


MIGRALION

Caractérisation de l'utilisation du golfe du lion
par les migrateurs terrestres et l'avifaune marine
à l'aide de méthodes complémentaires

**Lot 6 : Développement de méthodes permettant l'analyse des
différentes données produites dans le cadre du programme et issues
d'autres projets**
Premier rapport d'analyse méthodologique et des données acquises

Livable L32

Septembre 2022

Partenaires scientifiques et techniques



Financeurs du programme



Programme financé par le Ministère de la Transition
écologique, la Région Sud et la Région Occitanie

SUIVI DU DOCUMENT

Auteurs	Valentin Lauret, Olivier Gimenez, Aurélien Besnard	Centre d'Ecologie Fonctionnelle et Evolutive (CEFE) – EPHE
Coordination scientifique et technique du projet	Aurélien Besnard Sophie de Grissac (avant 03/2023) Yann Planque (après 05/2023)	CEFE – EPHE France Énergies Marines (FEM)
Coordination du projet	Alexandra Gigou (avant 11/2023) Claire Hébert	Office Français de la Biodiversité (OFB)
Titre du document	Migralion - Premier rapport d'analyse méthodologique et des données acquises – livrable L32 – Lot 6 : Développement de méthodes permettant l'analyse des différentes données produites dans le cadre du programme et issues d'autres projets	
Citation du document	LAURET, V., GIMENEZ, O., BESNARD, A. 2022. MIGRALION - Caractérisation de l'utilisation du golfe du lion par les migrateurs terrestres et l'avifaune marine à l'aide de méthodes complémentaires : Premier rapport d'analyse méthodologique et des données acquises. Lot 6 : Développement de méthodes permettant l'analyse des différentes données produites dans le cadre du programme et issues d'autres projets. <i>Rapport pour l'OFB</i> . 40 PP.	
Statut du document	Version finale	
Date de sauvegarde	09/2022	
Date de diffusion	09/2022 (version validée) puis 04/2024 (actualisation du format de suivi)	
Version du document	3	
Nom du fichier	MIGRALION_Lot 6_L32_Rapportmethodologique_2022-09.pdf	
Nombre de pages	40	
Niveau de diffusion	Consortium / Gouvernance/ Public	

HISTORIQUE DES CHANGEMENTS

Version	Date	Modifié par	Modifications / Sections
1	2022	Valentin Lauret, Olivier Gimenez Aurélien Besnard	Document initial
2	09/2022	Sophie de Grissac (FEM), Alexandra Gigou (OFB)	Relecture globale et corrections mineures dans l'ensemble du document, validation du rapport
3	23/04/2024	Jehanne Rivet (FEM), Claire Hébert (OFB)	Mise à jour du format Ajout page de suivi

APPROBATION

Version	Date	Approuvé par	Titre
Finale	24/04/2024	Yann Planque, FEM	Coordinateur scientifique et technique
Signature : 			
Finale	25/04/2024	Claire Hébert, OFB	Coordinatrice du projet
Signature : 			

Table des matières

I.	Synthèse bibliographique sur l'état de l'art de l'intégration de données	5
II.	Description des données mobilisables.....	10
A.	Données historiques	10
1.	SAMM.....	10
2.	PELMED	11
3.	Fermes pilotes.....	12
4.	Parc Naturel Marin du Golfe du Lion	12
5.	Recensement des colonies du GISOM	12
6.	Comptages côtiers pendant les périodes migratoires	13
7.	Recensement des études télémétriques du CEREMA	14
8.	Autres données issues des suivis télémétriques disponibles	15
B.	Données Migralion.....	15
1.	Télémétrie.....	15
2.	Radar	15
III.	La séparation entre oiseaux marins et migration	16
IV.	Oiseaux marins.....	18
A.	Méthodologie envisagée.....	18
B.	Intégration des données historiques et télémétrie : RSF, <i>occupancy</i>	19
1.	Processus latent	19
2.	Processus d'observation : comptages en mer	20
3.	Processus d'observation : télémétrie	21
C.	Ecriture du modèle pour la sterne caugek.....	22
D.	Oiseaux marins et altitude de vol	24
V.	Flux migratoires.....	25
A.	Données	26
B.	L'analyse par "dimension"	27
1.	La spatialisation des flux : les dimensions X et Y	27
2.	L'altitude de vol : la dimension Z	28
3.	Phénologie de la migration : la dimension T.....	29

C.	Formulation d'un modèle 4D et intégration	29
1.	Processus écologique	29
2.	Processus d'observation	30
3.	Anticipation des difficultés.....	32
VI.	Perspectives	34
A.	Dans l'attente des analyses écologiques par lot	34
B.	Incertitudes autour de la contribution des GLS	34
C.	Phénologie : étude spécifique ou déclinaison des autres modèles	34
D.	Quelles adaptations ?.....	35
VII.	Références.....	36

Table des figures

Figure 1 :	Campagnes de Survol Aérien de la Mégafaune Marine (SAMM), plan d'échantillonnage spatial et fréquence des espèces observées	11
Figure 2 :	Campagnes PELMED, plan d'échantillonnage spatial et fréquence des espèces observées	11
Figure 3 :	Campagnes MégaObs du Parc Naturel Marin (PNM) du Golfe du Lion, plan d'échantillonnage spatial et fréquence des espèces observées	12
Figure 4 :	Campagnes de recensement des colonies de laro-limicoles 2011-2021, plan d'échantillonnage spatial et fréquence des espèces observées	13
Figure 5 :	Localisation des sites de comptages de migration explorés à la date de Septembre 2022	14
Figure 6 :	Exemple de grille utilisable pour la modélisation de l'utilisation de l'espace par les oiseaux marins	20
Figure 7 :	Données de comptages et de télémétrie disponibles sur les Sternes caugek	23
Figure 8 :	Présentation schématique des données, covariables, et processus écologiques du modèle estimant les flux migratoires en 4 dimensions..	32

I. Synthèse bibliographique sur l'état de l'art de l'intégration de données

Dans de nombreuses situations d'étude de systèmes écologiques et de leur fonctionnement, les écologues disposent de plusieurs types d'observations ou d'ensembles de données décrivant toutes le même système (Kéry & Royle, 2020; Zipkin et al., 2019). En effet, l'étude d'un même système écologique peut s'aborder à travers différentes questions qui peuvent, de fait, conduire à la mise en place de plusieurs programmes de collectes de données de nature différente par différents partenaires (Zipkin & Saunders, 2018). Ainsi, plusieurs campagnes de collectes de données sur un même système peuvent être menées avec différentes méthodes d'acquisition. Le choix du protocole de collecte de données se réalise en fonction du coût qu'il représente, des spécialités et des capacités des partenaires qui le mettent en oeuvre, et des spécificités de chaque méthode de collecte pour répondre à la question écologique posée (Nichols & Williams, 2006). Par exemple, l'étude de la population de grands dauphins (*Tursiops truncatus*) de Méditerranée française se base sur des survols aériens quadrillant la zone côtière et pélagique, ainsi que sur des suivis côtiers à bord de bateaux (voir Box 1). Dans ce cas, les zones couvertes, les périodes d'acquisition, mais aussi les structures opérant ces suivis sont différentes (Lauret, Labach, Turek, et al., 2021). 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, par exemple des données de comptages et de données de détection/non-détection, qui contiennent tous deux des informations sur la distribution de l'espèce cible. Dans ces cas où plusieurs jeux de données différents sont disponibles, il semble pertinent de vouloir combiner leur analyse pour exploiter toutes les informations que les données contiennent (Besbeas et al., 2002; Miller et al., 2019; Schaub & Abadi, 2011; Zipkin & Saunders, 2018). Il existe plusieurs formes de combinaison de données dans la littérature allant de l'analyse séparée de jeux de données dont les résultats sont comparés ; à l'analyse à l'aide de modèles spécifiquement construits pour combiner simultanément plusieurs jeux de données dans un cadre formel unifié, modèles dits « intégrés » (Kéry & Royle, 2020). Intégrer formellement des données de plusieurs protocoles de suivi au sein d'une même analyse peut en effet être pertinent dans de nombreux contextes en écologie : des données de nature différente peuvent par exemple donner des informations complémentaires sur la structure et la dynamique d'une population (Schaub & Abadi, 2011) ; combiner plusieurs jeux de données peut permettre d'augmenter la couverture spatiale et temporelle de la population étudiée (Farr et al., 2020; Isaac et al., 2019; Schaub & Abadi, 2011), etc. Les deux principaux avantages des modèles intégrés sont une précision accrue 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 & Royle, 2020; Lauret, Labach, Authier, et al., 2021; Zipkin et al., 2019).

Le développement de modèles permettant l'analyse combinée de données issues de sources hétérogènes a été l'une des voies de recherche importante de l'écologie statistique des 20 dernières années. Diverses approches permettant de combiner les informations de plusieurs jeux de données ont émergé. Toutes consistent en une démarche de modélisation statistique permettant de tirer le meilleur parti de l'accumulation de données collectées

par des approches différentes sur un même système écologique (Amundson et al., 2014; Fletcher et al., 2019; Miller et al., 2019; Pacifici et al., 2019). On nomme communément cette approche *data fusion*, *data combination* ou *data integration* selon le type de données mobilisées et selon le degré de formalisation de la combinaison. Par exemple, on peut envisager plusieurs méthodes pour combiner les jeux de données A et B. La méthode la moins formelle consiste à comparer les résultats issus de l'analyse statistique séparée des jeux de données A et B, e.g. on superpose les cartes de résultats dans le cas d'une étude de la distribution et on discute des convergences/divergences entre les deux cartographies. Il ne s'agit pas au sens propre d'une combinaison dans l'analyse des données mais plutôt d'une approche de type « triangulation » (ou “*multiple lines of evidence*” en anglais (Cook et al., 2012 ; Bland et al., 2017)). Un niveau supérieur de combinaison consiste à utiliser explicitement les résultats de l'analyse du jeu de données A comme paramètre de l'analyse du jeu de données B. Le jeu de données A fourni des informations pour l'analyse du jeu de données B mais la réciproque n'est pas vraie. Par exemple, un jeu de données contenant les localisations des bateaux de pêches peut être utilisé comme variable dans un modèle estimant l'utilisation de l'espace par les grands dauphins (*Tursiops truncatus*) (Pirota et al., 2015). Enfin, le niveau d'intégration le plus avancé consiste en l'analyse simultanée des jeux de données de A et de B qui partagent des informations. Le modèle statistique est construit spécifiquement pour estimer un ou plusieurs paramètres écologiques qui sont communs aux deux jeux de données, c'est-à-dire des paramètres qui peuvent être extraits des deux jeux de données. On parle ici de modèles intégrés au sens strict (voir par exemple Quéroué et al. (2021)).

Le principe de base 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 & 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 accommodant les spécificités de chaque protocole de collecte (Fletcher et al., 2019; Zipkin et al., 2019). Ainsi, il y a un processus d'observation différent pour chaque jeu de données et protocole de collecte, e.g. des données de présence-absence seraient par exemple analysées via un modèle d'*occupancy* permettant de prendre en compte la détection imparfaite de l'espèce sur des sites (MacKenzie, 2006) alors que de données de comptages serait analysées en *distance sampling*, méthode permettant de modéliser la baisse de la probabilité de détection avec la distance à l'observateur (Buckland et al., 2005).

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. Ces développements permettent par exemple 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), voir Kéry & 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. Ils sont souvent utilisés pour combiner 1) un jeu de données important mais de qualité limitée (i.e. souvent constitué de données peu coûteuses ou faciles à collecter donc collectées avec un effort d'échantillonnage réduit ou constitué de données collectées sans protocole – sciences participatives), avec 2) un ensemble de données plus limité (dans le temps, dans l'espace), plus difficile à obtenir, mais qui contiendrait des informations de meilleure qualité. D'une part, les données fines peuvent permettre de caractériser et donc de corriger les faiblesses du jeu de données important et plus pauvre en information telles qu'un déséquilibre spatial dans l'effort d'échantillonnage (Dorazio, 2014). D'autre part, le jeu de données grossier peut s'étendre sur une échelle spatio-temporelle large au contraire du jeu de données fin et donc favoriser l'étude du système à large échelle après correction de ses biais via le jeu de données fin (Jiménez et al., 2022; Kéry & Royle, 2020; Lauret et al., 2022). Il est important de noter que le recul sur ces analyses intégrées et leur pertinence pousse aujourd'hui à anticiper la planification des suivis dans le but de mettre directement en œuvre des protocoles de collecte de données complémentaires (Zipkin & Saunders, 2018).

Box 1 : Les grands dauphins en Méditerranée française

L'étude du grand dauphin (*Tursiops truncatus*) est complexe car il s'agit d'une espèce très mobile, difficilement accessible à l'observation ([Labach, 2021](#)). En Méditerranée, les connaissances sur la population française de grand dauphin viennent principalement de deux programmes récents d'acquisition de données à large échelle. Premièrement, les campagnes de Suivi Aérien de la Mégafaune Marine (SAMM, voir section 2), collectent des données de comptages à la côte et dans les eaux pélagiques. Deuxièmement, un autre programme de collecte de données sur le grand dauphin en Méditerranée française couvre tout le plateau continental via un protocole de photo-identification des individus. Les survols aériens SAMM sont peu fréquents mais sont les seuls à échantillonner les eaux pélagiques. Les suivis côtiers effectués par photo-identification fournissent des informations plus détaillées et plus fréquentes car les programmes sont en général répétés annuellement ou mensuellement dans certaines zones. Chaque jeu de données analysé indépendamment produit des indicateurs écologiques à la précision limitée, voire parfois des résultats contradictoires ([Lauret et al., 2022](#)). Des développements méthodologiques récents ont conduit à intégrer l'analyse des données issues 1) des campagnes de survol aérien basées sur un protocole de type distance sampling ([Laran et al., 2017](#)) ; et 2) des suivis de photo-identification par bateau via un modèle de capture-recapture spatiale ([Pirodda et al., 2015](#) ; [Labach et al., 2021](#)).

