au maximum des jeux de données qui seront disponibles dans le projet. La simulation de jeux de données permet de maîtriser le processus d'observation (détection des individus à partir de survol aériens) et le processus latent (distribution des individus en lien avec des covariables environnementales) afin de vérifier si ces paramètres sont, par la suite, correctement estimés par les modèles intégrés envisagés.

Par exemple, pour simuler le processus latent, nous pouvons utiliser la distance à la côte (**Figure 2a**), la pente sous l'eau (**Figure 2b**), et la production primaire nette (**Figure 2c**) comme covariables expliquant la distribution des individus dans la zone d'étude. De simples relations linéaires peuvent être simulées entre la distribution des individus et les covariables environnementales : une relation linéaire négative entre la distance à la côte et la densité d'individus (**Figure 2d**), une relation linéaire positive entre la pente sous l'eau et la densité d'individus (**Figure 2e**) et une relation linéaire positive entre la production primaire (en log) et la densité d'individus (**Figure 2f**). On obtient ainsi la carte de distribution suivante (**Figure 2g**), à partir de laquelle on peut simuler des présences d'individus avec un IPP (**Figure 2h**).

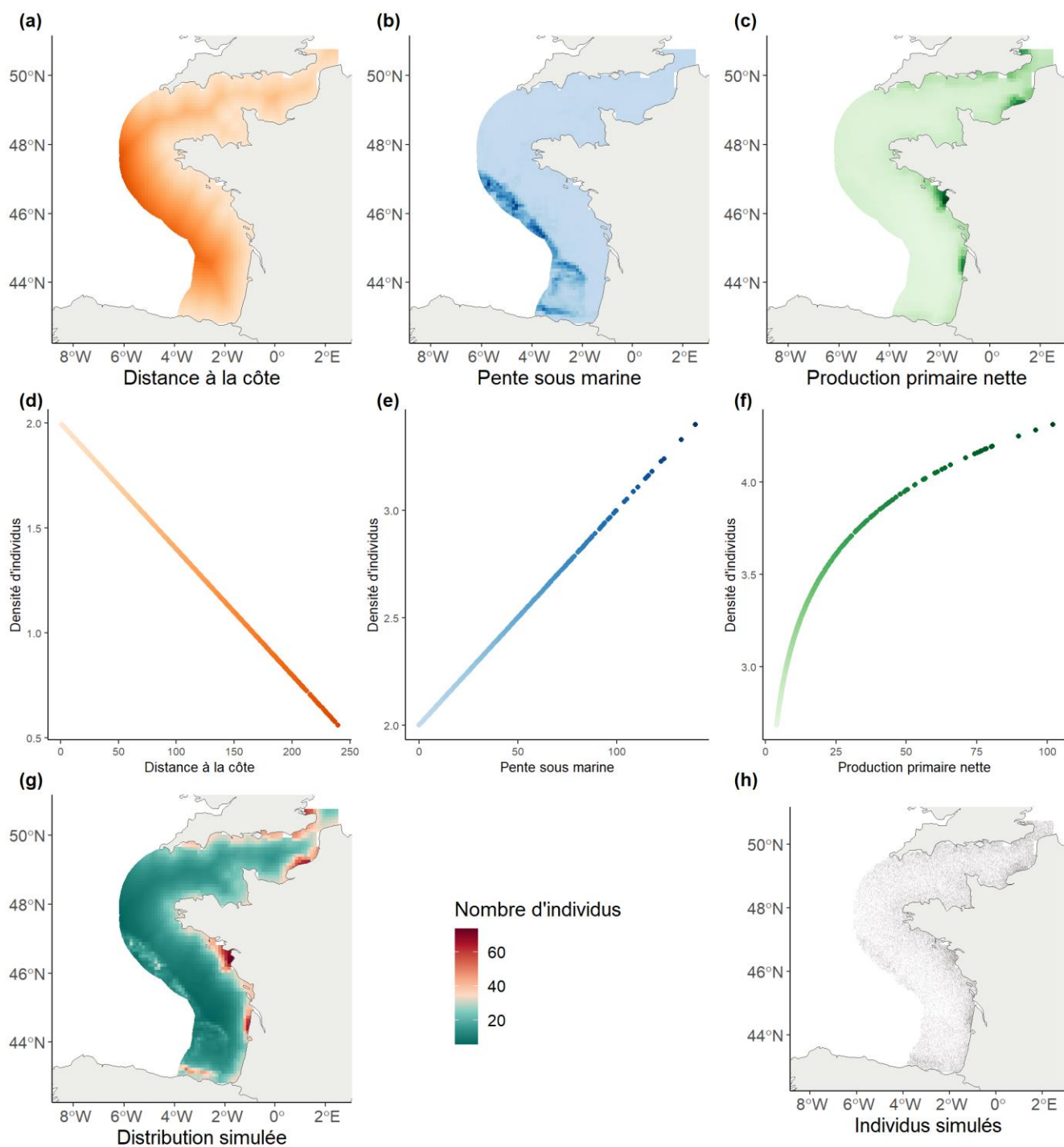


Figure 2. Simulation d'une distribution d'oiseaux marins en fonction de covariables environnementales (a) la distance à la côte, (b) la pente sous-marine, (c) la production primaire nette. Des relations linéaires ont été simulées entre la densité d'individus et les covariables environnementales (d) relation négative avec la distance à la côte, (e) relation positive avec la pente sous-marine, (f) relation positive avec la production primaire nette (en log). (g) simulation d'une distribution : à partir des trois relations avec les covariables environnementales, une distribution est simulée sur l'ensemble de la zone. (h) simulation d'individus : à partir de la densité d'individus simulée, des localisations d'individus (points noirs) ont été simulées à l'aide d'un processus de Poisson inhomogène.

Ensuite, pour simuler le processus d'observation, et ainsi créer des jeux de données simulés, nous utilisons les transects des campagnes de survols aériens identifiées par le lot 5 (**Figure 1**). Nous considérons une méthode de détection unique à

l'ensemble de ces jeux de données : la détection est considérée comme uniforme et parfaite dans les 200m de chaque côté du transect survolé. Ainsi nous pouvons simuler des détections d'individus à partir de notre carte de distribution simulée et des transects des campagnes aériennes (**Figure 3a**).

