


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 décrivant la méthode de suivi et de modélisation de l'espace par
l'avifaune en mer sur une large zone**

Livable L34

Mars 2024

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	Coline Canonne , Valentin Lauret, Louis Schroll, Sébastien Roques, Aurélien Besnard	France Énergies Marines (FEM) Centre d'Ecologie Fonctionnelle et Evolutive (CEFE) – EPHE
Coordination scientifique et technique du projet	Aurélien Besnard Yann Planque	CEFE – EPHE France Énergies Marines (FEM)
Coordination du projet	Claire Hébert	Office Français de la Biodiversité (OFB)
Titre du document	Migralion - Premier rapport décrivant la méthode de suivi et de modélisation de l'espace par l'avifaune en mer sur une large zone – livrable L34 - 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	CANONNE, C., LAURET, V., SCHROLL, L., ROQUES, S., BESNARD, A. 2024. 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 décrivant la méthode de suivi et de modélisation de l'espace par l'avifaune en mer sur une large zone. 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	16/05/2024	
Date de diffusion	Mai 2024	
Version du document	3	
Nom du fichier	MIGRALION_Lot6_L34_Rapportméthode1_2024-03.pdf	
Nombre de pages	40	
Niveau de diffusion	Consortium / Gouvernance/ Public	

HISTORIQUE DES CHANGEMENTS

Version	Date	Modifié par	Modifications / Sections
1	19/03/2024	Coline Canonne, Valentin Lauret, Louis Schroll, Sébastien Roques, Aurélien Besnard	Document initial
2	29/04/2024	Jehanne Rivet (FEM), Claire Hébert (OFB)	Relecture et corrections mineures dans l'ensemble du document
3	02/05/2024	Coline Canonne	Corrections mineures dans l'ensemble du document

APPROBATION

Version	Date	Approuvé par	Titre
Finale	16/05/2024	Yann Planque, FEM	
Signature : 			
Finale	16/05/2024	Claire Hébert, OFB	
Signature : 			

Table des matières

Objectifs du programme	5
Objectifs de ce rapport	5
Lot 6 : Développement de méthodes permettant l'analyse des différentes données produites dans le cadre du programme et issues d'autres projets	6
1 Rappels sur les objectifs du lot	6
1.1 Consortium du lot 6.....	6
1.2 Objectifs généraux du lot 6	6
2 Oiseaux marins : intégration des différentes campagnes d'acquisition de données	7
2.1 Développement d'un modèle intégrant les données de comptages/d'observation en mer	7
2.2 Développement d'un modèle intégrant télémétrie et données de comptages/d'observation en mer..	10
2.2.1 Contexte, enjeux	10
2.2.2 Présentation de la méthodologie intégrée	11
2.3 Oiseaux marins et altitude de vol.....	12
3 Spatialisation des flux de migrants	13
3.1 Intégration des données de radar météo et de radar à la côte via un IPP	13
3.2 Exemple d'application du modèle précédent sur des données simulées.....	15
3.3 Autres méthodes disponibles dans la littérature.....	16
4 Phénologie, altitudes et directions de vol de la migration	17
4.1 Migrants terrestres et phénologie de la migration	17
4.2 Migrants terrestres et altitude de vol.....	18
4.3 Migrants terrestres et directions de vol	19
Références bibliographiques.....	20
Annexes.....	24

Table des figures

Figure 1 : SIMULATION D'UNE INTENSITE MIGRATOIRE SUR LE GOLFE DU LION ET DE LA POSITION DISPOSITIFS DE RECOLTE DE DONNEES RADARS.	15
Figure 2 : PREDICTION DE L'INTENSITE MIGRATOIRE SUR TOUTE LA ZONE D'ETUDE UNIQUEMENT A PARTIR DES DONNEES ISSUES DES RADARS BIRDSCAN A LA COTE.	16
Figure 3 : PREDICTION DE L'INTENSITE MIGRATOIRE SUR TOUTE LA ZONE D'ETUDE A PARTIR DES DONNEES ISSUES DES RADARS METEOROLOGIQUES AINSI QUE DES RADARS BIRDSCAN A LA COTE.	16
Figure 4 : SCHEMA REPRESENTANT LE CADRE DE L'ANALYSE INTEGREE PRESENTEE DANS MEEHAN ET AL. 2022.	17

Table des tableaux

Tableau 1 : ILLUSTRATION DU PROCESSUS DE SELECTION DES COVARIABLES AVEC LES DONNEES SUR LA STERNE CAUGEK EN DEHORS DE SA PERIODE DE REPRODUCTION.	9
--	---

Objectifs du programme

Le constat de déficit de connaissances sur les migrateurs terrestres, les zones fonctionnelles des oiseaux marins et des chiroptères a été mis en évidence en Méditerranée française par la communauté scientifique, les associations environnementales et les gestionnaires d'espaces naturels, notamment en lien avec le développement de nouvelles activités anthropiques en mer.

Or, dans un contexte de développement de projets en mer Méditerranée, notamment de centrales éoliennes offshore dans le golfe du Lion, il est apparu essentiel d'acquérir des connaissances permettant de caractériser les flux migratoires et les fonctionnalités des zones en mer pour l'avifaune et les chiroptères, à l'échelle du golfe du Lion, que ce soit pour des espèces résidentes ou ponctuellement présentes.

Pour répondre à ces besoins, l'Office français de la biodiversité a lancé le programme MIGRALION qui a pour objectif d'étudier les thématiques suivantes : passages migratoires en mer de l'avifaune migratrice et des chiroptères (routes préférentielles, flux et altitudes) ainsi que l'utilisation en trois dimensions de l'espace marin par les oiseaux marins. Une vaste diversité de moyens humains et technologiques sont déployés de manière complémentaire : observations visuelles à terre et en mer, baguage, télémétrie, radars ornithologiques et de navigation, récepteurs/enregistreurs acoustiques et à ultrasons. L'ensemble des données collectées ainsi que les données historiques accessibles seront analysés grâce à une méthode d'analyse combinée de données multi-sources développée spécialement dans le cadre du programme. Les différents modules (lots) de l'étude, débutée le 31 mars 2021, sont portés par une douzaine de structures expertes (associatives, académiques, bureau d'étude, etc.) ayant collaboré à l'élaboration de propositions complémentaires pour les 6 lots.