Ainsi, les deux jeux de données participent aux estimations des paramètres écologiques d'intérêt que sont l'abondance et la densité de grand dauphin en Méditerranée française. Dans le modèle intégré, les données côtières détaillées apportent des informations sur la taille de la population tandis que les données de survol à large-échelle ont permis de déterminer la répartition spatiale des individus en caractérisant l'effet de la profondeur sur la densité des dauphins. L'analyse intégrée des deux sources de données a ainsi permis de réduire les limites d'échantillonnage de chaque jeu de données et a aidé à améliorer l'inférence écologique par rapport à une analyse séparée ([Lauret, Labach, Turek, et al., 2021](#)). Cette étude sur les grands dauphins en Méditerranée illustre l'intérêt de la modélisation spatiale intégrée pour réaliser des inférences écologiques à grande échelle sur les effectifs et la densité des populations.

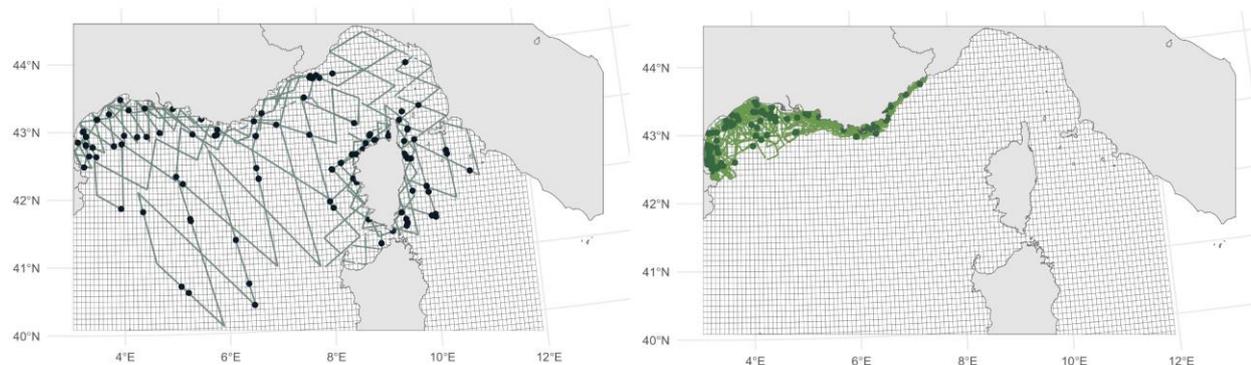


Figure box 1 : Plan d'échantillonnage des grands dauphins en Méditerranée française par les programmes SAMM et GDEGeM

Les développements méthodologiques pour la modélisation intégrée sont toujours en cours, en témoigne le nombre important de publications scientifiques dans le domaine, ou l'organisation de sessions dédiées à cette question lors des conférences internationales sur les méthodes statistiques appliquées à l'écologie. Les questions les plus récentes concernent l'inclusion des données issues de suivis télémétriques dans des modèles d'estimation d'abondance comme en témoigne le récent numéro spécial dédiée à l'intégration des modèles de déplacement aux modèles de capture-recapture spatiale dans le journal *Ecology* ([Converse et al., 2022](#)).

Concernant le problème plus spécifique de Migralion, à savoir l'utilisation du Golfe du Lion par l'avifaune, Matthiopoulos et al. (2022) ont très récemment insisté sur l'intérêt de l'analyse intégrée des données pour étudier l'utilisation de l'espace par oiseaux marins. Dans cet article, ils ont aussi passé en revue les possibilités d'analyse intégrée pour cette question et ont formulé plusieurs recommandations pour leur développement. On y retrouve l'importance de prendre en compte la détection imparfaite des individus ou des espèces, maintenant bien considérée dans la plupart des modèles statistiques utilisés en écologie (MacKenzie, 2006), ainsi que la nécessité de tenir compte des phénomènes de densité-dépendance lors de la modélisation de la dynamique des populations d'oiseaux marins. Les auteurs suggèrent aussi d'utiliser des approches de statistiques bayésiennes et des modèles hiérarchiques (voir sections suivantes pour des détails). Détection imparfaite, densité-dépendance, modélisation hiérarchique, et inférence Bayésienne sont de fait tous des aspects que nous maîtrisons et que nous avons décidé d'inclure dans les méthodes que nous proposons pour Migralion. De plus, cet article présente les perspectives de développement méthodologique à envisager pour le futur de la modélisation de l'utilisation de l'habitat des oiseaux marins. Il est souligné l'importance d'intégrer les données issues de suivis télémétriques avec les données classiquement utilisées pour estimer abondance et dynamique des populations, *i.e.* les données issues de comptages et issues de suivis de populations de type "capture - recapture", vu l'abondance de ces types de données chez les oiseaux marins (Matthiopoulos et al., 2022). Cet aspect constitue un des principaux défis que nous avons identifiés pour Migralion et qui sera prochainement traité dans le lot 6 (voir section 4), ce qui réaffirme la cohérence des approches que nous avons envisagé avec l'actualité de la littérature scientifique sur le sujet. D'autres études suggèrent aussi d'aller vers des développements méthodologiques complexes comme celle visant à la modélisation explicite de l'autocorrélation résiduelle par un champ gaussien (Martino et al., 2021), ou comme celle visant à lier télémétrie et sélection d'habitat en réconciliant les concepts de *step selection function* et d'*habitat selection function* (Michelot et al., 2019, 2020). Bien qu'attrayantes dans le cadre de Migralion, ces deux perspectives sont extrêmement ambitieuses en termes de développement méthodologique. Ce sont aujourd'hui des méthodes peu maîtrisées dans la littérature scientifique et nécessitant d'importants moyens de calcul. Ces deux approches ne semblent donc pas être des candidats idéals dans un travail de modélisation intégrant déjà des difficultés méthodologiques importantes et spécifiques. Ces questions pourront malgré tout être discutées avec le post-doctorant recruté pour l'analyse des données de télémétrie dans le lot 3.

Dans le cadre de Migralion, nous nous inspirons donc largement de méthodes d'analyse séparées des données déjà publiées dans la littérature scientifique tout en les adaptant aux spécificités de notre cas d'étude. En effet, la combinaison de certaines données nécessite des développements méthodologiques qui n'ont jamais été publiés, e.g. combiner télémétrie et données radar, et constitue donc un challenge déjà important. Dans la suite de ce rapport, nous passons en revue les différents jeux de données que nous mobilisons ainsi que les méthodologies pressenties pour leur analyse intégrée.

II. Description des données mobilisables

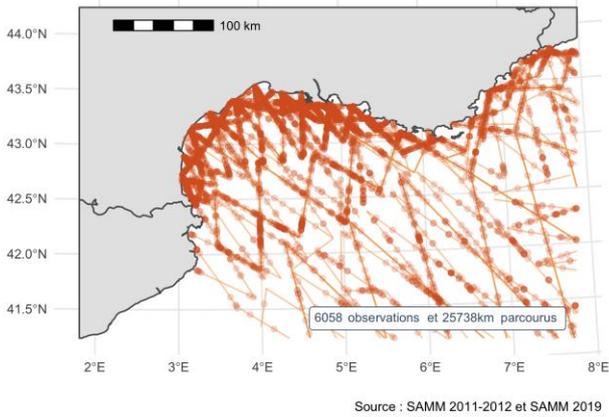
A. Données historiques

Le lot 1 porté par La Tour du Valat a consisté en un recensement des données disponibles sur l'avifaune dans le Golfe du Lion en amont des collectes de Migralion. Ce travail a permis de constituer un état des lieux des programmes de collecte mis en oeuvre jusque-là et dans certains cas les données ont pu être récupérées. Dans le cadre du lot 6, nous avons examiné le type de données récoltées par le lot 1 et identifié les jeux de données *a priori* pertinents pour nos travaux de modélisation intégrée. Dans cette section, nous présentons les jeux de données récupérés et mobilisables.

1. SAMM

Les Survol Aérien de la Mégafaune Marine **SAMM** réalisés par l'Observatoire Pelagis et l'Université de La Rochelle constituent un programme de suivi de la mégafaune marine et des activités en mer unique par son ampleur (Laran et al., 2017). Des transects de lignes en avion ont été réalisés sur l'ensemble de la Zone Économique Exclusive française en 2011-2012, et une deuxième fois en 2018-2019 pour la Méditerranée. Sur ces transects, des données écologiques (cétacés, oiseaux marins, thons), ainsi que sur les activités humaines sont relevées systématiquement. Le protocole d'observation est de type Distance Sampling : deux observateurs relèvent l'espèce détectée, ainsi que la distance perpendiculaire à l'avion. Une trentaine d'espèce d'oiseaux marins a été détectées en Méditerranée avec plusieurs centaines d'observations pour les espèces les plus communes sur les deux campagnes (Figure 1). Les suivis sont réalisés en Novembre, Décembre, Janvier, et Février pour la campagne hivernale ; et en Mai, Juin, Juillet et Aout pour la campagne estivale. Les dates ne couvrent pas les périodes migratoires et les espèces terrestres de petite taille ne sont pas identifiables par ce type de protocole. Elles fournissent par contre des informations pertinentes concernant les oiseaux marins résidant toute l'année dans le Golfe du Lion.

SAMM : Transects et observations d'oiseaux marins
 Survol aérien en transect de ligne



SAMM : Espèces détectées

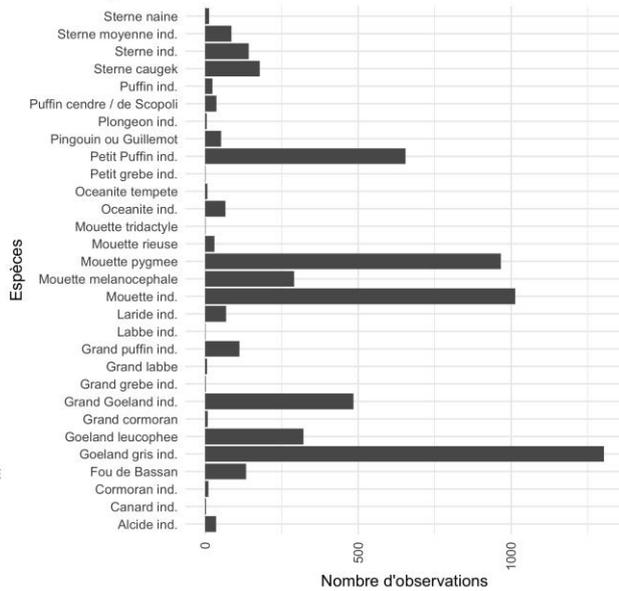
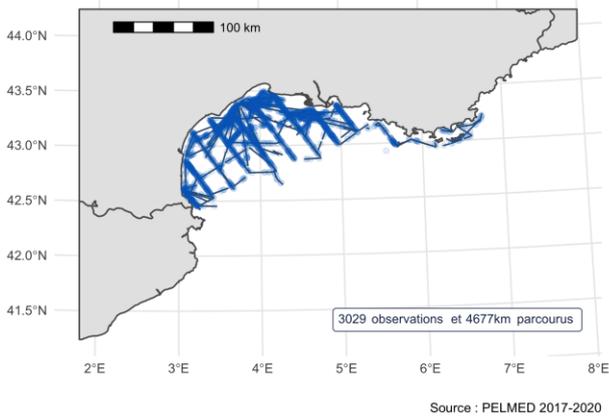


Figure 1 : Campagnes de Survol Aérien de la Mégafaune Marine (SAMM), plan d'échantillonnage spatial et fréquence des espèces observées

2. PELMED

Les campagnes annuelles **PELMED** sont menées par l'IFREMER de Sète et visent à l'évaluation des ressources halieutiques des poissons pélagiques de Méditerranée (Jean-Hervé, 1985). Tout au long de la campagne qui se déroule entre Juin et Septembre, les observateurs observent et comptent aussi les oiseaux et mammifères marins. Les campagnes PELMED quadrillent le Golfe du Lion en faisant des transects de lignes en bateau chaque année depuis 1985 (Figure 2).

PELMED : Transects et observations d'oiseaux marins
 Campagne en bateau par transect de ligne



PELMED : Espèces détectées

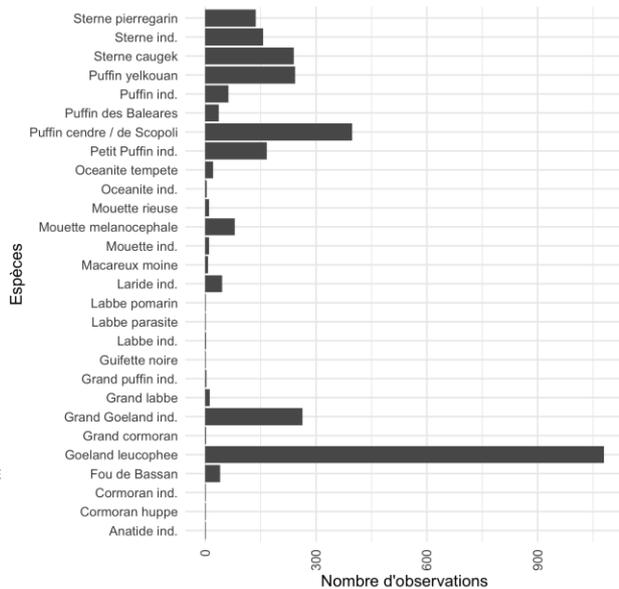


Figure 2 : Campagnes PELMED, plan d'échantillonnage spatial et fréquence des espèces observées

3. Fermes pilotes

Dans le cadre des études préliminaires à l'installation de fermes éoliennes pilotes, Biotope réalise des suivis scientifiques. Des transects en bateau et en avion ont ainsi permis de récolter de nombreuses observations d'oiseaux marins. Trois études ont été réalisées : **Provence Grand Large (PGL)**, et **EFGL, EOLMED**. Les sollicitations ont été faites auprès de l'OFB et des différents partenaires industriels propriétaires des données pour les obtenir. A l'heure de l'écriture de ce rapport, nous ne les avons pas reçues et ne pouvons donc pas en dire plus sur leur intérêt pour l'analyse intégrée.