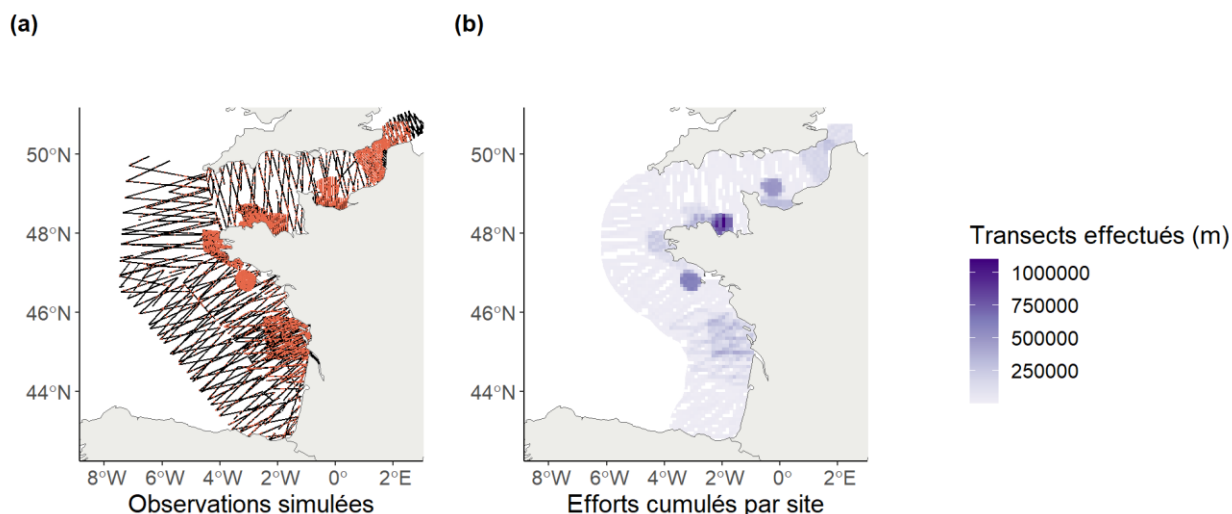


Figure 3. Simulation de jeux de données de transects aériens. (a) Simulation d'observations. Les points roses représentent les individus qui auraient été observés en suivant les transects aériens (traits noirs), en considérant une détection parfaite et uniforme dans les 200m de chaque côté de l'avion. (b) Efforts cumulés (transects effectués en mètres) dans chacun des sites de l'aire d'étude.

Une fois les jeux de données simulés, nous pouvons tester les différentes méthodes d'analyses de données envisagées afin de vérifier la compatibilité des méthodes avec les données pour estimer correctement la distribution des espèces.

Les simulations peuvent être facilement modifiées. Par exemple, le nombre de covariables environnementales ainsi que le lien entre covariables environnementales peuvent être modulés dans une infinité de combinaisons différentes. De la même manière, le processus de détection de chacun des survols peut être modifié. Nous allons concentrer nos simulations autour de quelques scénarios qui nous permettront notamment de vérifier la capacité de nos méthodes à :

- Identifier correctement les relations entre covariables environnementales et distribution des individus.
- Prendre en compte un échantillonnage très hétérogène. En effet, certaines zones ont un effort d'échantillonnage largement supérieur à des zones qui sont nettement moins couvertes (**Figure 3b**). Cette hétérogénéité peut être source de biais dans l'estimation des relations entre la distribution des individus et les covariables environnementales.
- Intégrer correctement des processus d'observations différents. En effet, selon les campagnes, différentes techniques d'observations des individus sont utilisées (digital, visuel, visuel complété par du digital) ajoutant de l'hétérogénéité dans les données.

3.2. Intégration de la télémétrie et campagnes aériennes en mer

La télémétrie et les comptages/observations en mer fournissent des informations spatiales complémentaires pour comprendre comment les animaux utilisent l'espace. D'une part, la télémétrie permet un suivi très détaillé de quelques individus, d'autre part les comptages permettent de suivre théoriquement l'ensemble de la population avec des résolutions spatiales et temporelles plus grossières. Toutefois, la comparaison et la combinaison de ces deux types de données n'est pas triviale bien qu'attrayante (Matthiopoulos et al. 2022). En effet, la télémétrie et les comptages se collectent sur des

échelles écologiques différentes et la différence fondamentale dans la nature des données enregistrées constitue un défi important pour comparer et combiner ces données (Michelot et al. 2019, Lindenmayer et al. 2022, Nathan et al. 2022).

3.2.1. Modélisation des données de télémétrie dans le cadre d'une modélisation intégrée

L'utilisation des données GPS pour décrire l'utilisation de l'habitat est abondamment documentée dans la littérature scientifique (Aarts et al. 2008, Johnson et al. 2013). Les données de télémétrie permettent de définir une fonction de sélection de l'habitat nommée *Resource Selection Function* (RSF) (Aarts et al. 2008, McLoughlin et al. 2010, Fieberg et al. 2021). Elles sont classiquement ajustées à l'aide d'une régression logistique dans laquelle les localisations des trajets de télémétrie sont codées comme des 1 et un jeu de points constituant l'habitat disponible (ou des points de "pseudo-absence") comme des 0. Appelons K l'ensemble de données qui stocke les données binaires 1/0. Les RSF opèrent dans un espace géographique continu, et K_s pour chaque point s dans S est modélisé comme un tirage de Bernoulli dans l'intensité IPP λ de l'équation :

$$K_s \sim \text{Bernoulli}(\lambda_s)$$

Pour tenir compte de la variation de la sélection de l'habitat entre les individus i , il est possible d'inclure un effet aléatoire individuel sur les paramètres de pente β'_s avec

$$\beta_i \sim \text{Normal}(\beta, \tau)$$

Pour $i \in [1, I]$

β étant le paramètre d'intérêt décrivant la force relative de sélection de l'habitat de la covariable X pour la population, τ représente la variance interindividuelle, c'est-à-dire l'hétérogénéité au sein de la population, de la sélection de l'habitat de la covariable X . La sélection des covariables environnementales peut être effectuée de la même manière que présentée dans la partie 3.1.2.3.

3.2.2. Présentation de la méthodologie intégrée développée dans MIGRALION

Les modèles de RSF (cf. 3.2.1) ainsi que les GLM de Poisson (cf. 3.1.2.1) estiment la sélection de l'habitat via une combinaison linéaire de prédicteurs environnementaux (biotiques ou abiotiques). Il est donc possible de comparer la force de sélection relative de l'habitat estimée par leurs coefficients β respectifs. Aarts et al. (2012) ont montré la convergence du coefficient de sélection relative de l'habitat β entre les différents types de modèles ouvrant ainsi des perspectives pour développer formellement un modèle intégré qui utilise conjointement les deux sources de données (télémétrie et comptages/observations en mer). Sur cette base un modèle intégré permettant d'estimer la sélection relative de l'habitat intégrant les données de télémétrie via des modèles de RSF et les données de comptage via un GLM de Poisson a été développé dans MIGRALION (Lauret et al. *soumis*).

Malgré le fait que les RSF et les GLM de Poisson quantifient la sélection de l'habitat par une régression linéaire généralisée sur les covariables X , ils n'estiment pas la même intensité absolue de l'IPP λ (Aarts et al. 2012, Northrup et al. 2022). Dans les RSF, l'intercept β_0 n'a pas de signification et les densités d'utilisation de l'espace résultante du processus de modélisation ne peuvent être interprétées qu'en relatif et non en absolu (Augar et al. 2015). Le GLM de Poisson et les RSF convergent uniquement pour estimer l'intensité relative de l'IPP (βX), c'est-à-dire que les β estimés par une RSF sont cohérents avec les β estimés par le GLM de Poisson (Aarts et al. 2012). Au contraire l'intercept des GLM peut permettre d'obtenir une intensité d'utilisation de l'espace absolue.