Ainsi ce vaste programme a pour ambition d'apporter des éléments de connaissance sur la faune volante terrestre et marine avec des données inédites sur la spatio-temporalité des déplacements locaux et les migrations des espèces à travers le golfe du Lion. Ces informations aideront à caractériser les menaces qui pèsent sur les populations, dont plusieurs sont en déclin. Il sera alors possible de déterminer comment optimiser le réseau d'espaces naturels protégés, en mer et à terre, d'améliorer les pratiques de gestion de ces espaces, d'ajuster si besoin les programmes de développement d'infrastructures, et de mieux réguler les activités anthropiques responsables de la dégradation de l'état des populations.

Objectifs de ce rapport

Ce rapport présente les méthodologies envisagées dans le cadre des analyses intégrées des données récoltées dans le programme MIGRALION. Les deux premiers volets concernent l'utilisation de l'habitat du golfe du Lion par les oiseaux marins, au travers de comptages et/ou de suivi GPS. Le troisième volet présente une étude par simulation d'analyse des flux migratoires se basant sur les données radars. Enfin, nous présentons les approches envisagées pour l'analyse intégrée des hauteurs et directions de vol, pour les oiseaux marins et les migrateurs terrestres.

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

1 Rappels sur les objectifs du lot

1.1 Consortium du lot 6

Ce document s'inscrit dans le cadre du lot n°6 du programme MIGRALION dont France Energies Marines (FEM), le Centre d'Ecologie Fonctionnelle et Evolutive de Montpellier (CEFE) et la Tour du Valat (TDV) sont titulaires.

L'institut France Energies Marines (FEM) est l'Institut pour la Transition Énergétique dédié aux énergies marines renouvelables (EMR). Sa mission : fournir, valoriser et alimenter l'environnement scientifique et technique nécessaire au développement de cette filière en pleine expansion. Fort d'une équipe pluridisciplinaire de plus de 50 collaborateurs et d'un modèle de collaboration public-privé, France Energies Marines s'appuie sur quatre programmes de R&D transversaux et complémentaires : caractérisation de sites, dimensionnement et suivi des systèmes, intégration environnementale et optimisation des parcs.

La Tour du Valat est un institut de recherche pour la conservation des zones humides méditerranéennes, basée en Camargue sous le statut d'une fondation privée reconnue d'utilité publique. L'objectif principal de la Tour du Valat est de changer le comportement et les décisions des gouvernements et de la société en général dans le bassin méditerranéen afin que les zones humides soient conservées et gérées de façon durable.

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

1.2 Objectifs généraux du lot 6

L'objectif de ce lot est de développer des méthodologies de traitement et d'analyses statistiques des données acquises par les moyens déployés dans les lots 3, 4 et 5 de manière à caractériser de manière robuste la migration de l'avifaune terrestre et l'utilisation de l'espace par l'avifaune marine au-dessus du golfe du Lion. Il s'agit en particulier de :

- Veiller à la cohérence des différents plans d'échantillonnage spatio-temporels prévus dans les lots 3, 4 et 5 pour s'assurer que les données obtenues pourront être comparées et agrégées dans des analyses combinées
- Compléter la synthèse bibliographique en analysant les données recensées dans le lot 1 et non traitées
- Proposer une analyse critique des jeux de données acquis et notamment de leur représentativité spatio-temporelle
- Identifier, sélectionner et développer une ou des méthodes d'analyse des données permettant leur analyse combinée pour répondre aux objectifs du programme MIGRALION

Aujourd'hui, aucune technologie ne permet de collecter l'ensemble des informations nécessaires pour décrire les flux migratoires en quatre dimensions (longitude, latitude, altitude et temps) à l'échelle d'une zone aussi grande que celle du golfe de Lion. Les lots 3 à 5 visent donc à utiliser différentes technologies permettant de collecter des informations complémentaires pour décrire la migration en Méditerranée. Prises indépendamment, ces données ne fournissent qu'une information partielle et limitée sur l'organisation

(notamment spatiale) des flux migratoires dans le golfe du Lion. Leur combinaison formelle dans un cadre d'analyse unifié est donc une nécessité pour le programme MIGRALION. Par ailleurs, la collecte d'informations de différentes natures sur l'utilisation de l'espace par les oiseaux marins (suivi GPS de différentes espèces aux écologies différentes, recensement des colonies, etc.) implique là aussi le développement de méthodes d'analyses permettant la combinaison à la fois des données de natures différentes mais aussi obtenues sur différentes espèces.

En quelques années, la combinaison de données issues de sources hétérogènes s'est imposée comme une démarche de modélisation permettant de tirer le meilleur parti de l'accumulation de données collectées par des approches différentes sur un même système écologique. Cette approche nommée (selon le type de données mobilisées mais aussi selon le degré de formalisation de la combinaison) data fusion, data combination ou data integration, permet en effet de modéliser, dans un même cadre formel, des données de sources différentes, collectées avec des techniques différentes et selon des plans d'échantillonnage eux aussi hétérogènes (voir par exemple Crawford *et al.*, 2018 ; Isaac *et al.*, 2020 ; Miller *et al.*, 2019 ; Pacifici *et al.*, 2019 ; Zipkin & Saunders, 2018). Des travaux récents, menés par les chercheurs coordinateurs de ce lot ont par exemple montré qu'il était possible de combiner des données de recensements standardisées et des données opportunistes pour acquérir une meilleure précision sur la distribution d'espèces (Renner *et al.*, 2019). Malgré tout, la combinaison de données reste une approche récente et novatrice dont les exemples sont restreints à quelques types de données bien spécifiques (l'approche est maintenant très bien développée en dynamique des populations via la combinaison de données de capture-recapture et de comptages (voir par exemple Arnold *et al.*, 2018 ou Riecke *et al.*, 2019) mais est balbutiante sur les questions de connectivité migratoire (voir par exemple Korner-Nievergelt *et al.*, 2017 ou Zhu *et al.*, 2020) et n'a, à notre connaissance, jamais été développée pour la modélisation des flux d'oiseaux migrants. Des articles récents ont discuté l'intérêt de combiner des données de flux migratoires issus de données collectées par des radars météorologiques avec des données de comptages sur des sites à la côte, mais ces articles ne proposent pas l'approche méthodologique permettant concrètement cette combinaison (Shibley *et al.*, 2018 ; Weisshaupt *et al.*, 2021). Certains travaux se sont intéressés à évaluer la cohérence de différentes sources de données (par exemple observateurs versus radar, Schmidt *et al.*, 2017) mais là encore, il ne s'agissait pas de combinaison. De ce fait, le travail à mener sur ce lot 6 visant à combiner/intégrer les nombreuses sources de données, hétérogènes, dans un même cadre relève d'une démarche de recherche pour laquelle il est à ce stade difficile de déterminer jusqu'où cette intégration sera possible. Le groupe de chercheurs mobilisés pour ce lot dispose cependant d'une expertise de très haut niveau sur ces questions d'intégration de données et maîtrise les différents outils de modélisation utilisés pour chaque source de données différente.