4. Parc Naturel Marin du Golfe du Lion

Depuis 2019, les agents du Parc Naturel Marin du Golfe du Lion effectuent des campagnes en mer (transects en bateau) visant à récupérer des données concernant les activités humaines et la mégafaune marine. Les campagnes ont principalement eu lieu en Novembre et à la fin du printemps (Mai, Juin, Juillet). Le PNM est particulièrement étendu pour une Aire Marine Protégée (AMP) et la surface couverte par ces suivis est complémentaire de celle couverte par les études des fermes pilotes (Figure 3). Les données ont été récupérées auprès du Parc Naturel Marin.

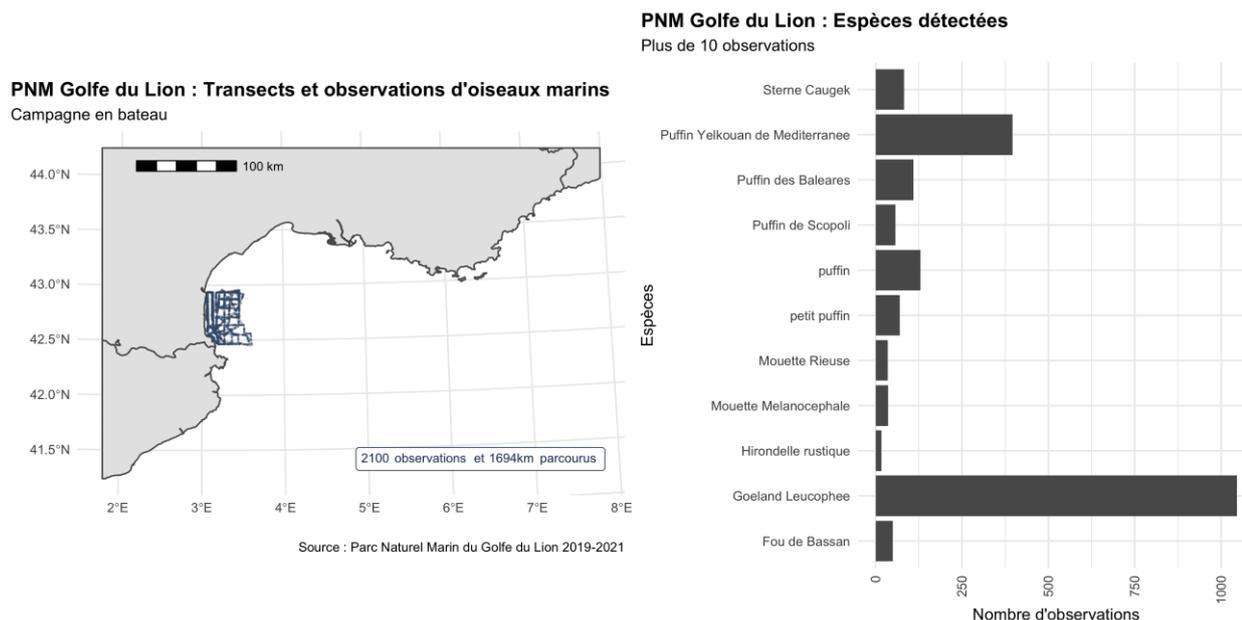


Figure 3 : Campagnes MégaObs du Parc Naturel Marin (PNM) du Golfe du Lion, plan d'échantillonnage spatial et fréquence des espèces observées

5. Recensement des colonies du GISOM

Tous les dix ans, le Groupement d'Intérêt Scientifique Oiseaux Marins (GISOM) effectue un recensement exhaustif des colonies d'oiseaux marins et de laro-limicoles nicheurs à l'échelle nationale. De plus, les

programmes [LIFE+ ENVOLL](#), et [LARIMED](#) ont permis d’avoir des recensements annuels sur le pourtour méditerranéen pour certaines espèces (sternes, mouettes, certains goélands comme le goéland railleur). Nous connaissons ainsi la localisation et la taille des colonies, des informations qui sont sans aucun doute déterminantes dans l’utilisation du Golfe du Lion par les oiseaux marins et qui pourront donc être mobilisées dans l’analyse intégrée (Figure 4).

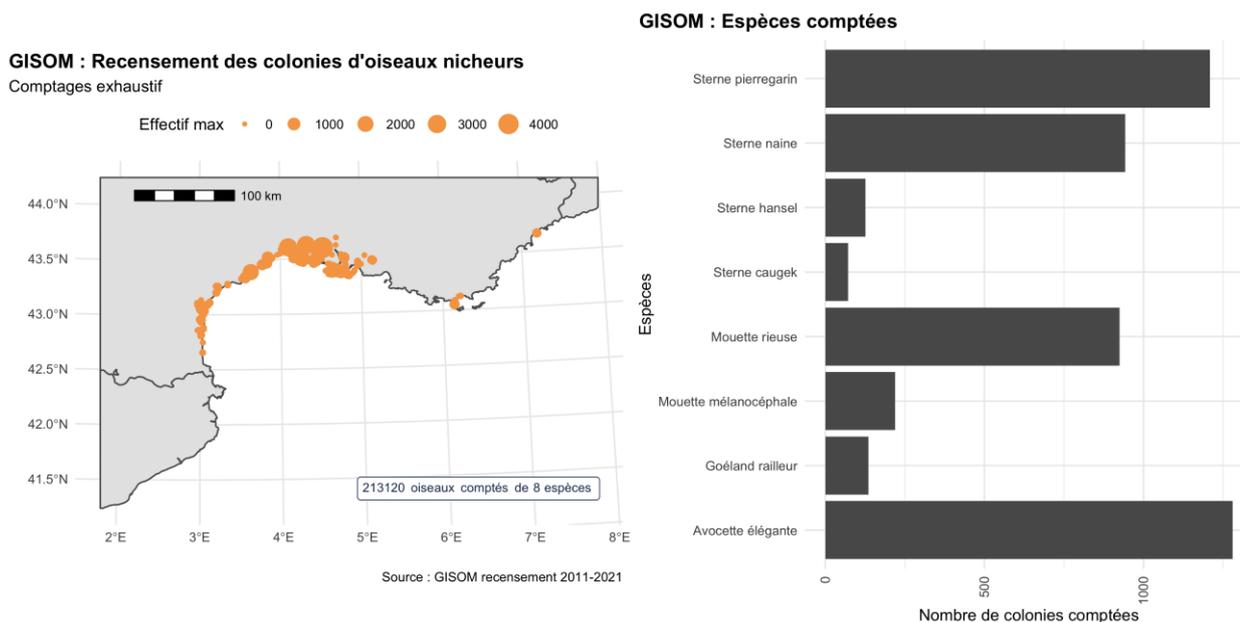


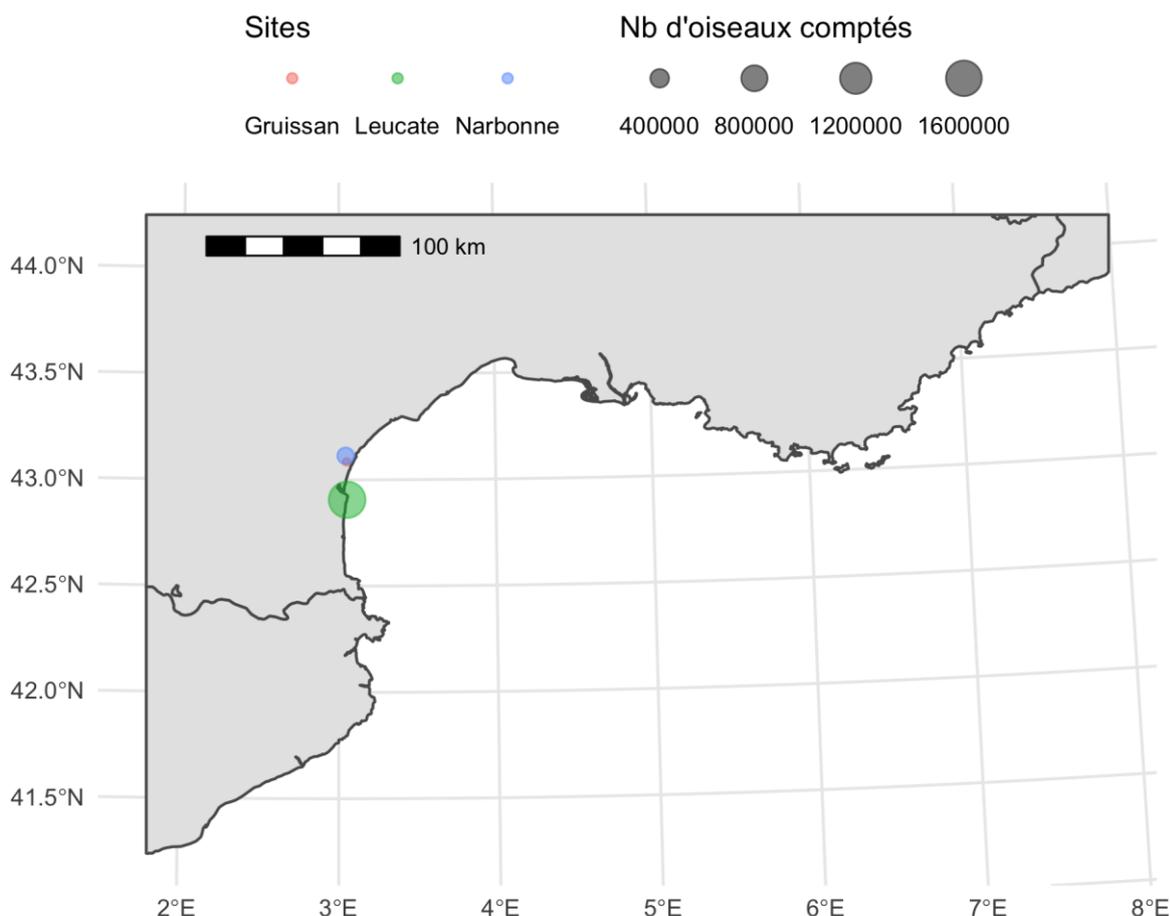
Figure 4 : Campagnes de recensement des colonies de laro-limicoles 2011-2021, plan d’échantillonnage spatial et fréquence des espèces observées

6. Comptages côtiers pendant les périodes migratoires

Le lot 1 de Migralion a recensé plusieurs sites de comptages côtiers des flux d’oiseaux migrateurs le long du littoral du Golfe du Lion. Les données sont récoltées par des associations principalement durant la période pré-nuptiale (voir le livrable du lot 1 pour plus de détail). Plusieurs jeux de données ont pu être récoltés et explorés pendant le début du lot 6. Pour chaque espèce, les observateurs comptent le nombre d’individus survolant le site de comptage. Récemment, la principale production de données de ce type, par son ampleur et sa qualité, correspond au site situé sur la falaise de Leucate qui a fourni trois saisons complètes de comptages journaliers en 2020, 2021, et 2022. Un autre site important correspond à l’étang de Canet mais les données n’ont pas été récupérées pour ce dernier site (Figure 5).

Comptages côtiers de la migration

Données récupérées depuis 2018



Source : LPO Aude, MedMigration

Figure 5 : Localisation des sites de comptages de migration explorés à la date de Septembre 2022

7. Recensement des études télémétriques du CEREMA

Sollicité par la Direction Générale de l'Énergie et du Climat (DGEC) du Ministère de la Transition Énergétique et Sociale, le Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement (CEREMA) a effectué un travail de récupération des données de suivis télémétriques acquises en Méditerranée depuis décembre 2011, en cherchant l'exhaustivité. Les données ont été collectées sur la période de décembre 2011 à décembre 2020 et portent sur 20 espèces d'oiseaux marins et non-marins utilisant le Golfe du Lion (CEREMA, 2021). Pendant Migralion, nous aurons la possibilité de solliciter l'utilisation des données provenant des études télémétriques présentées dans le rapport.

Voir le [portail internet](#) pour visualiser les données disponibles.

8. Autres données issues des suivis télémétriques disponibles

Les porteurs des lots 1, 3, et 6 sont en contact avec d'autres programmes de télémétrie sur des espèces non couvertes par Migralion (e.g. rapaces, œdicnèmes) qui utilisent le Golfe du Lion pour leurs trajets migratoires. Conjointement avec les porteurs du lot 3, nous prendront contact avec les propriétaires de ces jeux de données pour les récupérer et les inclure dans les analyses de Migralion.

B. Données Migralion

A la date d'écriture de ce rapport, voici les données récoltées dans le cadre de Migralion disponibles pour utilisation dans le lot 6.