Le modèle intégré va ainsi conjointement estimer les β . Le modèle intégré conserve la structure hiérarchique du GLM de Poisson et la structure de régression logistique de la RSF, incluant ainsi deux modèles similaires à l'IPP avec des intercepts spécifiques et des coefficients de sélection de l'habitat partagés.

Le sous-modèle du GLM de Poisson :

$$\log(\lambda_{counts}) = \beta_0^c + \beta X$$

$$N \sim \text{Poisson}(\lambda_{counts})$$

Le sous-modèle de la RSF :

$$\text{logit}(\lambda_{RSF}) = \beta_0^r + \beta X$$

$$K \sim \text{Binomial}(\lambda_{RSF}, w)$$

On notera que les ordonnées β_0^c du sous-modèle Poisson GLM et β_0^r du sous-modèle RSF sont différentes. Alors que les paramètres λ_{counts} et λ_{RSF} ne sont pas égaux, les β s reflétant la sélection relative pour les covariables d'habitat sont partagés entre les deux sous-modèles. Ainsi, nous pouvons obtenir des cartes de l'intensité d'utilisation relative de l'habitat en représentant βX pour une RSF, un Poisson GLM, et pour le modèle intégré. Dans le cas du Poisson GLM et du modèle intégré, il est possible d'estimer un nombre d'individus attendus par cellule de l'aire d'étude. Dans ce cas, il s'agit d'une valeur absolue et non relative de la sélection de l'habitat.

4. Spatialisation des flux d'oiseaux migrateurs

4.1. Intégration des données de radar météorologiques et de radar à la côte via un IPP.

Un des objectifs du programme MIGRATLANE consiste dans le développement de méthodes permettant l'analyse jointe de données radars (avifaunistique et météorologiques), comptages visuels à la côte, télémétrie et données acoustiques dans le but d'estimer la variation spatiale des flux migratoires à l'échelle de l'arc Atlantique Nord-Est.

La variation spatiale des flux migratoires est décrite par la densité d'utilisation de l'espace par les individus sur les dimensions X et Y correspondant respectivement aux coordonnées géographiques de Longitude et de Latitude. Les données GPS, les données de radars météorologiques et dans une moindre mesure les données acoustiques ainsi que les données de radars à la côte, nous renseignent sur cette variation des flux dans l'espace (Liechti et al. 2019). Les données GLS ne permettent pas la localisation des individus équipés avec la précision nécessaire pour ce type de question (Rakhimberdiev et al. 2016). Les données issues des radars à la côte ne fournissent, quant à elles, que des informations sur des intensités de flux migratoires à quelques points donnés à la côte. La majorité de l'information spatiale permettant de décrire les flux migratoires est ainsi contenue dans les données des radars météorologiques et les données GPS.

En développement conjoint entre MIGRATLANE et MIGRALION, la méthode d'analyse envisagée se trouve sous la forme d'un modèle intégré où le processus spatial que l'on cherchera à décrire se trouvera sous la forme d'un processus spatial de Poisson :

$$N_i \sim \text{Poisson}(\lambda_i)$$

où i représente la maille dans laquelle on veut estimer l'intensité d'utilisation de l'espace. Ce qui nous intéresse est alors l'estimation du paramètre λ_i . Pour estimer ce paramètre, le but est d'estimer la relation entre ce paramètre et des covariables spatiales comme la latitude et la longitude. On peut alors l'exprimer de la façon suivante :

$$\log(\lambda_i) = \mu_0 + f(x, y)$$

où $f(x, y)$ représente une spline sur les coordonnées x et y et 0 l'intercept. La spline $f(x, y)$, nécessite l'estimation d'un nombre k de coefficients β en fonction de la façon dont nous la paramétrons (nombres de points d'inflexions de la spline).

Dans le but d'estimer cette relation, différentes sources de données sont à notre disposition (radars à la côte, radar météorologiques et télémétrie). Chacune de ces sources de données présente des caractéristiques qui lui sont propres quant à la manière dont sont générées les données (processus de détection ou d'observation) ainsi que leurs couvertures spatiales. Dans ce cadre d'analyse, chaque source de données peut avoir son propre processus de détection que l'on peut modéliser comme suit :

$$Y_{i,j} \sim \text{Binomial}(N_i, p_{i,j})$$

où $p_{i,j}$, représentant la détection du dispositif de récolte de données j dans la maille i , peut être modélisée en fonction de covariables spatiales ou environnementales (z_i) :

$$p_{i,j} = \mu_p + \alpha_p * z_i$$

avec μ_p , représentant l'ordonnée à l'origine et α_p le coefficient de la relation linéaire.

Cette modélisation des processus de détection s'applique plus particulièrement aux sources de données générant des données de comptages ou pouvant être traduites comme des données de comptages (abondances relatives) : radars météorologiques et à la côte, acoustique.

4.1.1. Exemple d'application de l'IPP radar sur des données simulées

Dans cet exemple, nous prenons la zone d'étude de MIGRATLANE dans laquelle nous simulons un axe de migration tout d'abord nord-est/sud-ouest puis prenant une composante nord-sud en arrivant sur la pointe bretonne (**Figure 4**).

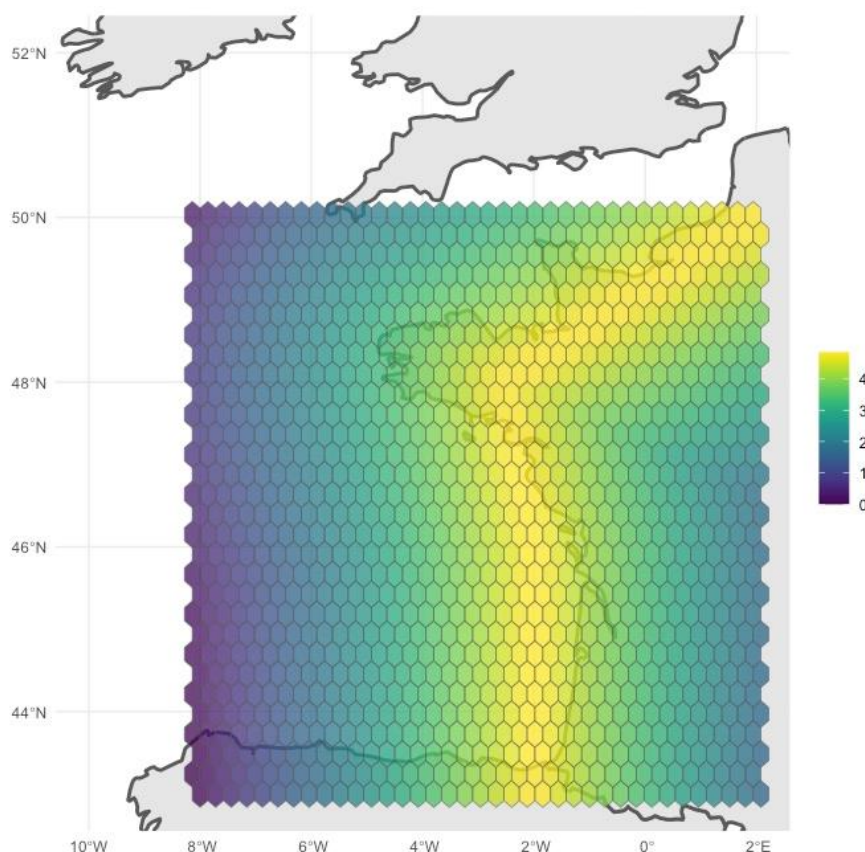


Figure 4. Simulation d'une intensité migratoire sur la zone d'étude de MIGRATLANE

De ces données simulées, nous générons deux jeux de données d'observation (simulation du processus de récolte de données) : le premier dans lequel nous avons accès à quatre radars météorologiques et le second dans lequel nous avons accès à quatre radars météorologiques ainsi que quatre radars *Birdscan* à la côte et quatre enregistreurs acoustiques (**Figure 5**).