2 Oiseaux marins : intégration des différentes campagnes d'acquisition de données

2.1 Développement d'un modèle intégrant les données de comptages/d'observation en mer

Des données de comptages/ d'observations en mer peuvent être utilisées pour caractériser l'usage de l'espace par les oiseaux marins dans le golfe du Lion. A partir de ces données, collectées par des observateurs en bateau ou en avion parcourant des transects, la distribution des espèces peut être prédite à l'aide de différentes méthodes de modélisation statistique généralement regroupées sous le terme générique de "*Species Distribution*" Models (Elith & Leathwick, 2009 ; Miller, 2010). Dans le cadre de MIGRALION, au vu de l'hétérogénéité des sources de données, nous avons fait le choix d'utiliser des modèles dits "*d'occupancy*" (MacKenzie *et al.*, 2002), en "dégradant" les données de comptages en simple données de détection / non-détection.

Dans le golfe du Lion, plusieurs suivis ont été réalisés par différentes organisations ces dernières années (cf livrable L33). Mobiliser l'ensemble de ces données disponibles permet de produire des prédictions plus fiables et plus précises que ce qui serait obtenu d'une analyse reposant sur un seul de ces jeux de données (Isaac *et al.*, 2020). Néanmoins, chaque jeu de données est associé à des sources de biais intrinsèques, notamment en ce qui concerne les sources d'erreur d'observations (effet de la distance à l'observateur, effort d'échantillonnage spatial, météo). Ceci limite la pertinence de l'utilisation d'un modèle unique pour toutes les sources de données (Miller *et al.*, 2019). Afin de dépasser cette contrainte, nous utilisons des modèles d'*occupancy* dits "intégrés" (Fletcher *et al.*, 2019). Il s'agit de modèles statistiques qui peuvent combiner plusieurs sources de données dans un unique cadre de modélisation hiérarchique (Tyre *et al.*, 2003 ; MacKenzie *et al.*, 2002). Ces modèles sont implémentés dans le logiciel R (R Core Team, 2024) dans le package *spOccupancy* (Doser *et al.*, 2022), suffisamment flexible pour les jeux de données dont nous disposons. C'est donc cette méthode et ce package que nous avons utilisé.

Le principe de ces modèles hiérarchiques est de séparer la modélisation du processus d'occurrence, commune à tous les jeux de données, du processus de détection, spécifique à chaque jeu de données. Si on note Z_j la présence ou l'absence d'une espèce sur le site j , alors cette variable d'occurrence résulte d'un processus de Bernoulli suivant :

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

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

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

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

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

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

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

Les covariables utilisées pour modéliser le processus d'occurrence décrivent l'habitat des espèces considérées et fournissent des indications indirectes (proxy) de l'abondance des ressources dans chaque site. Nous avons adopté un processus de sélection de modèle afin d'identifier les covariables les plus pertinentes pour prédire la distribution des différentes espèces d'oiseaux marins étudiées. À chaque étape (voir tableau ci-dessous), les modèles sont évalués en fonction de leur WAIC (Critère d'Information Largement Applicable, Watanabe 2010) et de leur performance prédictive, évaluée par une étape de la validation croisée et en utilisant la déviance comme critère de notation (Hooten & Hobbs, 2015).

Dans un premier temps, nous explorons la possibilité de transformer logarithmiquement chaque variable ou d'ajouter un effet quadratique dans un modèle contenant uniquement la variable testée (Tableau 1). Ensuite, en retenant les transformations logarithmiques ou les effets quadratiques des variables les plus appropriées, nous organisons les covariables en trois catégories : statiques, dynamiques et liées à la température

de l'eau. Pour chaque catégorie, toutes les combinaisons possibles des covariables sont testées, et la meilleure combinaison est sélectionnée. Les meilleures combinaisons de chaque bloc sont ensuite assemblées pour ajuster le modèle final. Enfin, dans une dernière étape, nous examinons si l'incorporation d'une structure d'autocorrélation spatiale améliore le modèle final.

Tableau 1 : ILLUSTRATION DU PROCESSUS DE SELECTION DES COVARIABLES AVEC LES DONNEES SUR LA STERNE CAUGEK EN DEHORS DE SA PERIODE DE REPRODUCTION.

Etape 1 : Tester l'ajout d'un effet quadratique ou une transformation logarithmique. Présenté ici avec la distance à la côte.								
Modèle	WAIC				Performance prédictive (validation croisée)			
	PELMED	Megaobs	Migralion	Total	PELMED	Megaobs	Migralion	Total
1	635	143	467	1245	644	145	471	1260
dist_to_shore	582	130	462	1174	589	132	456	1177
dist_to_shore + l(dist_to_shore)^2	582	130	463	1175	589	132	457	1178
log_dist_to_shore	571	126	451	1148	579	126	438	1143

Etape 2 : Tester plusieurs associations de covariables en séparant toutes les covariables en 3 blocs (covariables statiques, dynamiques et liées à la température). Seul le bloc des covariables statiques est représenté ici.								
Modèle	WAIC				Performance prédictive (validation croisée)			
	PELMED	Megaobs	Migralion	Total	PELMED	Megaobs	Migralion	Total
1	635	143	467	1245	643	145	472	1260
log_dist_to_shore	572	125	451	1148	579	127	439	1145
concavity	640	143	465	1248	646	145	476	1267
log_bathymetry	570	126	441	1137	580	127	438	1145
log_dist_to_shore + concavity	573	126	451	1150	581	127	441	1149
log_dist_to_shore + log_bathymetry	563	123	444	1130	573	124	437	1134
concavity + log_bathymetry	573	126	438	1137	583	127	436	1146
log_dist_to_shore + concavity + log_bathymetry	565	123	443	1131	577	124	438	1139