1. Télémétrie

- 9 Spatules blanches (*Platalea leucorodia*) ont été équipées de GPS en 2020, [trajets sur Movebank](#).
- 3 flamant roses (*Phoenicopterus roseus*) ont été équipés de GPS en 2020.
- 22 Sternes caugek (*Thalasseus sandvicensis*) ont été équipées de GPS en 2021 et 2022, [trajets sur Movebank](#).
- Plusieurs dizaines de GLS (*Geolocators Light Sensors*) ont été déployés mais les données ne seront disponibles qu'après recapture de l'individu lors de la prochaine saison de nidification, c'est-à-dire lors des années 2023-2024 au plus tôt pour avoir un nombre de données suffisant pour en faire l'analyse.

2. Radar

Les données radars côtiers et radars en mer issues de la première campagne de migration pré-nuptiale ont été récoltées par Biotope dans le cadre des lots 4 et 5 et sont en cours de traitement. Nous n'avons pas eu accès aux résultats pour l'instant. Nous devons nous réunir avec les porteurs des lots 4 et 5 durant l'Automne pour une présentation détaillée des résultats des premières campagnes. Nous discuterons à cette occasion de la répartition des analyses qui devront être réalisées dans un premier temps par Biotope avec les données radars seules, et dans un second temps par le lot 6 en combinaison avec les autres jeux de données (télémétrie, comptages).

III. La séparation entre oiseaux marins et migration

Les premières réflexions sur les données disponibles et sur les plans d'échantillonnages actuels de Migration 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 résident toute l'année autour du Golfe du Lion et utilisent la zone continuellement, ce qui permet, par exemple, de considérer des jeux de données collectés toute l'année, y compris hors des périodes migratoires (e.g. PELMED, SAMM). Les oiseaux marins utilisent le Golfe du Lion pour se nourrir, se déplacer entre les colonies de reproduction et les zones d'alimentation. Nous avons à faire à 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, nous chercherons donc à estimer une utilisation de l'espace qui englobe l'ensemble des activités des individus étudiés. A l'opposé, les oiseaux migrateurs utilisent le Golfe du Lion 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 du Golfe du Lion par les migrateurs restreint les périodes de collecte et les jeux de données disponible 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 ultime de l'utilisation du Golfe du Lion 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 un cas, nous cherchons à répondre à la question « **où sont les oiseaux marins ?** » ; et dans l'autre, « **où et quand passent les migrateurs ?** ». Dans les sections suivantes, nous détaillons les approches méthodologiques envisagées pour répondre à ces questions à partir de la combinaison des différents jeux de données (voir aussi Table 1).

Table 1 : Questions écologiques et données disponibles pour l'études des oiseaux marins et des migrateurs terrestres dans le cadre du lot 6 de Migralion. Le niveau de définition des données récoltées est noté entre parenthèses dans la dernière ligne du tableau. Par exemple, les comptages en mer permettent de récolter des données au niveau de l'espèce quand les radars renseignent au niveau du groupe d'espèces

	Oiseaux marins	Migrateurs terrestres
Ecologie dans le Golfe du Lion	Résident toute l'année dans le Golfe du Lion Nichent en colonies et se nourrissent en mer.	Utilisent le Golfe du Lion pour leur migration pré-nuptiale (Fev - Mai) et post-nuptiale (Aug - Oct). Zones de passage et effectifs mal connus, certains longent le trait de côte, d'autres traversent en mer.
Sources des jeux de données disponibles	Historique : SAMM, PELMED, PNM Golfe du Lion, Fermes pilotes, GISOM, GPS Migralion : Balises GPS, campagnes en mer	Historique : Comptages côtiers, balises GPS, GLS Migralion : Radars en mer, radars à la côte, balises GPS, GLS
Type de données	Comptages en mer (espèce) <i>Migration Traffic Rate</i> obtenues par radar (groupe de taille) Recensement exhaustif de l'effectif nicheur et de l'emplacement des colonies (espèce) Trajet des individus (individu)	<i>Migration Traffic Rate</i> : Nombre d'individus par unité de temps, avec altitude de vol (groupe de taille) Trajets des individus à différents niveaux de résolution : précis par GPS, dégradé par GLS (individu)
Estimation écologique	Carte de probabilité de l'utilisation de l'espace dans le Golfe du Lion, à l'échelle de l'espèce ou du groupe d'espèce	(Idéalement) carte des flux par unité de temps, à l'échelle du groupe d'espèce

IV. Oiseaux marins

A. Méthodologie envisagée

Les oiseaux marins étant résidents toute l'année dans la zone étudiée, plusieurs hypothèses des modèles classiques d'abondance sont respectées, comme par exemple celle concernant la fermeture de l'aire d'étude pendant la période d'échantillonnage avec peu d'oiseaux qui en sortent et qui y rentrent (hormis les variations démographiques qui sont faibles pour des espèces longévives comme les oiseaux marins, Wooller et al. (1992) ; Dobson & Jouventin (2010)). L'étude des oiseaux marins se faisant à partir de jeux de données collectés toute l'année, on pourra modéliser l'utilisation du Golfe du Lion toute l'année durant, et potentiellement explorer la saisonnalité de cette utilisation. Pour Migralion, les données disponibles sur ce groupe d'espèces se résument pour l'instant en deux catégories principales : les trajets GPS issus du lot 3, et les données de comptages en mer (issus des campagnes historiques et du lot 4 de Migralion). Le recensement et la localisation des effectifs nicheurs pourront aussi être utilisés comme des variables du modèle pouvant informer les phénomènes de densité-dépendance (Matthiopoulos et al., 2022). Nous considérerons aussi l'apport des données radar des lots 4 et 5 qualitativement dans un premier temps, et nous réfléchirons à une potentielle intégration formelle dans les modèles d'utilisation de l'espace bien qu'aucune piste ne se dégage pour l'instant car il n'est pas certain de pouvoir distinguer les oiseaux marins parmi les échos radar obtenus.

Premièrement, les balises GPS fournissent des informations détaillées (stratégie d'alimentation, de déplacement, de repos) à une résolution fine spatialement et temporellement et sur plusieurs mois mais seulement sur une poignée d'individus d'une espèce. En effet, sur les milliers d'individus que compte la population de sternes caugek (*Thalasseus sandvicensis*) dans le Golfe du Lion, seulement une vingtaine est équipée, et uniquement sur quelques colonies dont celles de l'Etang de Thau. Nous avons donc des informations précises sur l'utilisation de l'espace grâce à ce type de données mais sur un échantillon réduit d'individus de la population. D'autre part, les données de comptages en mer ont couvert l'ensemble du Golfe du Lion mais seulement pendant une période restreinte de l'année avec souvent un seul passage par zone. Le principal défi de l'intégration de données pour les oiseaux marins réside donc dans la combinaison entre données de télémétrie et comptages en mer, deux types d'information fondamentalement différents par nature.

Plusieurs développements méthodologiques récents ont émergé pour intégrer l'analyse de données issues de la télémétrie et issues de comptages (Matthiopoulos et al., 2022) que nous explorons actuellement pour les appliquer dans le cadre de Migralion. Naturellement, nos travaux se concentrent sur les espèces d'oiseaux marins pour lesquelles des données de GPS sont disponibles actuellement (e.g. Sterne caugek, Puffin de Yelkouan - *Puffinus yelkouan*, Puffin de Scopoli - *Calonectris diomedea*) mais la méthode développée a pour vocation à être ensuite généralisée à l'ensemble des espèces pour lesquelles les données seront suffisantes.

B. Intégration des données historiques et télémétrie : RSF, *occupancy*

Pour développer la méthode permettant l'intégration des données de télémétrie et de comptages, nous avons constitué un groupe de travail formé de Nicolas Courbin, David Grémillet, Charlotte Lambert, Valentin Lauret, et Aurélien Besnard. Ensemble, nous avons identifié plusieurs pistes de modélisation potentielles. Comme décrit dans la Section 1, les modèles intégrés prennent souvent la forme de modèles hiérarchiques. Nous utilisons ce terme de **modèle hiérarchique** pour décrire un ensemble couplé de modèles ordinaires qui sont conditionnellement liés les uns aux autres (Kéry & Royle, 2016). Par conditionnel, nous entendons que la distribution de probabilité d'une variable aléatoire dépend de l'autre variable aléatoire, généralement notée $f(y|z)$ ou $[y|z]$, indiquant que les résultats de la variable aléatoire y dépendent du résultat de la variable aléatoire z . Généralement en écologie, les modèles hiérarchiques contiennent une composante pour les observations (ou données), puis une composante pour décrire les variables latentes ou les résultats d'un processus écologique (Kéry & Royle, 2016). Dans les modèles intégrés, nous avons plusieurs modèles d'observation, un pour chaque type de données Kéry & Royle (2016). Parmi les modèles hiérarchiques couramment utilisés en écologie, on trouve les modèles d'*occupancy* (Mackenzie et al., 2002), les modèles *N-mixture* (Royle et al., 2004), le *distance sampling* (Silllett et al., 2012), et bien d'autres (Kéry & Schaub, 2012).

1. Processus latent

Le processus écologique latent décrit, dans notre cas, l'utilisation de l'espace par l'espèce d'oiseau marin considérée, utilisation que nous modélisons en fonction de variables environnementales. On appellera λ la probabilité d'utilisation de l'espace qu'on estime à l'aide d'une régression logistique sur les C variables environnementales *habitat*.

$$\text{logit}(\lambda) = \alpha_0 + \sum_i^C \alpha_i \text{habitat}_i$$

Pour appliquer ce modèle au cas des oiseaux marins du Golfe du Lion, nous définissons comme zone d'étude une aire allant du Cap Creus au Sud jusqu'à la frontière italienne à l'Est. Cette zone d'étude est ensuite divisée en nombreuses cellules (ou mailles) conjointes (Figure 6). L'estimation des paramètres écologiques se fait à l'échelle de chaque cellule, on estime donc un λ_n différent pour chacune des N cellules de la grille. La résolution de la grille (correspondant au nombre et à la taille des cellules) est à définir en fonction du nombre de données à analyser (disposer de beaucoup de données permet de faire un maillage fin avec beaucoup de cellules) et du temps de calcul acceptable (les temps de calculs augmentent fortement avec le nombre de cellules).

Zone d'étude pour le lot 6 de Migralion

Golfe du Lion

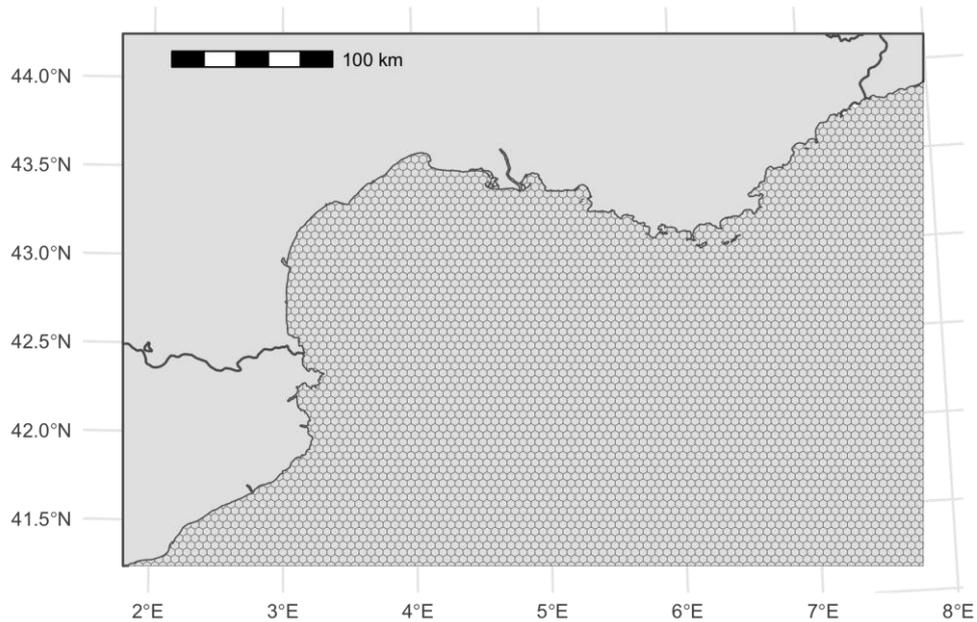


Figure 6 : Exemple de grille utilisable pour la modélisation de l'utilisation de l'espace par les oiseaux marins

L'utilisation de l'espace est décrite en chaque cellule de la grille avec $z = 1$ si l'espèce utilise la cellule, et $z = 0$ si l'espèce n'utilise pas la cellule. Le paramètre z est modélisé par un tirage dans une distribution de Bernoulli de paramètre λ .

$$z \sim \text{Bernoulli}(\lambda)$$

2. Processus d'observation : comptages en mer

Les données relevées lors des campagnes de comptages en mer sont résumées en présence-absence pour être liées à λ via un processus d'observation de type *occupancy* (MacKenzie, 2006). Les données récoltées dans chaque cellule échantillonnée, c-à-d les présences notées $y = 1$ et les absences notées $y = 0$ sont modélisées via une distribution de Bernoulli de paramètre $z p$, où p est la probabilité de détecter l'espèce sachant qu'elle utilise la cellule.

$$y \sim \text{Bernoulli}(z p)$$

La probabilité de détection peut être estimée de plusieurs manières : en fonction de la distance d'observation par rapport au transect lorsque que cette information est notée, comme dans le cas du *Distance Sampling* (Buckland et al., 2005) pour les données SAMM ; en fonction du nombre de prospections par cellule dans le cas des modèles d'*occupancy* (MacKenzie, 2006) ; ou en fonction d'autres variables pouvant affecter la détection (e.g. la distance parcourue dans chaque cellule, les conditions météorologiques).