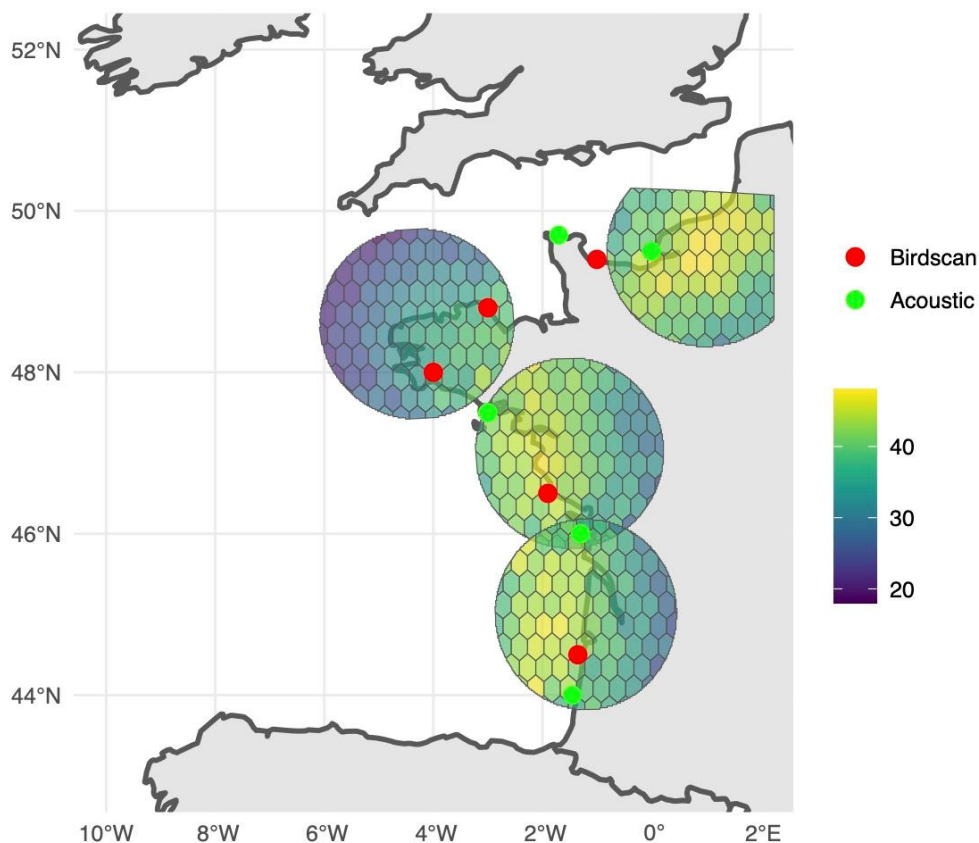


Figure 5. Simulation de récolte de données par les différents dispositifs dans la zone d'étude (Birdscan, acoustique et radars météorologiques).

Avec ces deux jeux de données différents, nous utilisons notre modèle précédent pour tenter d'estimer la relation entre l'intensité migratoire et les coordonnées géographiques. Une fois cette relation estimée, nous comparons la capacité de prédiction des deux modèles sur toute la zone d'étude (zone dans laquelle nous avons simulé l'intensité migratoire dans la **Figure 4**).

Dans l'ensemble, le patron de migration simulé est bien estimé par notre modèle avec les deux jeux de données. L'intégration des radars à la côte ainsi que des enregistreurs acoustiques ne semble pas améliorer de manière notable la prédiction du patron spatial d'intensité migratoire. Ces simulations, bien que simplifiées, confortent notre hypothèse quant à la faible contribution de ces deux dispositifs pour renseigner sur la variation spatiale d'intensité de flux migratoires.

Sur les deux prédictions (**Figure 6**), nous pouvons aussi noter que les zones de faible intensité migratoire sont globalement surestimées par rapport à notre intensité simulée (**Figure 4**) : le gradient de valeur ne descend pas sous la valeur 20 dans les prédictions alors qu'il devrait aller jusqu'à 0. Cette surestimation des zones de faibles intensité migratoire est probablement due au fait que nous n'avons pas ou peu échantillonné ces zones dans nos simulations (placement des radars météorologiques).

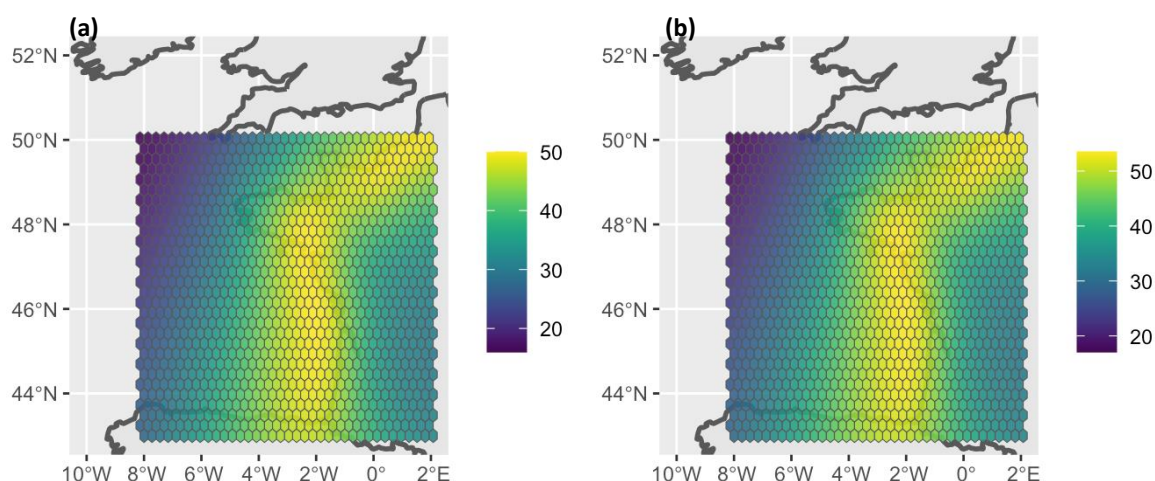


Figure 6. Prédiction de l'intensité migratoire sur toute la zone d'étude avec (a) seulement les radars météorologiques (b) les radars météorologiques ainsi que les radars à la côte et les enregistreurs acoustiques.