Etape 3 : Tester l'ajout des différents blocs ensemble								
Modèle	WAIC				Performance prédictive (validation croisée)			
	PELMED	Megaobs	Migralion	Total	PELMED	Megaobs	Migralion	Total
1	635	143	467	1245	643	145	472	1260
log_dist_to_shore + log_bathymetry	563	123	445	1131	572	124	436	1132
mean_CHL + mean_SSH	536	123	406	1065	540	120	391	1051
mean_winter_SST + mean_spring_SST + mean_summer_SST	547	124	409	1080	559	122	402	1083
log_dist_to_shore + log_bathymetry + mean_CHL + mean_SSH	536	118	408	1062	543	116	392	1051
log_dist_to_shore + log_bathymetry + mean_winter_SST + mean_spring_SST + mean_summer_SST	542	121	414	1077	553	119	402	1074
mean_CHL + mean_SSH + mean_winter_SST + mean_spring_SST + mean_summer_SST	532	120	406	1058	538	119	393	1050
log_dist_to_shore + log_bathymetry + mean_CHL + mean_SSH + mean_winter_SST + mean_spring_SST + mean_summer_SST	532	120	408	1060	541	119	394	1054

Etape 4 : Tester l'ajout d'une structure d'autocorrélation spatiale								
Modèle	WAIC				Performance prédictive (validation croisée)			
	PELMED	Megaobs	Migralion	Total	PELMED	Megaobs	Migralion	Total
Sans effet spatial	532	120	409	1061	540	119	395	1054
Effet spatial exponentiel	517	112	357	986	657	148	469	1274
Effet spatial sphérique	519	113	360	992	664	150	468	1282
Effet spatial gaussien	530	119	392	1041	657	150	469	1276

2.2 Développement d'un modèle intégrant télémétrie et données de comptages/d'observation en mer

2.1.1 Contexte, enjeux

La télémétrie et les comptages/observations en mer fournissent des informations spatiales utiles pour comprendre comment les animaux utilisent l'espace. L'accessibilité croissante des dispositifs de télémétrie et l'abondance des comptages de population rendent les données de déplacement et de comptages largement disponibles pour estimer la sélection de l'habitat (Matthiopoulos *et al.*, 2022 ; Northrup *et al.*, 2022 ; Thurfjell *et al.*, 2014). Les données de comptage et de déplacement coexistent ainsi dans de multiples contextes écologiques (par exemple, sur les ongulés : Gamelon *et al.* (2021), Fradin & Chamaillé-Jammes (2023), Chandler *et al.* (2022) ; ou les oiseaux marins : Laran *et al.* (2017), Courbin *et al.* (2023)). Ces deux sources de données permettent aussi de capturer des aspects complémentaires de la sélection de l'habitat, la télémétrie permettant un suivi très détaillé de quelques individus alors que les comptages permettent de suivre théoriquement l'ensemble de la population avec des résolutions temporelles plus grossières. Toutefois, la comparaison et la combinaison de ces deux types de données n'est pas triviale bien qu'attrayante (Matthiopoulos *et al.*, 2022). Télémétrie et comptages se collectent sur des échelles écologiques différentes et la différence fondamentale dans la nature des données enregistrées constitue ainsi un défi important pour comparer et combiner ces données (Lindenmayer *et al.*, 2022 ; Nathan *et al.*, 2022 ; Michelot *et al.*, 2019).

L'intégration (ou la combinaison) de multiples jeux de données est un axe de recherche intense de l'écologie statistique actuelle (Zipkin *et al.*, 2021 ; Kéry & Royle, 2020). Le principe de base de la plupart des modèles intégrés est que les jeux de données multiples décrivent le même processus écologique sous-jacent. La plupart des modèles intégrés incorporent ainsi un processus écologique identique décrit par l'ensemble de jeux de données impliqués, et un processus d'observation différent correspondant à chaque protocole de surveillance (Kéry & Royle, 2020 ; Fletcher *et al.*, 2019). Les deux principaux avantages des modèles intégrés sont i) une plus grande précision des paramètres d'intérêt et ii) dans certaines situations, la possibilité d'estimer des paramètres qui ne pourraient pas être déduits de chaque source de données seule (Kéry & Royle, 2020 ; Zipkin *et al.*, 2019). Les développements méthodologiques sur l'intégration des données ont permis de combiner de nombreux types de données différents (voir Kéry & Royle (2020) pour une revue). Récemment, les données de mouvement ont été incluses dans les modèles de capture-recapture spatialisé afin de mieux informer le processus des centres d'activité (Converse *et al.*, 2022). En utilisant des données de comptage, de présence et/ou d'absence collectées dans le cadre de multiples programmes de surveillance, les modèles intégrés de distribution d'espèces modélisent statistiquement un processus conjoint d'abondance ou de distribution via une régression linéaire des covariables environnementales, informant ainsi conjointement les coefficients β de sélection de l'habitat (Strebel *et al.*, 2022 ; Miller *et al.*, 2019). Les modèles de distribution et de déplacement des espèces permettent tous deux d'estimer la force de sélection relative de l'habitat par rapport à un autre. Aarts *et al.* (2012) ont par ailleurs montré la convergence des modèles de Ressource Selection Function RSF (ou Step Selection Function, SSF) utilisés classiquement pour modéliser les données de déplacement/télémétrie avec les modèles de processus ponctuels inhomogènes classiquement utilisés pour l'analyse des données de comptage. Pour analyser des données de comptages, via ces processus ponctuels inhomogènes, on ajuste un modèle linéaire généralisé (GLM) avec une distribution de Poisson. Un GLM de Poisson permet d'estimer le nombre d'individus attendus en chaque point de l'aire d'étude. Le GLM de Poisson nécessite de discrétiser l'aire d'étude en une multitude de cellules (voir article en annexe). Comme les modèles de RSF ainsi que les GLM de Poisson estiment la sélection de l'habitat via une combinaison linéaire de prédicteurs environnementaux (biotiques ou

abiotiques), il est possible de comparer la force de sélection relative de l'habitat estimée par leurs coefficients β respectifs. Cette convergence du coefficient de sélection relative de l'habitat β entre les différents types de modèles ouvre des perspectives pour développer formellement un modèle intégré qui utilise conjointement les deux sources de données (télémétrie et comptages/observations en mer).

Nous avons sur cette base développé un modèle intégré permettant d'estimer la sélection relative de l'habitat intégrant les données de télémétrie via des modèles de *Resource Selection Function* (RSF) et les données de comptage via un modèle linéaire généralisé (GLM) de Poisson.

2.1.2 Présentation de la méthodologie intégrée

Un processus ponctuel de Poisson inhomogène (PPI) est un modèle qui décrit la distribution stochastique de points aléatoires dans l'espace où la densité attendue des points est corrélée à des covariables spatiales. Il estime une fonction d'intensité qui renvoie λ_s , le nombre attendu de points à chaque emplacement infinitésimal s dans l'espace S . L'intensité attendue λ peut s'écrire sous la forme d'une fonction log-linéaire des prédicteurs environnementaux X :

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

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

Pour estimer la sélection de l'habitat à partir des données de comptage, nous ajustons un GLM de Poisson sur un espace géographique discret où le comptage attendu N dans la maille suit l'intensité λ d'un IPP (Aarts *et al.*, 2012). Nous avons utilisé des données de comptage collectées sur plusieurs mailles pour constituer N un ensemble de données de comptage. N_s représente les comptages attendus dans la maille s et est estimé comme un tirage dans une loi de Poisson dans λ_s .