3. Processus d'observation : télémétrie

L'utilisation des données GPS pour décrire l'utilisation de l'habitat est abondamment documentée dans la littérature scientifique qui fait état de plusieurs méthodes différentes (Aarts et al., 2008; Johnson et al., 2008). Les données de télémétrie vont permettre de définir une fonction de sélection de l'habitat (*Habitat Selection Function* / *Resource Selection Function*, Aarts et al. (2008); Fieberg et al. (2021); McLoughlin et al. (2010)). Les localisations de GPS (codées $t = 1$) sont analysées simultanément à des points de « pseudo-absences » (codés $t = 0$) c'est-à-dire des localisations tirées au hasard dans la zone d'étude. La modélisation de ces données binomiales (0/1) se fait via un tirage dans une loi de Bernoulli de paramètre λ , le paramètre commun décrivant la probabilité d'utilisation de l'espace.

$$t \sim \text{Bernoulli}(\lambda)$$

L'idée de ces méthodes repose sur la comparaison de l'habitat utilisé par les individus aux localisations relevées par les GPS, avec des localisations dites « nulles » caractérisant des zones de l'habitat non-utilisées par les individus. Il existe plusieurs façons de générer ces localisations « nulles » à considérer pour le modèle de sélection d'habitat (Barbet-Massin et al., 2012; Johnson et al., 2008; McLoughlin et al., 2010), et à ce jour, plusieurs méthodes sont envisagées pour Migralion.

- Comme première possibilité, une étude de l'autocorrélation spatio-temporelle des localisations issues des suivis GPS permet de définir un pas de temps auquel sous-échantillonner les localisations pour que celles-ci constituent des étapes de sélections d'habitat indépendantes. Ensuite, pour chaque point GPS, plusieurs points nuls (« pseudo-absence ») sont générés dans la zone d'étude (Trevail et al., 2021).
- En deuxième option, il est possible de garder l'intégralité des trajets GPS récoltés. Pour chaque trajet réalisé par un individu, 10 trajets fictifs sont simulés en suivant les mêmes caractéristiques (distance, *turning angle*, etc) et en partant du même point que les trajets réels. Les trajets simulés ne suivent alors aucune sélection de l'habitat disponible et peuvent ainsi être considérés comme des modèles « nuls » dans l'analyse (Courbin et al., 2018; Raymond et al., 2015).
- Une troisième possibilité permet, là encore, de garder l'intégralité des trajets collectés par GPS en générant des points de pseudo-absence via le déplacement des trajets ou des couches d'habitat relativement l'un à l'autre (Cushman, 2010). Une approche similaire fait pivoter les trajets GPS autour d'un barycentre situé arbitrairement dans la zone d'étude. Cette rotation permet d'échantillonner de nouvelles localisations « nulles » aux différents endroits où les trajets sont confrontés à de nouveaux habitats.
- Quatrièmement, les données de télémétrie peuvent être mobilisées dans une *integrated step selection analysis* où la sélection de l'habitat s'opère à chaque « pas » en fonction de l'habitat disponible défini par les caractéristiques du trajet et de variables d'habitat. Cette méthode récemment développée présente un cadre d'analyse complexe et très gourmand en ressources de calcul pour combiner l'estimation d'un

modèle de déplacement et de sélection d'habitat à partir de données GPS (Avgar et al., 2015; Michelot et al., 2019).

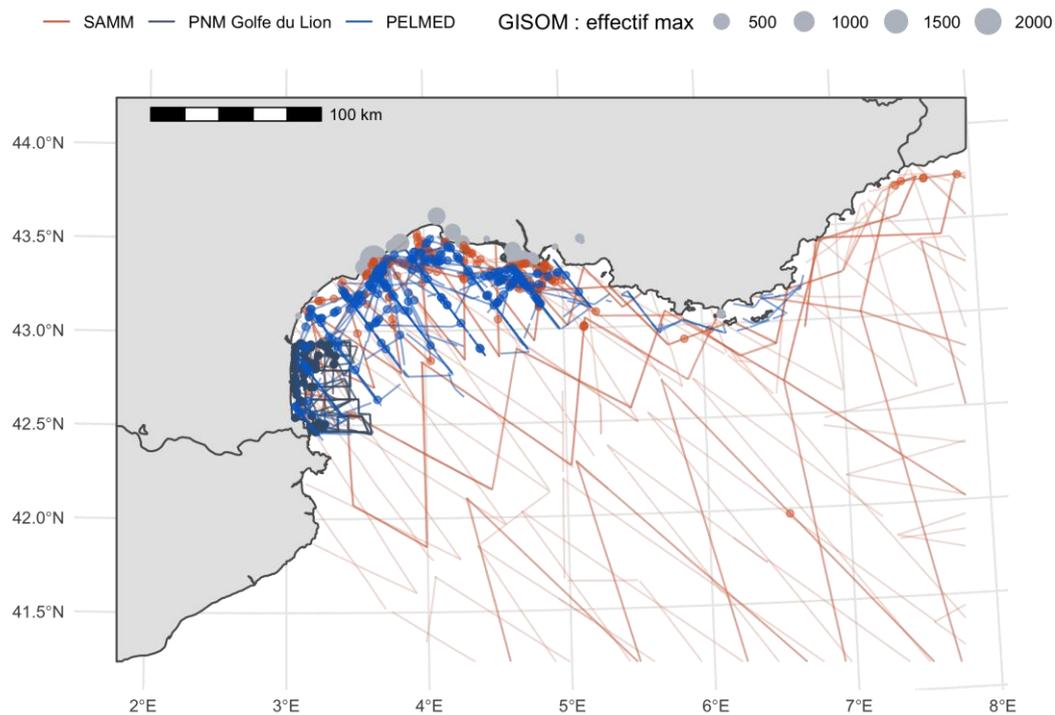
Le travail de modélisation des données GPS seules sera réalisé par le/la postdoctorant·e recruté·e par le lot 3 qui commencera à l'automne 2022. Nous menons les réflexions ci-dessus sur le choix de ce modèle nul en étudiant les différentes possibilités afin que le choix de la méthode soit explicite lors du début du contrat de postdoc du lot 3. Une fois la méthode d'analyse identifiée, les porteurs du lot 3 seront en charge de la réalisation de ce modèle de sélection d'habitat que nous nous chargeront de combiner avec les données de comptages.

C. Ecriture du modèle pour la sterne caugek

Comme premier cas d'étude nous avons choisi la sterne caugek au vu des données disponibles déjà abondantes. Dans un modèle intégré, nous avons intégré les données de comptages issues des campagnes PELMED, Megaobs, et SAMM, et les trajets de télémétrie issue du lot 3 (Figure 7).

Sterne caugék : données de comptages

Transect et détections



Sternes caugék : traces GPS

22 individus équipés

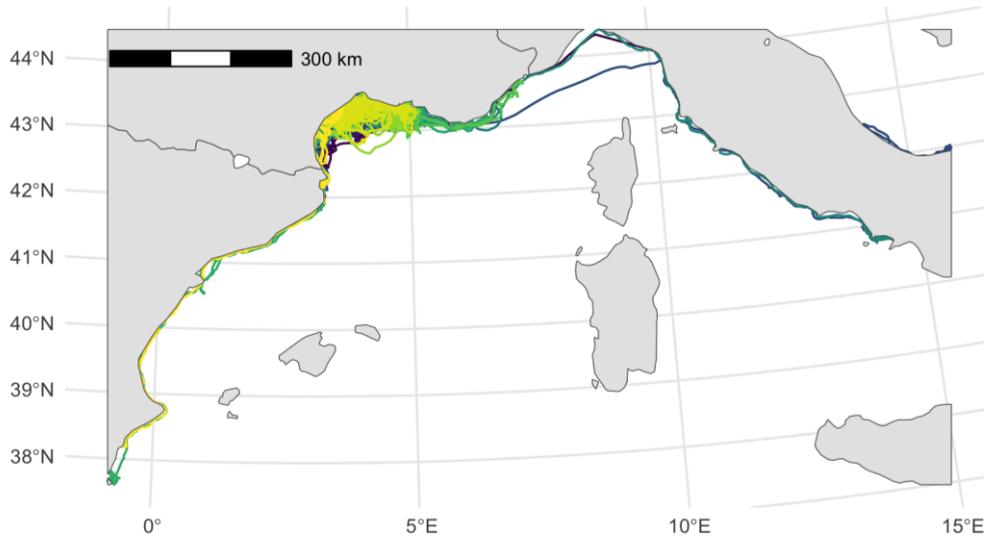


Figure 7 : Données de comptages et de télémétrie disponibles sur les Sternes caugék

A ce jour, le « squelette » du modèle intégré est construit et permettra de réaliser les véritables analyses écologiques en coopération avec le lot 3 durant l'automne. Les codes sont [disponibles sur Github](#).

Une fois, les travaux sur les données de télémétrie réalisés par le lot 3, nous intégreront les deux types de données via le modèle explicité ci-dessus pour obtenir des cartes d'utilisation de l'espace.

D. Oiseaux marins et altitude de vol

Bien que les données de comptages (e.g. SAMM, PELMED) ne renseignent pas l'altitude de vol, les données GPS issues des oiseaux marins équipés pourront nous apporter des informations. Il sera possible d'analyser l'altitude de vol des oiseaux marins équipés en fonction de variables environnementales et météorologiques (voir sections suivantes). Les données issues des radars embarqués pendant les campagnes en mer fourniront aussi les altitudes de vol. Néanmoins, il n'est pas encore certain que le niveau de définition des données soit suffisant pour distinguer les oiseaux marins des migrateurs terrestres, ce qui rendrait inutilisable ces données pour une analyse spécifique aux oiseaux marins.

Si toutefois les données le permettent, nous réaliserons une analyse intégrée combinant les altitudes mesurées par GPS avec celles mesurées par radar. L'altitude de vol z sera modélisée via un modèle linéaire généralisée 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 et par radar informeront le même processus écologique décrivant l'altitude de vol z .

$$\log(z) = \beta_0 + \beta_1 \text{météo} + \beta_2 \text{topographie}$$

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. On pourra estimer l'altitude de vol moyenne ainsi qu'une variance associée (Ross-Smith et al., 2016). Plus simplement, on pourra aussi diviser les altitudes de vol potentielles 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 (voir section suivante). Dans l'idéal, il sera possible d'obtenir une prédiction des altitudes de vol en fonction des conditions atmosphériques et de paramètres géographiques.

V. Flux migratoires

Pour avoir une connaissance parfaite du passage migratoire dans le Golfe du Lion, il faudrait connaître le nombre d'oiseaux en chaque point de la zone d'étude, à chaque instant, et à chaque altitude. Lorsqu'une couverture globale de la zone d'étude dans le temps et l'espace est assurée, il est possible de modéliser explicitement les flux migratoires en suivant les groupes d'oiseaux pendant leur passage dans la zone d'étude, i.e. on détecte l'entrée d'un groupe d'oiseaux, on suit son déplacement, puis on note sa sortie de la zone d'étude. C'est l'approche utilisée pour décrire les flux migratoires terrestres avec les radars météorologiques (Nilsson et al., 2019; Nussbaumer et al., 2021). Dans le cas de Migralion, il est naturellement impossible d'obtenir une mesure exhaustive du nombre d'oiseaux à chaque instant dans tout le Golfe du Lion. La modélisation explicite des flux est illusoire au vu de la couverture du Golfe du Lion par les données acquises. De plus, dans le contexte de planification éolienne, l'estimation des flux absolus n'est pas indispensable pour identifier les zones de risque les plus importantes ou hiérarchiser le risque associé à différentes zones. Ainsi, nous pressentons que la plupart des estimations se feront sur la base de flux relatifs et non de flux absolus. Nous pourrions identifier les zones et périodes de plus fort passage relativement à l'ensemble de la zone mais probablement pas estimer un effectif précis d'oiseaux en migration. Les flux migratoires que nous estimons ici se rapportent donc plus précisément à **la densité d'utilisation de l'espace au cours du temps**.

Estimer la densité d'utilisation de l'espace par les oiseaux en migration nécessite de prendre en compte plusieurs dimensions spatio-temporelles dans lesquelles les flux migratoires varient. Tout d'abord, il faut tenir compte de la spatialisation des flux dans le Golfe du Lion car les passages migratoires ne seront pas identiques entre Banyuls et la Camargue par exemple. De plus, dans le contexte éolien, la question de l'altitude de vol des oiseaux est de première importance pour estimer les potentiels risques de collision. Enfin, la phénologie de la migration doit être considérée explicitement car l'intensité des flux migratoires varie aussi avec le temps. Si on résume, il faudrait donc décrire les flux migratoires en quatre dimensions, x, y, z, t :

- la latitude (y) couvrant tout le Golfe du Lion
- la longitude (x) couvrant tout le Golfe du Lion
- l'altitude de vol (z) entre le niveau du sol ou de la mer et une altitude maximale à définir.
- le temps (t) s'étalant du début à la fin de la période migratoire (probablement défini comme la période de fonctionnement des radars à la côte)

De ce fait, notre objectif est de modéliser les flux migratoires en quatre dimensions en intégrant les jeux de données disponibles. Un tel modèle constituerait l'idéal à atteindre pour décrire au mieux les flux migratoires. Nous allons présenter dans la section suivante l'apport des différents jeux de données à chaque "dimension" puis, détaillerons la formulation mathématique d'un tel modèle "4D". Plusieurs modèles seront construits pour chaque groupe d'espèces au niveau de définition le plus fin mesuré par les radars (e.g. passereaux, oiseaux d'eau, grands

oiseaux). La métrique écologique décrivant la densité d'utilisation par les oiseaux migrateurs par unités de temps et d'espace sera appelée $\mu_{x,y,z,t}$, en fonction de la longitude x , de la latitude y , de l'altitude z , et du temps t .