Cet exemple sur données simulées démontre la pertinence de continuer dans le développement du cadre d'analyse décrit dans la partie précédente pour modéliser et prédire l'utilisation de l'espace par les oiseaux migrateurs sur l'arc Atlantique Nord-Est.

4.2. Spatialisation des flux à partir des données de télémétrie

Malgré les potentiels biais liés aux différentes espèces ou zones de provenance des oiseaux équipés de GPS, il est tout de même possible d'utiliser la télémétrie pour estimer l'utilisation de l'espace par les oiseaux migrateurs. Ceci peut être réalisé à l'aide de Resource Selection Function (RSF, cf. 4.2.1. pour un exemple de simulation). La méthode de modélisation est quasi-similaire à ce qui est réalisé pour les oiseaux marins, à la différence que le lien avec l'habitat sera modélisé à l'aide d'une spline sur les coordonnées géographiques (cf. 4.1.).

Cette formulation ouvre la porte à une possibilité de développement d'un modèle intégré avec les données radars, acoustiques et télémétriques. Il a en effet été montré que les modèles utilisés pour analyser des données GPS (RSF; cf. 3.2.1.) peuvent aussi être écrits sous la forme d'un processus spatial de Poisson inhomogène (IPP; Aarts et al. 2012). Dans ce cas, les données GPS des oiseaux migrateurs équipés seraient utilisées dans un autre compartiment du modèle qui sera écrit sous la forme d'une RSF. Ainsi, les paramètres $\beta_{1,...,k}$ de la spline sur les coordonnées géographiques dans l'IPP-radar pourraient être partagés et estimés conjointement à ceux de la spline sur les coordonnées géographiques dans la RSF. Cependant, la structure mathématique de la construction d'une spline pose des défis méthodologiques qui demandent un temps de développement afin de pouvoir réaliser un modèle intégré où les paramètres de la spline sur les coordonnées géographiques seraient partagées entre l'IPP et la RSF.

Dans le cas où les verrous méthodologiques auront pu être levés, ce modèle intégré permettrait de mieux estimer l'utilisation de l'espace par les oiseaux migrateurs en intégrant toute l'information disponible sur les migrateurs récoltée dans le cadre du programme MIGRATLANE.

4.2.1. Exemple d'application d'une RSF sur des localisations simulées

Dans cet exemple, nous prenons la même zone d'étude que précédemment avec le même flux migratoire fictif. Dans le cadre d'une RSF les données simulées ne sont pas des comptages ou mesures d'intensité migratoire mais des points de localisations d'individus équipés de balises GPS. Les données simulées montrent donc des localisations simulées sur notre axe migratoire fictif (**Figure 7**).

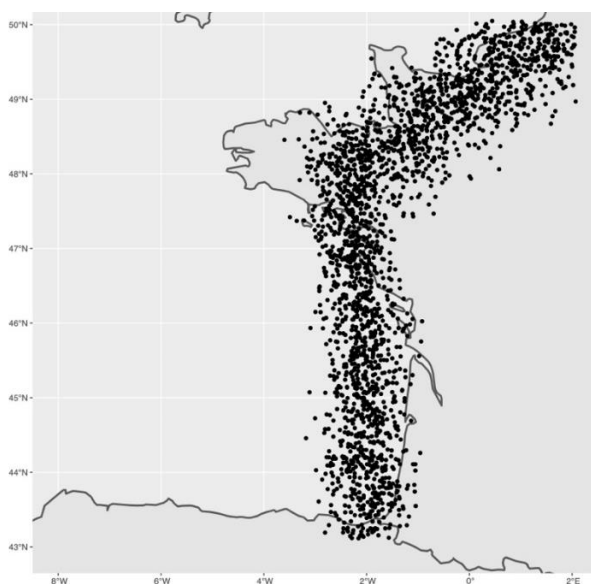


Figure 7. Données de localisation simulées sur un axe migratoire fictif similaire à celui de l'IPP radar.

Dans le cadre d'une RSF, il faut ensuite simuler des points d'habitats disponibles (pseudo-absences) de façon aléatoire dans la zone d'étude. Ces deux catégories de points sont ensuite utilisées dans le modèle de RSF pour différencier présences et pseudo-absences. Comme dans la partie précédente sur l'IPP basé sur les données radar (cf. 4.1.), nous utilisons une spline sur les coordonnées géographiques pour expliquer l'utilisation de l'espace par les oiseaux migrants. On peut voir sur la **Figure 8**, que le patron de migration simulé est bien réestimé par le modèle. Ceci ouvre la porte à une future intégration du modèle RSF avec le modèle IPP comme énoncé précédemment.

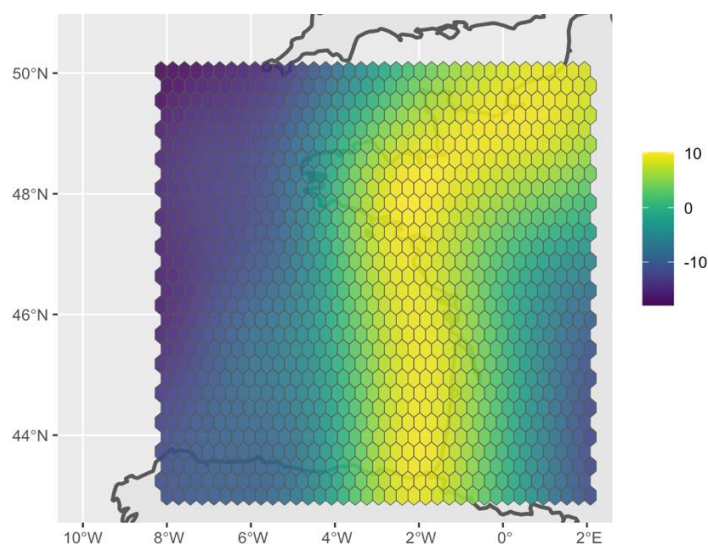


Figure 8. Résultat de l'utilisation de l'espace estimé par le modèle RSF à partir des données simulées.

4.3. Autres méthodes disponibles dans la littérature

Différentes approches modélisant les flux migratoires sont déjà disponibles dans la littérature. En effet, des auteurs ont récemment développé une méthode d'intégration de données provenant à la fois de science participative (eBird), de baguage et de télémétrie pour caractériser des patrons de migration à l'échelle du continent américain (Meehan et al. 2022, cf. **Figure 9**). Plus précisément, les auteurs conduisent tout d'abord des analyses de connectivité migratoire à l'aide des données de baguage et de télémétrie (Schuster et al. 2019, Lin et al. 2020). Ensuite, avec ces résultats de connectivité migratoire, ils simulent des trajets entre aires de reproduction et d'hivernage (Least-cost path analysis, Adriaensen et al. 2003). Enfin, à travers l'utilisation de GAMMs (Generalized Additive Mixed Models), les auteurs réalisent une analyse jointe des données de télémétrie, de présence/absence (eBird) et des trajets simulés (least-cost paths) afin de représenter les patrons spatio-temporels de déplacements d'une espèce à l'échelle d'une voie de migration.