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

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

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

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

Les RSF utilisés pour modéliser les données de télémétrie sont classiquement ajustées à l'aide d'une régression logistique dans laquelle les localisations des trajets de télémétrie sont codées comme des 1 et un jeu de points constituant l'habitat disponible (ou des points de "pseudo-absence") comme des 0. Appelons K l'ensemble de données qui stocke les données binaires 1/0. Les RSF opèrent dans un espace géographique continu, et K_s pour chaque point s dans S est modélisé comme un tirage de Bernoulli dans l'intensité IPP λ de l'équation :

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

Pour tenir compte de la variation de la sélection de l'habitat entre les individus I , nous avons inclus un effet aléatoire individuel sur les paramètres de pente β'_s avec

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

pour $i \in [1, I]$

β étant le paramètre d'intérêt décrivant la force relative de sélection de l'habitat de la covariable X pour la population, τ représente la variance interindividuelle, c'est-à-dire l'hétérogénéité au sein de la population, de la sélection de l'habitat de la covariable X .

Malgré le fait que les RSFs et les GLMs de Poisson quantifient la sélection de l'habitat par une régression linéaire généralisée sur les covariables X , ils n'estiment pas la même intensité absolue de l'IPP λ (Northrup *et al.*, 2022 ; Aarts *et al.*, 2012). Dans les RSF, l'intercept β_0 n'a pas de signification et les densités d'utilisation de l'espace résultantes du processus de modélisation ne peuvent être interprétées qu'en relatif et non en absolu (Avgar *et al.*, 2017). Le GLM de Poisson et les RSFs convergent uniquement pour estimer l'intensité relative de l'IPP (βX), c'est-à-dire que les β estimés par une RSF sont cohérents avec les β estimés par le GLM de Poisson (Aarts *et al.*, 2012). Au contraire, l'intercept des GLMs peut permettre d'obtenir une intensité d'utilisation de l'espace absolue.

Ainsi, nous construisons un modèle intégré qui estime conjointement les β qui reflètent la sélection dans l'IPP. Le modèle intégré conserve la structure hiérarchique du GLM de Poisson et la structure de régression logistique de la RSF, incluant ainsi deux modèles similaires à l'IPP avec des intercepts spécifiques et des coefficients de sélection de l'habitat partagés.

Le sous-modèle du GLM de Poisson :

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

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

Le sous-modèle de la RSF :

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

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

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

2.3 Oiseaux marins et altitude de vol

Trois types de données nous fournissent des informations concernant les hauteurs de vol. Tout d'abord, bien que la majorité des données de comptages (e.g. SAMM, PELMED) ne contiennent pas cette information, cette donnée a été renseignée lors des comptages par observations visuelles des campagnes en mer MIGRALION.

Néanmoins, cette information n'étant pas disponible pour tous les oiseaux détectés, des analyses préliminaires seront nécessaires pour s'assurer de l'utilisation potentielle de ces données dans le cadre d'une analyse intégrée. A noter également que les hauteurs relevées de cette façon sont peu précises, et ne recourent que partiellement les altitudes mesurées par radar. Ensuite, nous disposons des hauteurs de vol pour les individus équipés de GPS. Il sera possible d'analyser l'altitude de vol des oiseaux marins équipés en fonction de variables environnementales et météorologiques (Péron *et al.*, 2020 ; Ross-Smith *et al.*, 2016). Enfin, 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 .

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

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

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

3 Spatialisation des flux de migrateurs

3.1 Intégration des données de radar météo et de radar à la côte via un IPP

Un des objectifs du programme MIGRALION consiste à développer des méthodes permettant l'analyse jointe de données radars (avifaunistiques), comptages (transects dans le golfe du Lion) et télémétrie dans le but d'estimer la variation spatiale des flux migratoires à l'échelle du golfe du Lion.

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, et dans une moindre mesure les radars à la côte, nous renseignent sur cette variation des flux dans l'espace (Liechti *et al.*, 2019). Cependant, les données GPS/GLS ne peuvent pas correctement renseigner spatialement 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 (ils se trouvent quasi exclusivement dans la moitié Est de la zone d'étude : 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 (Rakhimberdiev *et al.*, 2016). Les données issues des radars à la côte ne fournissent quant à eux que des informations sur des intensités de flux migratoires sur deux localisations géographiques en simultané. Les possibilités de modélisation intégrée des flux migratoires semblent donc limitées dans le cadre de MIGRALION.

MIGRATLANE, le pendant Atlantique du programme MIGRALION, lancé en 2022, s'attache aussi à décrire spatialement l'intensité des flux migratoires à l'échelle de l'arc atlantique français. Le lot 6 de ces deux programmes étant sous la responsabilité de la même équipe de recherche, les développements méthodologiques sont conjoints. Cependant, à l'inverse de MIGRALION, l'intégration des données provenant de radars météorologiques est prévue dans le programme MIGRATLANE. Les données récoltées dans le cadre de MIGRALION ne comportant que peu d'informations spatiales sur l'intensité des flux migratoires, l'intégration des données provenant des radars météorologiques sur le golfe du Lion permettrait d'avoir plus d'informations spatiales sur l'intensité des flux migratoires et d'améliorer nos estimations d'utilisation de l'espace (voir l'exemple de la partie suivante). Ainsi, la méthode présentée ci-après intègre l'information provenant de radars météorologiques dans le cadre d'analyse envisagée, même si à ce stade, la disponibilité de ces données pour MIGRALION est incertaine. En effet, ce développement méthodologique semble être le seul à même de permettre une modélisation intégrée formelle des flux migratoires.

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

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

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

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

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

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

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

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

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

Cette modélisation des processus de détection s'applique plus particulièrement aux sources de données générant des données de comptages ou pouvant être traduites comme des données de comptages : radars météorologiques et à la côte, campagnes en mer.

Malgré les biais liés aux zones de provenance des oiseaux équipés de GPS dans le cadre de MIGRALION, l'intégration de ces données dans ce cadre d'analyse est envisageable. Il a en effet été montré que les modèles utilisés pour analyser des données GPS (Resource Selection Function; voir sections précédentes) peuvent aussi être écrits sous la forme d'un processus spatial de Poisson non-homogène (Aarts *et al.*, 2012). Dans ce cas, les données GPS des oiseaux migrateurs équipés pourront être intégrés dans le modèle pour permettre de mieux estimer λ_i et les paramètres associés au calcul de ce dernier (μ_0 et $\beta_{1,\dots,k}$).