Les données récoltées ne couvrant pas l'entièreté du Golfe du Lion pendant l'ensemble de la période, il s'agira de modéliser les flux migratoires en fonction de covariables explicatives pour pouvoir prédire l'intensité des flux dans les zones non couvertes et pendant les périodes non-échantillonnées. Nous modéliserons le flux migratoire μ en fonction de plusieurs paramètres :

- Environnementaux : incluant des variables abiotiques (e.g. distance à la côte, profondeur). Les variables biotiques (e.g. disponibilité en ressources) sont peu déterminantes pendant les déplacements en migration sur une courte période temporelle.
- Météorologiques : direction, sens, et force des vents, visibilité, température, etc..

Il est toutefois important de noter dès maintenant que sans corrélation notable entre les différentes sources de données, il sera impossible de prédire l'intensité, même relative, des flux migratoires. Dans le cas où la quantité et la qualité des données récoltées ne permettraient pas d'alimenter un modèle d'une telle ampleur (par exemple trop faible couverture géographique, échantillonnage temporellement trop réduit, etc), nous pourrions simplifier le modèle 4D pour obtenir des inférences moins globales comme par exemple modéliser le flux migratoire par unité de temps mais seulement en certains points, ou encore modéliser spatialement les altitudes de vol en fonction de la météo, etc. Dans les sections suivantes, nous présentons : A) les données disponibles pour la modélisation des flux migratoires dans le Golfe du Lion, B) l'approche utilisée pour analyser chaque dimension X, Y, Z et T séparément, avec la contribution des différents jeux de données disponibles, C) la formulation mathématique d'un modèle 4D intégrant les quatre dimensions.

A. Données

Les données concernant les migrateurs terrestres qui utilisent le Golfe du Lion durant les périodes migratoires peuvent être séparées en deux catégories. Nous les présentons ci-dessous avec les dimensions x , y , z , et t qu'elles informent.

- **Les données à l'échelle de l'individu ou de l'espèce**
 - les données GPS qui fournissent des informations sur la position (x , y) et l'altitude (z) d'un individu à chaque instant (t).
 - les données GLS qui informent sur l'altitude (z) d'un individu à chaque instant (t) mais qui sont imprécises géographiquement.
 - les comptages à la côte qui mesurent le nombre d'individus en migration chaque jour (t) mais en une seule localisation (x , y) donc fournissant peu d'information spatiale et sans information sur l'altitude.
- **Les données à l'échelle du groupe d'espèces**

- les données issues des radars à la côte qui fournissent des flux d’oiseaux à chaque altitude (z) en chaque instant (t) mais seulement en deux localisations (x, y) à la fois et donc avec peu d’information spatiale.
- les données issues des campagnes radar en mer qui fournissent des flux d’oiseaux dans tout le Golfe du Lion (x, y, z) mais pendant une fenêtre temporelle restreinte.
- les données issues des radars météo qui fournissent des informations spatiales et temporelles car ils fonctionnent continuellement et ont une grande portée (plusieurs dizaines de km). Les données issues des radars météo pourraient alimenter tous les compartiments spatio-temporels étudiés exceptée la mesure de l’altitude de vol (x, y, t). Nous n’avons pas accès à ces données pour l’instant mais le programme **SEMAFOR** pourrait nous permettre de les obtenir dans les prochaines années.

B. L’analyse par “dimension”

1. La spatialisation des flux : les dimensions X et Y

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 campagnes en mer, dans une moindre mesure les radars à la côte, et dans un cas idéal les données issues des radars météorologiques nous renseignent aussi sur cette variation des flux dans l’espace (Liechti et al., 2019). Cependant, les données GPS/GLS ne peuvent pas informer spatialement sur les flux migratoires à l’échelle du Golfe du Lion car les individus équipés ne sont pas répartis de façon homogène sur l’ensemble de la zone à ce stade de l’étude (ils se trouvent particulièrement en Camargue et autour de Montpellier). Par ailleurs les données GLS ne permettent pas la localisation des individus équipés avec la précision nécessaire pour ce type de question (Phillips et al., 2004). Les données issues des radars à la côte fournissent des informations spatiales mais seulement sur deux localisations géographiques en simultané.

La modélisation spatiale des flux intégrera donc les données issues des campagnes en mer et des radars à la côte. Durant les périodes de campagnes en mer, l’intégration des données issues des lots 4 et 5 permettra de comparer les flux mesurés à la côte avec les flux observés en mer. Cette analyse intégrée permettrait, si elle s’avère efficace, de prédire spatialement l’intensité du flux migratoire en fonction de paramètres météorologiques et géographiques.

On estimera alors $\mu_{x,y}$ l’intensité du flux migratoire en chaque cellule de la grille (Figure 6). Nous prendrons en compte les variations spatiales de la migration par l’inclusion de covariables environnementales abiotiques (e.g. la distance à la côte, profondeur) et météorologiques (e.g. vitesse moyenne du vent) influençant potentiellement la migration. Nous estimerons aussi un paramètre lissant sur les coordonnées géographiques de chaque cellule de la zone d’étude. Par exemple, le flux $\mu_{x,y}$ en x et y pourra être estimé comme :

$$\log(\mu_{x,y}) = \alpha_0 + \alpha_1 \text{distance à la côte}_{x,y} + \alpha_2 \text{vent moyen}_{x,y} + f(x,y)$$

où α_0 , α_1 , et α_2 sont à estimer, tout comme $f(.)$ fonction lissante.

Les flux migratoires mesurés par radar renseigneront les valeurs de $\mu_{x,y,t}$ via un processus de Poisson (voir section sur le modèle 4D) si on considère que les mesures radar sont précises. Toutefois, on pourra prendre en compte la détection imparfaite via un processus d'observation. Par exemple, soit $Rmer_{x,y}$ le nombre d'oiseaux mesurés par le radar en mer en x, y . On estimera :

$$N_{x,y} \sim \text{Poisson}(\mu_{x,y})$$

$$Rmer_{x,y} \sim \text{Binomial}(N_{x,y}, p_{x,y})$$

avec $p_{x,y}$ la probabilité individuelle de détection estimée en fonction de covariables (voir modèles sur les oiseaux marins).

Dans ce cas, l'apport des données télémétriques dans l'estimation des flux migratoire sera en premier lieu qualitatif, pour comparer les trajets observés par GPS aux prédictions de ce modèle intégré.

2. L'altitude de vol : la dimension Z

L'altitude de vol est mesurée par les différentes sources de données disponibles dans Migralion. Les GPS/GLS du lot 3 relèveront les altitudes de vol des oiseaux, tout comme les radars provenant des lots 4 et 5. 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 et en mer. Au contraire des radars ornithologiques, les altitudes ne sont pas estimées par les observateurs lors des comptages côtiers, ni par les radars météorologiques à notre connaissance. Nous avons donc à notre disposition les GPS/GLS et les radars ornithologiques pour mesurer l'altitude de vol.

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 estimer l'altitude de vol moyenne ainsi qu'une variance associée (Ross-Smith et al., 2016), ou on pourra diviser les altitudes de vol potentielles deux classes : une classe d'altitudes de vol "à risque" correspondant à une zone risque de collision, et une classe d'altitudes de vol sans risque.

Pour prendre en compte les variations d'altitude de vol, nous incluerons des covariables environnementales abiotiques (e.g. la distance à la côte, profondeur) et météorologiques (e.g. vitesse du vent au moment de la mesure, pression atmosphérique) 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 et par radar 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.

Dans le cas d'une analyse des flux incluant les autres dimensions X, Y, et T, la formulation de l'altitude de vol devra prendre une autre formulation que nous décrivons dans la section sur le modèle 4D.

3. Phénologie de la migration : la dimension T

Hormis les campagnes en mer qui sont réalisées durant quelques jours par période migratoire, les données issues de la télémétrie et des radars (à la côte et météo) sont collectées durant toute la migration.

L'intensité migratoire varie au cours du temps et ces variations seront prises en compte explicitement par l'inclusion d'une dimension temporelle t dans les modèles que nous réaliserons. Ainsi, le flux μ estimé dépendra de t (voir section suivante), et on calculera μ_t , μ_{zt} , μ_{xyt} , ou μ_{xyzt} en fonction du nombre de dimensions du modèle considéré. Le niveau de définition du pas de temps t , entre t_0 le début de la période migratoire et T la fin de la période, reste à définir en fonction des l'abondance et de la qualité des données disponibles.

C. Formulation d'un modèle 4D et intégration

Lors de l'analyse des flux migratoires, les variations spatiales (x , et y), altitudinales (z), et temporelles (t) sont à prendre en compte. Comme expliqué dans les sections précédentes, chaque source de variations est informée par un ensemble de données différentes et nécessite une formulation mathématique particulière pour permettre son analyse. Dans l'idéal, il serait possible d'intégrer dans un même modèle les quatre dimensions x, y, z, t dont la modélisation séparée a été décrite précédemment. Nous expliquons sa formulation dans la section suivante puis détaillons les potentielles limites.

1. Processus écologique

En intégrant les quatre dimensions détaillées ci-dessus, nous tentons de modéliser le flux migratoire $\mu_{x,y,z,t}$ décrivant la densité d'utilisation de chaque unité d'espace, d'altitude, et de temps par les oiseaux migrateurs.

Pour prendre en compte les variations de μ dans les différentes dimensions, nous incluons des covariables environnementales et météorologiques. La phénologie de la migration est prise en compte via la dimension temporelle t . Nous pourrions considérer des relations non-linéaires entre les différentes variables et μ en utilisant des modèles additifs généralisés (GAM). Les GAM peuvent aussi être utiles pour prendre en compte l'influence de l'auto-corrélation spatiale en incluant un paramètre non linéaire sur les coordonnées géographiques x et y . Ainsi, μ est calculé pour chaque x, y, z, t via une régression log-linéaire des covariables :

$$\log(\mu_{x,y,z,t}) = \gamma_0 + \gamma_1 \text{environnement}_{x,y} + \gamma_2 \text{météo}_{x,y,t} + f(x, y)$$

Dans cette équation, γ_0 , γ_1 , et γ_2 sont les coefficients à estimer et $f(\cdot)$ est une fonction lissante des coordonnées géographiques.

2. Processus d'observation

a) Données radar et comptages

On divisera chaque jeu de données selon les résolutions retenues pour chacune des quatre dimensions. Par exemple pour le radar en mer, on obtiendra $R_{mer,x,y,z,t}$ le nombre d'oiseaux compté en x, y, z, t .

Les flux migratoires mesurés seront liés à μ par un tirage dans une distribution de Poisson. Par exemple pour le radar en mer :

$$R_{mer,x,y,z,t} \sim \text{Poisson}(\mu_{x,y,z,t})$$

Les données de comptages visuels pourront être modélisées de la même manière en regroupant les espèces. La détection imparfaite pourra être prise en compte via un tirage Binomial si l'analyse des données radar révèle des imprécisions de mesure (voir section sur la spatialisation des flux).

b) Données télémétriques

Les données individuelles de télémétrie sont de nature fondamentalement différente de celles des flux μ estimés dans le modèle en quatre dimensions. La grandeur μ correspond à une densité d'oiseaux par unité d'espace et de temps. Ainsi, μ n'est pas compatible avec des trajets GPS ou GLS qui représentent seulement la trajectoire d'un individu. Pour rendre compatibles les données GPS/GLS avec μ , nous développerons des modèles spécifiques pour les données GPS/GLS que nous intégrerons ensuite dans le modèle 4D. Il faudra aussi filtrer les données issues des GPS/GLS pour garder uniquement les déplacements des oiseaux pendant la période de migration.

Bien que les données GPS/GLS du lot 3 concernent un nombre réduit d'individus, elles relèvent les altitudes de vol des oiseaux avec une grande profondeur temporelle vu que les individus équipés sont suivis pendant tout leur trajet migratoire. Ainsi, les données GPS/GLS permettent d'étudier l'influence des conditions météorologiques, géographiques et environnementales sur les altitudes de vol, comme détaillé dans la section précédente.

Dans le modèle 4D, nous utiliserons les altitudes de vol mesurées par les GPS/GLS et regroupées par groupe d'espèces. Nous essaierons de modéliser les altitudes de vol en fonction des variables météorologiques et environnementales pour obtenir la **probabilité individuelle d'altitude de vol**, i.e. la probabilité qu'un oiseau vole à une altitude donnée sachant les conditions environnementales et météorologiques. Ensuite, nous injecterons la prédiction sur l'altitude de vol dans le modèle 4D pour compléter les informations sur la répartition verticale des flux (i.e. le long de la dimension z).