Cette approche méthodologique paraît, en théorie, particulièrement adaptée aux programmes MIGRALION/MIGRATLANE. Cependant, ces deux programmes ne disposent pas d'autant de données que celles mobilisées par cette approche publiée dans la littérature, notamment d'un point de vue de l'homogénéité spatiale de l'échantillonnage. Une des sources de données principales de cette étude, tant en termes de couverture spatiale que de quantité de données, est celle provenant de la base de données de science participative eBird. Les auteurs ont ainsi accès à des informations de présence/absence ainsi que de densités à l'échelle de l'ensemble de la zone d'étude. Dans le cadre de MIGRATLANE on pourrait imaginer que cette source de données soit cependant remplacée par les résultats du modèle intégrant les différentes sources de données radar (cf. 4.1).

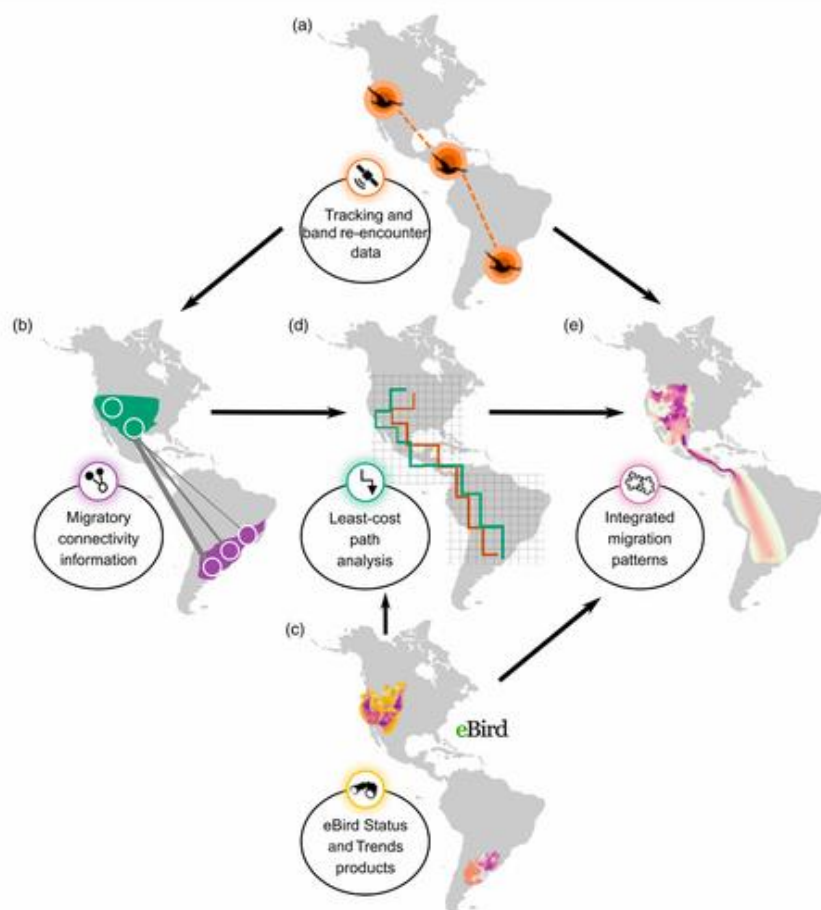


Figure 9. Schéma représentant le cadre de l'analyse intégrée présentée dans Meehan et al. (2022)..

5. Altitudes et directions de vol

5.1. Oiseaux marins et altitude de vol

Concernant les oiseaux marins, nous disposons seulement des hauteurs de vol pour les individus équipés de GPS. Il sera possible d'analyser l'altitude de vol des oiseaux marins équipés en fonction de variables environnementales et météorologiques (Ross-Smith et al. 2016, Péron et al. 2020).

L'altitude de vol z sera modélisée via un modèle linéaire généralisé avec un lien de type logarithmique en fonction des conditions météorologiques (e.g. vitesse du vent, heure de la journée, température) et de variables topographiques (e.g. distance à la côte). Les différentes données d'altitude collectées par GPS/GLS informeront le processus écologique décrivant l'altitude de vol z .

$$\log(z) = \beta_0 + \beta_1 \text{"météo"}_{x,y,t} + \beta_2 \text{"topographie"}_{x,y}$$

où β_0 , β_1 , et β_2 sont les paramètres à estimer.

Des corrections pourront être apportées pour caractériser le degré de risque vis-à-vis de l'éolien. Il sera aussi possible de modéliser directement la distribution des hauteurs de vol à l'aide de modèles dits à "espace d'états" et cela en fonction de covariables environnementales (conditions atmosphériques et de paramètres géographiques) comme pour le GLM (Ross-Smith et al. 2016). Si les données sont trop peu nombreuses pour permettre l'ajustement d'un tel modèle, il sera possible de diviser les altitudes de vol potentielles en deux classes : une classe d'altitudes de vol "à risque" correspondant à la potentielle zone de collision, et une classe d'altitudes de vol sans risque. Il s'agira ensuite d'estimer la probabilité de voler dans chaque classe d'altitudes à l'aide d'un GLM basé sur une distribution binomiale des données (cf. section suivante).

5.2. Migrateurs terrestres et altitude de vol

L'altitude de vol des migrateurs terrestres est mesurée par les différentes sources de données disponibles dans MIGRATLANE. Les GPS/GLS du lot 2 relèveront les altitudes de vol des oiseaux, tout comme les radars provenant du lot 4. Le lot 3 permettra aussi d'obtenir des données sur les altitudes de vol à l'aide d'observateurs réalisant des comptages à la côte. Les données de GPS et de GLS mesurent l'altitude de vol des oiseaux équipés tout au long de leur parcours. Les radars verticaux ornithologiques relèvent, quant à eux, les altitudes de vol des oiseaux à la côte en un point donné.

Comme pour les oiseaux marins, deux approches pour l'analyse de l'altitude de vol seront considérées (voir section sur les oiseaux marins). On pourra modéliser la distribution des hauteurs de vol à l'aide de modèles à "espace d'états" (Ross-Smith et al. 2016), ou on pourra diviser les altitudes de vol potentielles en deux classes : une classe d'altitudes de vol "à risque" correspondant à une zone à risque de collision, et une classe d'altitudes de vol sans risque et modéliser la probabilité que les oiseaux soient dans l'une ou l'autre classe à l'aide du modèle linéaire généralisé avec une distribution binomiale des données.

Pour prendre en compte les variations d'altitude de vol au cours de la migration, nous inclurons des covariables environnementales abiotiques (e.g. la distance à la côte, profondeur), météorologiques (e.g. vitesse du vent au moment de la mesure, pression atmosphérique) et phénologique (heure de la journée) influençant potentiellement l'altitude de vol.