3.2 Exemple d'application du modèle précédent sur des données simulées

Dans cet exemple, nous prenons le cas du golfe du Lion, dans lequel nous simulons un axe de migration sud-ouest/nord-est passant légèrement au large des côtes (Figure 1). De ces données simulées, deux jeux de données d'observations radars sont générés : un avec seulement les radars BirsScan sur la côte, et un autre avec les radars BirdScan ainsi que les radars météorologiques disponibles dans la région (Figure 1).

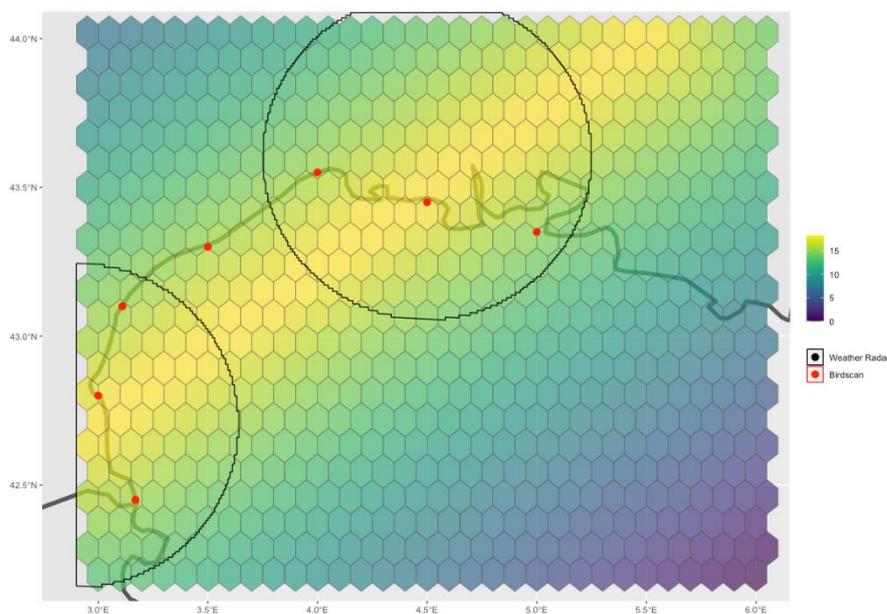


Figure 1 : SIMULATION D'UNE INTENSITE MIGRATOIRE SUR LE GOLFE DU LION ET DE LA POSITION DISPOSITIFS DE RECOLTE DE DONNEES RADARS.

Avec ces deux jeux de données différents, nous utilisons notre modèle précédent pour tenter d'estimer la relation entre l'intensité migratoire et les coordonnées géographiques. Une fois cette relation estimée, nous comparons la capacité de prédiction des deux modèles sur toute la zone d'étude (zone dans laquelle nous avons simulé l'intensité migratoire Figure 1). Si on compare la prédiction issue du modèle n'utilisant que les données des radars BirdScan à la côte (Figure 2) avec les données simulées (Figure 1), on voit que le patron spatial est mal estimé (l'axe de migration sud-ouest/nord-est n'apparaît pas sur la prédiction). Ceci est dû à une mauvaise estimation de la relation entre l'intensité migratoire et les coordonnées géographiques, et ce probablement à cause du manque d'information spatiale générée par les radars à la côte (les informations de ce type de radar étant très ponctuelles). Si on s'intéresse maintenant à la comparaison entre la prédiction issue du modèle avec radars météorologiques en plus (Figure 3) et les données simulées (Figure 1), on remarque que le patron d'intensité migratoire est beaucoup mieux estimé. On retrouve cet axe sud-ouest/nord-est présent dans les simulations. Cet exemple sur données simulées démontre la pertinence du cadre d'analyse décrit dans la partie précédente mais il met aussi en évidence le manque d'information spatiale produite par les données de radars BirdScan à la côte.

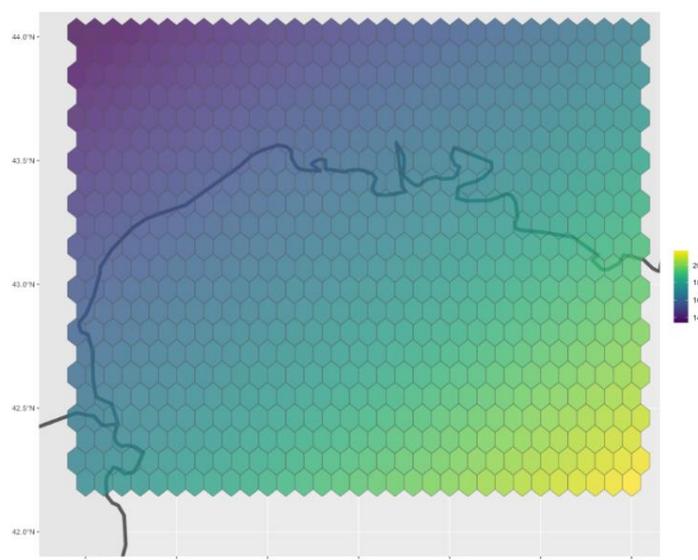


Figure 2 : PREDICTION DE L'INTENSITE MIGRATOIRE SUR TOUTE LA ZONE D'ETUDE UNIQUEMENT A PARTIR DES DONNEES ISSUES DES RADARS BIRDSCAN A LA COTE.

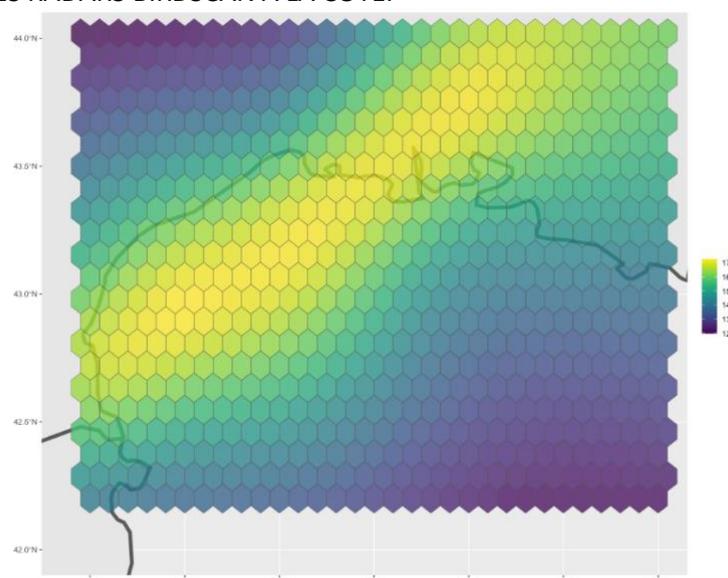


Figure 3 : PREDICTION DE L'INTENSITE MIGRATOIRE SUR TOUTE LA ZONE D'ETUDE A PARTIR DES DONNEES ISSUES DES RADARS METEOROLOGIQUES AINSI QUE DES RADARS BIRDSCAN A LA COTE.