Formulation du modèle d'altitude

Comme détaillé précédemment, l'altitude de vol peut être modélisée de façon continue en assignant une altitude moyenne et une dispersion autour de cette moyenne (Ross-Smith et al., 2016). Cependant, cette formulation est technique et gourmande en temps de calcul. Pour faciliter l'écriture et l'estimation des paramètres, les données d'altitude z collectées par GPS/GLS seront divisées en Z classes d'altitudes dans notre modèle en quatre dimensions (e.g. $z_1 = [0; 100m]$, $z_2 = [100; 200m]$, jusqu'à l'altitude maximale mesurée). Si une analyse séparée de l'altitude de vol est menée, il sera possible de garder la formulation continue comme expliqué dans la section dédiée. Vu le contexte éolien de Migralion, nous pourrions considérer deux classes d'altitudes : une classe d'altitude "à risque" correspondant à la potentielle zone de collision, et une classe d'altitude sans risque. Soit la classe d'altitude mesurée z qui sera modélisée par un tirage dans une distribution multinomiale de dimension Z et de paramètre π :

$$z \sim \text{Multinomial}(1, \pi)$$

où π correspond au vecteur standardisé des probabilités de voler dans chaque classe d'altitude.

$$\pi = \left(\frac{p_1}{\sum_{z=1}^Z p_z}, \dots, \frac{p_Z}{\sum_{z=1}^Z p_z} \right)$$

Chaque, p_z est modélisé via une régression linéaire généralisée avec un lien de type logit.

$$\text{logit}(p_z) = \beta_0 + \beta_1 \text{meteo}_{x,y,t} + \beta_2 \text{environnement}_{x,y}$$

où β_0 , β_1 et β_2 sont les paramètres à estimer. Avec ce modèle, on estime l'effet de la météo et de l'environnement sur la probabilité de voler dans une certaine classe d'altitude via les données télémétriques.

c) La vraisemblance jointe sur l'altitude de vol

Avec les données télémétriques on estime "quelle est la probabilité qu'un oiseau vole dans chaque classe d'altitude ?", tandis que les données issues des radars nous informent sur la répartition du flux selon les classes d'altitudes.

Pour intégrer ces deux métriques, la solution est de considérer que la répartition dans chaque classe d'altitude du flux mesuré par radar suit la même loi Multinomiale que la probabilité de voler dans chaque classe d'altitude pour les altitudes relevées par GPS/GLS. Ainsi, on retrouve π_z , le vecteur des probabilités de voler dans chaque classe d'altitude, probabilités qui seront alors estimées conjointement par les deux sources de données.

Formulation

Soit $Rmer_{x,y,z,t}$ les données observées par le radar en mer aux coordonnées x, y , à l'altitude z et au temps t et $N_{x,y,t}$ la somme des flux μ en x, y , et t .

$$N_{x,y,t} = \sum_z^Z \mu_{x,y,z,t}$$

En gardant la même division en Z classes d'altitude que pour le modèle ajusté sur les données télémétriques, on pourra estimer la dimension z de μ via un tirage dans une distribution Multinomiale de taille $N_{x,y,t}$ et de paramètre π .

$$Rmer_{x,y,z,t} \sim \text{Multinomial}(N_{x,y,t}, \pi)$$

$$\text{où } \pi = \left(\frac{p_1}{\sum_{z=1}^Z p_z}, \dots, \frac{p_Z}{\sum_{z=1}^Z p_z} \right)$$

$$\text{et } \text{logit}(p_z) = \beta_0 + \beta_1 \text{meteo}_{x,y,t} + \beta_2 \text{environnement}_{x,y}$$

Ainsi, π et p_z correspondent aux mêmes probabilités individuelles d'altitude de vol que celles estimées par les données de télémétrie.

Pour une visualisation graphique et simplifiée du modèle, voir Figure 9.

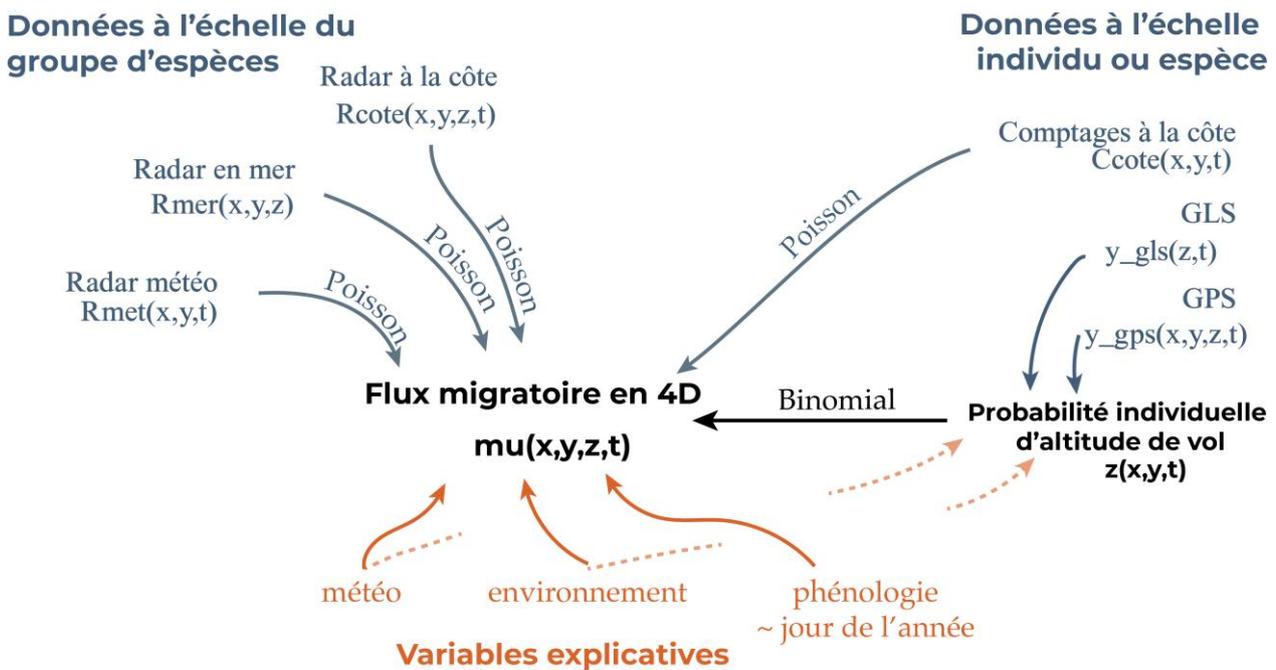


Figure 8 : Présentation schématique des données, covariables, et processus écologiques du modèle estimant les flux migratoires en 4 dimensions. Les dimensions x et y correspondent aux coordonnées géographiques, z à l'altitude de vol, et t au temps.

3. Anticipation des difficultés

Ce modèle 4D intégrant toutes les sources de données disponibles est réalisable en théorie (comme le montre sa formulation mathématique ci-dessus). Cependant, le modèle comprend un grand nombre de paramètres à estimer conjointement et plusieurs difficultés sont à prévoir.

- La quantité de données récoltées pourrait être trop faible pour une caractérisation écologique de cette ampleur avec plusieurs milliers d'unités d'espace pendant plusieurs centaines d'unités temporelles. Une trop forte résolution va augmenter le nombre d'unité spatio-temporelles sans information et risque de compliquer l'estimation des paramètres liant μ aux covariables. Il pourra être envisager de diminuer la résolution spatiale ou temporelle du modèle mais il ne faudra pas dépasser une résolution minimale au-delà de laquelle les estimations seraient trop grossière pour être utile dans le contexte de Migration.
- Le temps de calcul pourrait rendre irréalisable l'estimation des paramètres d'un tel modèle. Il est possible qu'un modèle de cette ampleur prenne plusieurs semaines à converger. Ce temps de calcul dépendra des résolutions spatiales et temporelles retenues, du nombre de variables écologiques et météorologiques incluses, etc. Un temps de calcul de plusieurs semaines serait rédhibitoire pour une exploration écologique fine qui nécessite de lancer les modèles plusieurs dizaines de fois.
- Les données récoltées pourraient ne pas permettre de corréler les flux à des covariables spatiales ou temporelles. Par exemple, les flux de migration pourraient ne pas être dépendant des variables météorologiques et environnementales retenues, ou les données récoltées caractériseraient des flux qui ne pourraient pas être reliés. Dans un tel cas où le signal écologique serait faible, le modèle ne pourrait pas produire d'estimation statistiquement fiable.

Si ces difficultés émergent, nous envisagerons de diviser les analyses en plusieurs études par dimensions comme présenté dans les sections précédentes.

VI. Perspectives

Les pistes de modélisation décrites dans les sections précédentes constituent une base méthodologique prometteuse mais dont l'application reste incertaine. En effet, à la date d'écriture de ce rapport, nous n'avons pas les données pour développer les outils de modélisation les précédemment décrits ni pour en explorer ou en tester la faisabilité. Les interactions avec les porteurs des lots 3 à 5 nous renseignent cependant sur la quantité et la qualité des données prochainement disponibles. Nous rédigeons dans cette section, les perspectives pressenties dès lors que les premières données seront disponibles.

A. Dans l'attente des analyses écologiques par lot

Les données radar issues des lots 4 et 5 permettent d'obtenir des flux d'oiseaux à la côte et en mer. Au jour de l'écriture de ce livrable, nous n'avons pas encore pu consulter les données des campagnes de la période pré-nuptiale 2022.

L'analyse par Biotope des données de radar à la côte permettra de comparer les passages migratoires mesurés par les deux radars côtiers. Mettre en évidence une corrélation des flux entre le radar mobile qui parcourt le littoral et le radar fixe de La Palissade (Camargue) serait encourageant car un tel lien permettrait de prédire l'intensité de la migration le long de la côte à partir des données de migration relevées au niveau du radar fixe. Cependant, d'après les premiers retours informels de l'équipe de Biotope, les flux migratoires seraient peu corrélés entre la Camargue et le reste du littoral.

L'analyse par chaque responsable de lot des données récoltées lors de la période pré-nuptiale 2022 permettra de discriminer plus précisément ce qui est atteignable en termes de modélisation intégrée de ce qui ne l'est pas. Plus largement, les résultats descriptifs obtenus lors de cette première campagne d'acquisition permettront de produire des recommandations pour les années 2023 et 2024 sur les lots 4 et 5 dans la mesure de ce qui est modifiable (e.g. localisation du radar fixe, dates des passages des campagnes en mer, etc).

B. Incertitudes autour de la contribution des GLS

La quantité et la qualité de l'information qui sera extraite des oiseaux équipés de GLS ne pourra pas être évaluée car les oiseaux ne sont pas encore recapturés. La quantité d'information dépendra de la capacité à recapturer les individus équipés. La qualité des données de GLS permettra de fournir des informations sur l'altitude de vol et la phénologie des espèces mais la précision des données GLS sera probablement trop limitée pour décrire les déplacements à l'échelle du Golfe du Lion, en particulier lors des périodes de migration à proximité des équinoxes lorsque de les GLS ne permettent plus la localisation des individus (Phillips et al., 2004).

C. Phénologie : étude spécifique ou déclinaison des autres modèles

La temporalité des flux migratoires sera explorée explicitement dans les modèles d'utilisation de l'espace et de modélisation du flux migratoires via l'inclusion d'une dimension temporelle t .

Cependant, en fonction de l'analyse de la phénologie par le lot 3 à partir des données GPS/GLS, il sera éventuellement possible de développer une analyse intégrée sur la phénologie avec les autres sources de données (e.g. notamment les comptages visuels côtiers).

D. Quelles adaptations ?

Enfin, la période d'analyse qui se dessine pour l'automne et l'hiver de 2022- 2023 s'avère cruciale pour obtenir un recul critique sur la première année de collecte. L'analyse détaillée des différentes données issues des radars côtiers permettront de mieux cerner les analyses que nous pourrons espérer développer dans le lot 6. Sans pouvoir mener d'analyse complète, nous devons collectivement évaluer au moins qualitativement les données récupérées, identifier les difficultés rencontrées pour répondre aux questions écologiques de Migralion et se lancer dans une réflexion concernant l'éventuelle modification des protocoles de suivis pour l'année 2023.

VII. Références

- Aarts, G., MacKenzie, M., McConnell, B., Fedak, M., & Matthiopoulos, J. (2008). Estimating space-use and habitat preference from wildlife telemetry data. *Ecography*, 31(1), 140–160. <https://doi.org/10.1111/j.2007.0906-7590.05236.x>
- Amundson, C. L., Royle, J. A., & Handel, C. M. (2014). A hierarchical model combining distance sampling and time removal to estimate detection probability during avian point counts. *The Auk*, 131(4), 476–494. <https://doi.org/10.1642/AUK-14-11.1>
- Avgar, T., Potts, J. R., Lewis, M. A., & Boyce, M. S. (2015). Integrated step selection analysis: Bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, 12.
- Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3(2), 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Besbeas, P., Freeman, S. N., Morgan, B. J. T., & Catchpole, E. A. (2002). Integrating Mark-Recapture-Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters. *Biometrics*, 58(3), 540–547. <https://doi.org/10.1111/j.0006-341X.2002.00540.x>
- Bland, L. M., Regan, T. J., Dinh, M. N., Ferrari, R., Keith, D. A., Lester, R., Mouillot, D., Murray, N. J., Nguyen, H. A., & Nicholson, E. (2017). Using multiple lines of evidence to assess the risk of ecosystem collapse. *Proceedings of the Royal Society B: Biological Sciences*, 284(1863), 20170660. <https://doi.org/10.1098/rspb.2017.0660>
- Buckland, S. T., Anderson, D. R., Burnham, K. P., & Laake, J. L. (2005). Distance sampling. *Encyclopedia of Biostatistics*, 2.
- CEREMA. (2021). *Étude avifaune en Méditerranée – Valorisation des données télémétriques*. 79. https://geolittoral.din.developpement-durable.gouv.fr/telechargement/AO6_Avifaune/Rapport_%C3%89tude%20avifaune%20Med_t%C3%A9l%C3%A9m%C3%A9trique.V5_10062021.pdf
- Converse, S. J., McClintock, B. T., & Conn, P. B. (2022). Special Feature: Linking capture–recapture and movement. *Ecology*, n/a(n/a), e3770. <https://doi.org/10.1002/ecy.3770>
- Cook, C. N., Carter, R. W. (Bill)., Fuller, R. A., & Hockings, M. (2012). Managers consider multiple lines of evidence important for biodiversity management decisions. *Journal of Environmental Management*, 113, 341–346. <https://doi.org/10.1016/j.jenvman.2012.09.002>
- Courbin, N., Besnard, A., Péron, C., Saraux, C., Fort, J., Perret, S., Tornos, J., & Grémillet, D. (2018). Short-term prey field lability constrains individual specialisation in resource selection and foraging site fidelity in a marine predator. *Ecology Letters*, 21(7), 1043–1054. <https://doi.org/10.1111/ele.12970>
- Cushman, S. A. (2010). Animal Movement Data: GPS Telemetry, Autocorrelation and the Need for Path-Level Analysis. In S. A. Cushman & F. Huettmann (Eds.), *Spatial Complexity, Informatics, and Wildlife Conservation* (pp. 131–149). Springer Japan. https://doi.org/10.1007/978-4-431-87771-4_7
- Dobson, F. S., & Jouventin, P. (2010). The trade-off of reproduction and survival in slow-breeding seabirds. *Canadian Journal of Zoology*, 88(9), 889–899. <https://doi.org/10.1139/Z10-054>