Dans le cas d'une analyse spécifique décrivant la distribution de l'altitude de vol, les différentes données d'altitude collectées par GPS/GLS, radar et observateurs pourront informer un même processus écologique décrivant l'altitude de vol z . Cette analyse passera par une régression linéaire généralisée avec un lien de type logarithmique.

Par exemple :

$$\log(z) = \beta_0 + \beta_1 \text{"vent"}_{x,y,t} + \beta_2 \text{"distance à la côte"}_{x,y}$$

où β_0 , β_1 , et β_2 sont les paramètres à estimer.

Dans l'idéal, cette analyse permettra d'obtenir une prédiction de l'altitude de vol en fonction des conditions atmosphériques et de paramètres géographiques (Shamoun-Baranes et al. 2017, Panuccio et al. 2019). Ce modèle pourra également servir de base pour caractériser des différences entre groupes d'espèces et détailler la phénologie de vol, tant au niveau du cycle journalier qu'au cours de la saison (La Sorte et al. 2015, Wills 2017).

5.3. Migrateurs terrestres et directions de vol

Les données issues des suivis GPS, suivis radars et comptages à la côte des lots 2, 3 et 4 nous renseignent sur les directions de vol des migrateurs. Les données GPS fournissent les directions de vol des oiseaux équipés tout au long de leur parcours. Les radars verticaux ornithologiques et les observateurs relèvent, quant à eux, les directions de vol à la côte.

Dans le cas d'une analyse spécifique décrivant la distribution des directions de vol, les différentes données collectées par GPS, radar et observateurs pourront informer un même processus écologique décrivant la direction de vol d . Afin de prendre en compte la nature circulaire de cette variable (mesure d'angles), cette analyse passera par un système de deux régressions linéaires généralisées :

$$\begin{cases} \cos(d) = \beta_0 + \beta_1 \text{"vent"}_{x,y,t} + \beta_2 \text{"distance à la côte"}_{x,y} \\ \sin(d) = \beta_0 + \beta_1 \text{"vent"}_{x,y,t} + \beta_2 \text{"distance à la côte"}_{x,y} \end{cases}$$

où β_0 , β_1 , et β_2 sont les paramètres à estimer.

L'interprétation des sorties du modèle se fait par back-transformation des sinus et cosinus, et grâce à l'utilisation de plusieurs indices quantifiant les changements de directions à différents points de la courbe de réponse (Cremers and Klugkist 2018, Cremers et al. 2018).

Cette analyse permettra d'obtenir une prédiction des directions de vol en fonction des conditions atmosphériques et de paramètres géographiques. Ce modèle pourra également servir de base pour caractériser des différences entre groupes d'espèces et détailler la phénologie des directions, tant au niveau du cycle journalier qu'au cours de la saison (Cabrera-Cruz et al. 2017, Welcker et al. 2017, Hirschhofer et al. 2023).

Il sera également intéressant de regarder le lien entre direction, hauteur et vitesse de vol. Le croisement de ces différentes régressions permettra potentiellement de distinguer des typologies de déplacements plus ou moins caractéristiques des passages migratoires (Archibald et al. 2017).

Références bibliographiques

- Aarts, G., J. Fieberg, and J. Matthiopoulos. 2012. Comparative interpretation of count , presence – absence and point methods for species distribution models. *Methods in Ecology and Evolution* 3:177–187.
- Aarts, G., M. MacKenzie, B. McConnell, M. Fedak, and J. Matthiopoulos. 2008. Estimating space-use and habitat preference from wildlife telemetry data. *Ecography* 31:140–160.

- Adriaensen, F., J. P. Chardon, G. De Blust, E. Swinnen, S. Villalba, H. Gulinck, and E. Matthysen. 2003. The application of 'least-cost' modelling as a functional landscape model. *Landscape and urban planning* 64:233–247.
- Archibald, K. M., J. J. Buler, J. A. Smolinsky, and R. J. Smith. 2017. Migrating birds reorient toward land at dawn over the Great Lakes, USA. *The Auk: Ornithological Advances* 134:193–201.
- Avgar, T., J. R. Potts, M. A. Lewis, and M. S. Boyce. 2015. Integrated step selection analysis: Bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution* 7:619–630.
- Besbeas, P., S. N. Freeman, B. J. T. Morgan, and E. A. Catchpole. 2002. Integrating Mark-Recapture-Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters. *Biometrics* 58:540–547.
- Cabrera-Cruz, S. A., T. J. Mabey, and R. Villegas-Patraca. 2017. Patterns of nocturnal bird migration in southern Mexico. *Revista mexicana de biodiversidad* 88:867–879.
- Cremers, J., and I. Klugkist. 2018. One direction? A tutorial for circular data analysis using R with examples in cognitive psychology. *Frontiers in psychology* 9:396689.
- Cremers, J., K. T. Mulder, and I. Klugkist. 2018. Circular interpretation of regression coefficients. *British Journal of Mathematical and Statistical Psychology* 71:75–95.
- Dormann, C. F., J. M. Calabrese, G. Guillerá-Arroita, E. Matechou, V. Bahn, K. Bartoń, C. M. Beale, S. Ciuti, J. Elith, K. Gerstner, J. Guelat, P. Keil, J. J. Lahoz-Monfort, L. J. Pollock, B. Reineking, D. R. Roberts, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, S. N. Wood, R. O. Wüest, and F. Hartig. 2018. Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs* 88:485–504.
- Doser, J. W., A. O. Finley, M. Kéry, and E. F. Zipkin. 2022. spOccupancy : An R package for single- and integrated spatial occupancy models. *Methods in Ecology and Evolution* 13:1670–1678.
- Elith, J., and J. R. Leathwick. 2009. Species Distribution Models : Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* 40:677–697.
- Fieberg, J., J. Signer, B. Smith, and T. Avgar. 2021. A 'How to' guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology* 90:1027–1043.
- Fletcher, R. J., T. J. Hefley, E. P. Robertson, B. Zuckerberg, R. A. McCleery, and R. M. Dorazio. 2019. A practical guide for combining data to model species distributions. *Ecology* 100:e02710.
- Hirschhofer, S., F. Liechti, P. Ranacher, R. Weibel, and B. Schmid. 2023. High-intensity bird migration along Alpine valleys calls for protective measures against anthropogenically induced avian mortality. *Remote Sensing in Ecology and Conservation*.
- Isaac, N. J. B., M. A. Jarzyna, P. Keil, L. I. Dambly, P. H. Boersch-supan, E. Browning, S. N. Freeman, N. Golding, G. Guillerá-arroita, P. A. Henrys, S. Jarvis, J. J. Lahoz-Monfort, J. Pagel, O. L. Pescott, R. Schmucki, E. G. Simmonds, and R. B. O'Hara. 2020. Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology & Evolution* 35:56–67.
- Jiménez, J., F. Díaz-Ruiz, P. Monterroso, J. Tobajas, and P. Ferreras. 2022. Occupancy data improves parameter precision in spatial capture–recapture models. *Ecology and evolution* 12:e9250.
- Johnson, D. S., M. B. Hooten, and C. E. Kuhn. 2013. Estimating animal resource selection from telemetry data using point process models. *Journal of Animal Ecology* 82:1155–1164.
- Kéry, M., and J. A. Royle. 2020. Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models. Academic Press.
- Lauret, V., Courbin, N. and Besnard A. (in progress). Comparing and integrating movement and counts data to estimate animal habitat selection.