3.3 Autres méthodes disponibles dans la littérature

Différentes approches modélisant les flux migratoires sont déjà disponibles dans la littérature. En effet, des auteurs ont récemment développé une méthode d'intégration de données provenant à la fois de science participative (eBird), de baguage, et de télémétrie, pour caractériser des patrons de migration à l'échelle du continent américain (Meehan *et al.*, 2022, voir Figure 4). Plus précisément, les auteurs conduisent tout d'abord des analyses de connectivité migratoire à l'aide des données de baguage et de télémétrie (Lin *et al.*, 2020 ; Schuster *et al.*, 2019). Ensuite, avec ces résultats de connectivité migratoire, ils simulent des trajets entre aires de reproduction et d'hivernage (Least-cost path analysis, Adriaenssen *et al.*, 2003). Enfin, à travers l'utilisation de GAMMs (Generalized Additive Mixed Models), les auteurs réalisent une analyse jointe des données de télémétrie, de présence/absence (eBird) et des trajets simulés (least-cost paths) afin de représenter les patrons spatio-temporels de déplacements d'une espèce à l'échelle d'une voie de migration.

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

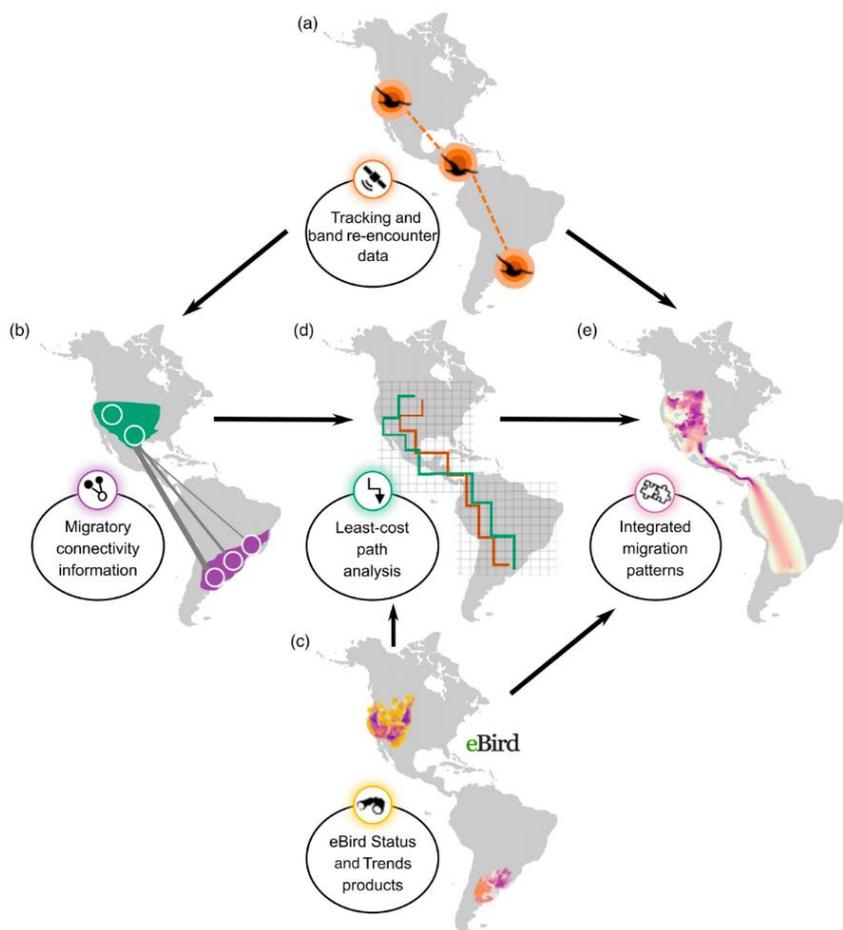


Figure 4 : SCHEMA REPRESENTANT LE CADRE DE L'ANALYSE INTEGREE PRESENTEE DANS MEEHAN ET AL. 2022.

4 Phénologie, altitudes et directions de vol de la migration

4.1 Migrateurs terrestres et phénologie de la migration

Le golfe du Lion est un carrefour migratoire pour de nombreuses espèces qui se reproduisent en Europe et rejoignent l'Afrique ou le pourtour méditerranéen durant l'hiver (Barboutis *et al.*, 2022). Bien que l'on définisse classiquement deux périodes migratoires "pré-nuptiale" (février à juin) et "post-nuptiale" (août à novembre), le flux de migrants n'est pas constant au cours de ces périodes. En effet, on observe des pics migratoires qui correspondent à des nuits de passages intenses coïncidant généralement avec de bonnes conditions météorologiques (Haest *et al.*, 2019).

Il existe différentes approches méthodologiques pour définir les périodes de migrations. Ce choix est important car la fenêtre temporelle d'étude pour les variables réponses considérées (MTR, directions et hauteurs de vol) dépend de la définition des dates de début et fin. Dans un cadre intégratif, les différentes sources de données apportent des informations complémentaires sur ces dates pour notre zone d'étude, et leur calcul nécessitera des approches méthodologiques différentes.

Dans le cas des données radar BirdScan, une possibilité est de construire un index combinant intensité de migration (MTR) et concentration des directions (Hirschhofer *et al.*, 2024). Concernant le traitement des données télémétriques qui décrivent la trajectoire d'individus en continu, plusieurs approches pourront être envisagées, selon l'échelle spatiale considérée, en se basant par exemple sur les longueurs de pas ou concentration des directions (Thompson *et al.*, 2024). Il sera intéressant de tester un éventuel effet de la distance entre sites sur l'autocorrélation temporelle entre pics migratoires, autrement dit voir si des sites proches dans l'espace sont plus synchrones.

4.2 Migrateurs terrestres et altitude de vol

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.