- Dorazio, R. M. (2014). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data: Imperfect detection and survey bias in presence-only data. *Global Ecology and Biogeography*, 23(12), 1472–1484. <https://doi.org/10.1111/geb.12216>
- Farr, M. T., Green, D. S., Holekamp, K. E., & Zipkin, E. F. (2020). Integrating distance sampling and presence-only data to estimate species abundance. *Ecology*. <https://doi.org/10.1002/ecy.3204>
- Fieberg, J., Signer, J., Smith, B., & Avgar, T. (2021). A “How to” guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, 90(5), 1027–1043. <https://doi.org/10.1111/1365-2656.13441>
- Fletcher, R. J., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology*, e02710. <https://doi.org/10.1002/ecy.2710>
- Isaac, N. J. B., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., Henrys, P. A., Jarvis, S., Lahoz-Monfort, J., Pagel, J., Pescott, O. L., Schmucki, R., Simmonds, E. G., & O’Hara, R. B. (2019). Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology & Evolution*. <https://doi.org/10.1016/j.tree.2019.08.006>
- Jean-Hervé, B. (1985). *PELMED - PELAGIQUES MEDITERRANÉE*. <https://doi.org/10.18142/19>
- Jiménez, J., Díaz-Ruiz, F., Monterroso, P., Tobajas, J., & Ferreras, P. (2022). Occupancy data improves parameter precision in spatial capture–recapture models. *Ecology and Evolution*, 12(8), e9250. <https://doi.org/10.1002/ece3.9250>
- Johnson, D. S., Thomas, D. L., Ver Hoef, J. M., & Christ, A. (2008). A General Framework for the Analysis of Animal Resource Selection from Telemetry Data. *Biometrics*, 64(3), 968–976. <https://doi.org/10.1111/j.1541-0420.2007.00943.x>
- Kéry, M., & Royle, J. (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models: Vols. Ch 10* (1st ed.). Elsevier. <https://doi.org/10.1016/C2015-0-04070-9>
- Kéry, M., & Royle, J. A. (2016). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS*. Elsevier/AP, Academic Press is an imprint of Elsevier.
- Kéry, M., & Schaub, M. (2012). *Bayesian Population Analysis using WinBUGS*. 443.
- Labach, H. (2021). *Stratégie de conservation du Grand Dauphin (Tursiops truncatus) en Méditerranée française* [PhD thesis].
- Labach, H., Azzinari, C., Barbier, M., Cesarini, C., Daniel, B., David, L., Dhermain, F., Di-Méglio, N., Guichard, B., Jourdan, J., Lauret, V., Robert, N., Roul, M., Tomasi, N., & Gimenez, O. (2021). Distribution and abundance of common bottlenose dolphin (*Tursiops truncatus*) over the French Mediterranean continental shelf. *Marine Mammal Science*, n/a(n/a). <https://doi.org/10.1111/mms.12874>
- Laran, S., Pettex, E., Authier, M., Blanck, A., David, L., Dorémus, G., Falchetto, H., Monestiez, P., Van Canneyt, O., & Ridoux, V. (2017). Seasonal distribution and abundance of cetaceans within French waters- Part I: The North-Western Mediterranean, including the Pelagos sanctuary. *Deep Sea Research Part II: Topical Studies in Oceanography*, 141, 20–30. <https://doi.org/10.1016/j.dsr2.2016.12.011>

- Lauret, V., Labach, H., Authier, M., & Gimenez, O. (2021). Using single visits into integrated occupancy models to make the most of existing monitoring programs. *Ecology*, 848663. <https://doi.org/10.1101/848663>
- Lauret, V., Labach, H., Turek, D., Laran, S., & Gimenez, O. (2021). Spatial integrated models foster complementarity between monitoring programs in producing large-scale ecological indicators. *bioRxiv*.
- Lauret, V., Labach, H., Turek, D., Laran, S., & Gimenez, O. (2022). Integrated spatial models foster complementarity between monitoring programmes in producing large scale bottlenose dolphin indicators. *Animal Conservation*, 11. <https://doi.org/10.1111/acv.12815>
- Liechti, F., Aschwanden, J., Blew, J., Boos, M., Brabant, R., Dokter, A. M., Kosarev, V., Lukach, M., Maruri, M., Reyniers, M., Schekler, I., Schmaljohann, H., Schmid, B., Weisshaupt, N., & Sapir, N. (2019). Cross-calibration of different radar systems for monitoring nocturnal bird migration across Europe and the Near East. *Ecography*, 42(5), 887–898. <https://doi.org/10.1111/ecog.04041>
- MacKenzie, D. I. (Ed.). (2006). *Occupancy estimation and modeling: Inferring patterns and dynamics of species*. Elsevier.
- Mackenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). *Estimating site occupancy rates when detection probabilities are less than one*. 83(8), 8.
- Markones, N., Dierschke, V., & Garthe, S. (2010). Seasonal differences in at-sea activity of seabirds underline high energetic demands during the breeding period. *Journal of Ornithology*, 151(2), 329–336. <https://doi.org/10.1007/s10336-009-0459-2>
- Martino, S., Pace, D. S., Moro, S., Casoli, E., Ventura, D., Frachea, A., Silvestri, M., Arcangeli, A., Giacomini, G., Ardizzone, G., & Jona Lasinio, G. (2021). Integration of presence-only data from several sources: A case study on dolphins' spatial distribution. *Ecography*, 44(10), 1533–1543. <https://doi.org/10.1111/ecog.05843>
- Matthiopoulos, J., Wakefield, E., Jeglinski, J. W. E., Furness, R. W., Trinder, M., Tyler, G., Mccluskie, A., Allen, S., Braithwaite, J., & Evans, T. (2022). Integrated modelling of seabird-habitat associations from multi-platform data: A review. *Journal of Applied Ecology*, 59(4), 909–920. <https://doi.org/10.1111/1365-2664.14114>
- McLoughlin, P. D., Morris, D. W., Fortin, D., Vander Wal, E., & Contasti, A. L. (2010). Considering ecological dynamics in resource selection functions. *Journal of Animal Ecology*, 79(1), 4–12. <https://doi.org/10.1111/j.1365-2656.2009.01613.x>
- Michelot, T., Blackwell, P. G., Chamaillé-Jammes, S., & Matthiopoulos, J. (2020). Inference in MCMC step selection models. *Biometrics*, 76(2), 438–447. <https://doi.org/10.1111/biom.13170>
- Michelot, T., Blackwell, P. G., & Matthiopoulos, J. (2019). Linking resource selection and step selection models for habitat preferences in animals. *Ecology*, 100(1), e02452. <https://doi.org/10.1002/ecy.2452>
- Miller, D. A. W., Pacifici, K., Sanderlin, J. S., & Reich, B. J. (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution*, 10(1), 22–37. <https://doi.org/10.1111/2041-210X.13110>
- Navarro, J., Louzao, M., Igual, J. M., Oro, D., Delgado, A., Arcos, J. M., Genovart, M., Hobson, K. A., & Forero, M. G. (2009). Seasonal changes in the diet of a critically endangered seabird and the importance of trawling discards. *Marine Biology*, 156(12), 2571–2578. <https://doi.org/10.1007/s00227-009-1281-3>

- Nichols, & Williams. (2006). Monitoring for conservation. *Trends in Ecology & Evolution*, 21(12), 668–673. <https://doi.org/10.1016/j.tree.2006.08.007>
- Nilsson, C., Dokter, A. M., Verlinden, L., Shamoun-Baranes, J., Schmid, B., Desmet, P., Bauer, S., Chapman, J., Alves, J. A., Stepanian, P. M., Sapir, N., Wainwright, C., Boos, M., Górska, A., Menz, M. H. M., Rodrigues, P., Leijnse, H., Zehntindjiev, P., Brabant, R., ... Liechti, F. (2019). Revealing patterns of nocturnal migration using the European weather radar network. *Ecography*, 42(5), 876–886. <https://doi.org/10.1111/ecog.04003>
- Nussbaumer, R., Bauer, S., Benoit, L., Mariethoz, G., Liechti, F., & Schmid, B. (2021). Quantifying year-round nocturnal bird migration with a fluid dynamics model. *Interface*, 8.
- Pacifici, K., Reich, B. J., Miller, D. A. W., & Pease, B. S. (2019). Resolving misaligned spatial data with integrated species distribution models. *Ecology*, 100(6). <https://doi.org/10.1002/ecy.2709>
- Phillips, R., Silk, J., Croxall, J., Afanasyev, V., & Briggs, D. (2004). Accuracy of geolocation estimates for flying seabirds. *Marine Ecology Progress Series*, 266, 265–272. <https://doi.org/10.3354/meps266265>
- Pirotta, E., Thompson, P. M., Cheney, B., Donovan, C. R., & Lusseau, D. (2015). Estimating spatial, temporal and individual variability in dolphin cumulative exposure to boat traffic using spatially explicit capture-recapture methods: Variability in wildlife exposure to disturbance. *Animal Conservation*, 18(1), 20–31. <https://doi.org/10.1111/acv.12132>
- Quéroué, M., Barbraud, C., Barraquand, F., Turek, D., Delord, K., Pacoureaux, N., & Gimenez, O. (2021). Multispecies integrated population model reveals bottom-up dynamics in a seabird predator–prey system. *Ecological Monographs*, n/a(n/a). <https://doi.org/10.1002/ecm.1459>
- Raymond, B., Lea, M.-A., Patterson, T., Andrews-Goff, V., Sharples, R., Charrassin, J.-B., Cottin, M., Emmerson, L., Gales, N., Gales, R., Goldsworthy, S. D., Harcourt, R., Kato, A., Kirkwood, R., Lawton, K., Ropert-Coudert, Y., Southwell, C., Hoff, J. van den, Wienecke, B., ... Hindell, M. A. (2015). Important marine habitat off east Antarctica revealed by two decades of multi-species predator tracking. *Ecography*, 38(2), 121–129. <https://doi.org/10.1111/ecog.01021>
- Ross-Smith, V. H., Thaxter, C. B., Masden, E. A., Shamoun-Baranes, J., Burton, N. H. K., Wright, L. J., Rehfish, M. M., & Johnston, A. (2016). Modelling flight heights of lesser black-backed gulls and great skuas from GPS: A Bayesian approach. *Journal of Applied Ecology*, 53(6), 1676–1685. <https://doi.org/10.1111/1365-2664.12760>
- Royle, J. A., Dawson, D. K., & Bates, S. (2004). Modelin abundance effects in distance sampling. *Ecology*, 85(6), 1591–1597. <https://doi.org/10.1890/03-3127>
- Schaub, M., & Abadi, F. (2011). Integrated population models: A novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology*, 152(S1), 227–237. <https://doi.org/10.1007/s10336-010-0632-7>
- Sillett, T. S., Chandler, R. B., Royle, J. A., Kéry, M., & Morrison, S. A. (2012). Hierarchical distance-sampling models to estimate population size and habitat-specific abundance of an island endemic. *Ecological Applications*, 22(7), 1997–2006. <https://doi.org/10.1890/11-1400.1>
- Sun, C. C., Royle, J. A., & Fuller, A. K. (2019). Incorporating citizen science data in spatially explicit integrated population models. *Ecology*, 100(9). <https://doi.org/10.1002/ecy.2777>

- Trevail, A. M., Green, J. A., Bolton, M., Daunt, F., Harris, S. M., Miller, P. I., Newton, S., Owen, E., Polton, J. A., Robertson, G., Sharples, J., & Patrick, S. C. (2021). Environmental heterogeneity promotes individual specialisation in habitat selection in a widely distributed seabird. *Journal of Animal Ecology*, *90*(12), 2875–2887. <https://doi.org/10.1111/1365-2656.13588>
- Wooller, R. D., Bradley, J. S., & Croxall, J. P. (1992). Long-term population studies of seabirds. *Trends in Ecology & Evolution*, *7*(4), 111–114. [https://doi.org/10.1016/0169-5347\(92\)90143-Y](https://doi.org/10.1016/0169-5347(92)90143-Y)
- Zipkin, E. F., Inouye, B. D., & Beissinger, S. R. (2019). Innovations in data integration for modeling populations. *Ecology*, e02713. <https://doi.org/10.1002/ecy.2713>
- Zipkin, E. F., & Saunders, S. P. (2018). Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation*, *217*, 240–250. <https://doi.org/10.1016/j.biocon.2017.10.017>