- Liechti, F., J. Aschwanden, J. Blew, M. Boos, R. Brabant, A. M. Dokter, V. Kosarev, M. Lukach, M. Maruri, and M. Reyniers. 2019. Cross-calibration of different radar systems for monitoring nocturnal bird migration across Europe and the Near East. *Ecography* 42:887–898.
- Lin, H.-Y., R. Schuster, S. Wilson, S. J. Cooke, A. D. Rodewald, and J. R. Bennett. 2020. Integrating season-specific needs of migratory and resident birds in conservation planning. *Biological conservation* 252:108826.
- Lindenmayer, D. B., T. Lavery, and B. C. Scheele. 2022. Why we need to invest in large-scale, long-term monitoring programs in landscape ecology and conservation biology. *Current Landscape Ecology Reports* 7:137–146.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255.
- Markones, N., V. Dierschke, and S. Garthe. 2010. Seasonal differences in at-sea activity of seabirds underline high energetic demands during the breeding period. *Journal of Ornithology* 151:329–336.
- Matthiopoulos, J., E. Wakefield, J. W. E. Jeglinski, R. W. Furness, M. Trinder, G. Tyler, A. McCluskie, S. Allen, J. Braithwaite, and T. Evans. 2022. Integrated modelling of seabird-habitat associations from multi-platform data: A review. *Page Journal of Applied Ecology*.
- McLoughlin, P. D., D. W. Morris, D. Fortin, E. Vander Wal, and A. L. Contasti. 2010. Considering ecological dynamics in resource selection functions. *Journal of animal ecology* 79:4–12.
- Meehan, T. D., N. L. Michel, E. J. Knight, L. Taylor, S. P. Saunders, W. V. Deluca, J. L. Deppe, M. F. Jimenez, N. E. Seavy, M. A. Smith, C. Witko, M. E. Akresh, D. R. Barber, J. C. Beasley, J. L. Belant, E. M. Bayne, R. O. Bierregaard, K. L. Bildstein, T. J. Boves, S. P. C. A. Celis-murillo, D. T. King, S. A. Mackenzie, K. P. Mcfarland, O. E. Rhodes, J. Christopher, R. E. Norvell, and C. B. Wilsey. 2022. Integrating data types to estimate spatial patterns of avian migration across the Western Hemisphere. *Ecological Applications* 32:e2679.
- Michelot, T., P. G. Blackwell, and J. Matthiopoulos. 2019. Linking resource selection and step selection models for habitat preferences in animals. *Ecology* 100:1–12.
- Miller, D. A. W., K. Pacifici, J. S. Sanderlin, and B. J. Reich. 2019. The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution* 10:22–37.
- Miller, J. 2010. Species distribution modeling. *Geography Compass* 4:490–509.
- Nathan, R., C. T. Monk, R. Arlinghaus, T. Adam, J. Alós, M. Assaf, H. Baktoft, C. E. Beardsworth, M. G. Bertram, and A. I. Bijleveld. 2022. Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science* 375:eabg1780.
- Navarro, J., M. Louzao, J. M. Igual, D. Oro, A. Delgado, J. M. Arcos, M. Genovart, K. A. Hobson, and M. G. Forero. 2009. Seasonal changes in the diet of a critically endangered seabird and the importance of trawling discards. *Marine Biology* 156:2571–2578.
- Northrup, J. M., E. Vander Wal, M. Bonar, J. Fieberg, M. P. Laforge, M. Leclerc, C. M. Prokopenko, and B. D. Gerber. 2022. Conceptual and methodological advances in habitat-selection modeling: guidelines for ecology and evolution. *Ecological Applications* 32:1–31.
- Panuccio, M., G. Dell'Omo, G. Bogliani, C. Catoni, and N. Sapir. 2019. Migrating birds avoid flying through fog and low clouds. *International Journal of Biometeorology* 63:231–239.
- Péron, G., J. M. Calabrese, O. Duriez, C. H. Fleming, R. García-Jiménez, A. Johnston, S. A. Lambertucci, K. Safi, and E. L. C. Shepard. 2020. The challenges of estimating the distribution of flight heights from telemetry or altimetry data. *Animal Biotelemetry* 8:1–13.
- R Core Team. 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing,

Vienna, Austria.

- Rakhimberdiev, E., N. R. Senner, M. A. Verhoeven, D. W. Winkler, W. Bouten, and T. Piersma. 2016. Comparing inferences of solar geolocation data against high-precision GPS data: annual movements of a double-tagged black-tailed godwit. *Journal of Avian Biology* 47:589–596.
- Ross-Smith, V. H., C. B. Thaxter, E. A. Masden, J. Shamoun-Baranes, N. H. K. Burton, L. J. Wright, M. M. Rehfish, and A. Johnston. 2016. Modelling flight heights of lesser black-backed gulls and great skuas from GPS: A Bayesian approach. *Journal of Applied Ecology* 53:1676–1685.
- Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. *Ecology* 84:777–790.
- Schaub, M., and F. Abadi. 2011. Integrated population models: a novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology* 152:227–237.
- Schaub, M., and M. Kéry. 2021. Integrated Population Models.
- Schuster, R., S. Wilson, A. D. Rodewald, P. Arcese, D. Fink, T. Auer, and J. R. Bennett. 2019. Optimizing the conservation of migratory species over their full annual cycle. *Nature communications* 10:1754.
- Shamoun-Baranes, J., F. Liechti, and W. M. G. Vansteelant. 2017. Atmospheric conditions create freeways, detours and tailbacks for migrating birds. *Journal of Comparative Physiology A* 203:509–529.
- La Sorte, F. A., W. M. Hochachka, A. Farnsworth, D. Sheldon, B. M. Van Doren, D. Fink, and S. Kelling. 2015. Seasonal changes in the altitudinal distribution of nocturnally migrating birds during autumn migration. *Royal Society Open Science* 2:150347.
- Sun, C. C., J. A. Royle, and A. K. Fuller. 2019. Incorporating citizen science data in spatially explicit integrated population models. *Ecology* 100:e02777.
- Watanabe, S. 2010. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research* 11:3571–3594.
- Welcker, J., M. Liesenjohnann, J. Blew, G. Nehls, and T. Grünkorn. 2017. Nocturnal migrants do not incur higher collision risk at wind turbines than diurnally active species. *Ibis* 159:366–373.
- Wills, G. V. 2017. Monitoring Avian Migration with Dedicated Vertical-looking Radar. University of Exeter (United Kingdom).
- Zipkin, E. F., B. D. Inouye, and S. R. Beissinger. 2019. Innovations in data integration for modeling populations. *Ecology* 100:1–3.
- Zipkin, E. F., and S. P. Saunders. 2018. Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation* 217:240–250.