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

Pour prendre en compte les variations d'altitude de vol au cours de la migration, nous inclurons des covariables environnementales abiotiques (e.g. la distance à la côte, profondeur) 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 (Panuccio *et al.*, 2019 ; Shamoun-Baranes *et al.*, 2017). Ce modèle pourra également servir de base pour caractériser des différences entre groupes d'espèces et détailler la phénologie de vol, tant au niveau du cycle journalier qu'au cours de la saison (Wills, 2017 ; La Sorte *et al.*, 2015).

4.3 Migrateurs terrestres et directions de vol

Les données issues des suivis GPS et suivis radars des lots 4 et 5 nous renseignent sur les directions de vol des migrants. Les données GPS fournissent les directions de vol des oiseaux équipés tout au long de leur parcours. Les radars verticaux ornithologiques relèvent, quant à eux, les directions de vol à la côte et en mer.

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

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

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

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

Cette analyse permettra d'obtenir une prédiction des directions de vol en fonction des conditions atmosphériques et de paramètres géographiques. Ce modèle pourra également servir de base pour caractériser des différences entre groupes d'espèces et détailler la phénologie des directions, tant au niveau du cycle journalier qu'au cours de la saison (Hirschhofer *et al.*, 2024 ; Cabrera-Cruz *et al.*, 2017 ; Welcker *et al.*, 2017).

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

Références bibliographiques

- Aarts, G., Fieberg, J., & Matthiopoulos, J. (2012). Comparative interpretation of count, presence–absence and point methods for species distribution models. *Methods in Ecology and Evolution* 3: 177–187.
- Adriaensen, F., Chardon, J. P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., & Matthysen, E. (2003). The application of 'least-cost' modelling as a functional landscape model. *Landscape and urban planning* 64: 233–247.
- Archibald, K. M., Buler, J. J., Smolinsky, J. A. and Smith, R. J. (2017). Migrating birds reorient toward land at dawn over the Great Lakes, USA. *The Auk* 134: 193–201.
- Arnold, T. W., Clark, R. G., Koons, D. N., & Schaub, M. (2018). Integrated population models facilitate ecological understanding and improved management decisions. *The Journal of Wildlife Management* 82: 266–274.
- Avgar, T., Lele, S. R., Keim, J. L., & Boyce, M. S. (2017). Relative Selection Strength: Quantifying effect size in habitat- and step-selection inference. *Ecology and Evolution* 7: 5322–5330.
- Barboutis, C., Navarrete, E., Karris, G., Xirouchakis, S., Fransson, T. and Bounas, A. (2022). Arriving depleted after crossing of the Mediterranean: obligatory stopover patterns underline the importance of Mediterranean islands for migrating birds. *Animal Migration* 9: 14–23.
- Cabrera-Cruz, S. A., Mabee, T. J. and Villegas-Patraca, R. 2017. Patterns of nocturnal bird migration in southern Mexico. *Revista Mexicana de Biodiversidad* 88: 867–879.
- Chandler, R. B., Crawford, D. A., Garrison, E. P., Miller, K. V., & Cherry, M. J. (2022). Modeling abundance, distribution, movement and space use with camera and telemetry data. *Ecology* 103: e3583.
- Converse, S. J., McClintock, B. T., & Conn, P. B. (2022). Special Feature: Linking capture-recapture and movement. *Ecology*: e3770.
- Courbin, N., Besnard, A., Boncourt, E., & Grémillet, D. (2023). Flight heights of Scopoli's shearwaters *Calonectris diomedea* in the context of offshore wind farm developments. *BioRxiv* 2023:05.
- Crawford, B. A., Moore, C. T., Norton, T. M. and Maerz, J. C. (2018). Integrated analysis for population estimation, management impact evaluation, and decision-making for a declining species. *Biological Conservation* 222: 33–43.
- Cremers, J., Mulder, K. T. and Klugkist, I. (2018). Circular interpretation of regression coefficients. *Brit J Math & Statis* 71: 75–95.
- Cremers, J. and Klugkist, I. (2018). One Direction? A Tutorial for Circular Data Analysis Using R With Examples in Cognitive Psychology. *Front. Psychol.* 9: 2040.
- Doser, J.W., Finley, A.O., Kéry, M., Zipkin, E.F., (2022). spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models. *Methods Ecol. Evol.* 13: 1670–1678.
- Elith, J., Leathwick, J.R., (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu. Rev. Ecol. Evol. Syst.* 40: 677–697.
- Fletcher Jr., 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* 100: e02710.

- Fradin, G., & Chamaillé-Jammes, S. (2023). Hogs sleep like logs: Wild boars reduce the risk of anthropic disturbance by adjusting where they rest. *Ecology and Evolution* 13: e10336.
- Gamelon, M., Nater, C. R., Baubet, É., Besnard, A., Touzot, L., Gaillard, J.-M., Lebreton, J.-D., & Gimenez, O. (2021). Efficient use of harvest data: A size-class-structured integrated population model for exploited populations. *Ecography* 44: 1296–1310.
- Haest, B., Hüppop, O., Van De Pol, M. and Bairlein, F. (2019). Autumn bird migration phenology: A potpourri of wind, precipitation and temperature effects. *Global Change Biology* 25: 4064–4080.
- Hirschhofer, S., Liechti, F., Ranacher, P., Weibel, R. and Schmid, B. (2024). High-intensity bird migration along Alpine valleys calls for protective measures against anthropogenically induced avian mortality. *Remote Sens Ecol Conserv: rse2.377*.
- Hooten, Mevin B, and N Thompson Hobbs. (2015). A Guide to Bayesian Model Selection for Ecologists. *Ecological Monographs* 85: 3–28.
- 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., (2020). Data Integration for Large-Scale Models of Species Distributions. *Trends Ecol. Evol.* 35: 56–67.
- 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.
- Korner-Nievergelt, F., Prévot, C., Hahn, S., Jenni, L., & Liechti, F. (2017). The integration of mark re-encounter and tracking data to quantify migratory connectivity. *Ecological Modelling* 344: 87-94.
- 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.
- La Sorte, F. A., Hochachka, W. M., Farnsworth, A., Sheldon, D., Van Doren, B. M., Fink, D. and Kelling, S. 2015. Seasonal changes in the altitudinal distribution of nocturnally migrating birds during autumn migration. *R. Soc. open sci.* 2: 150347.
- Liechti, F., Aschwanden, J., Blew, J., Boos, M., Brabant, R., Dokter, A. M., ... & Sapir, N. (2019). Cross-calibration of different radar systems for monitoring nocturnal bird migration across Europe and the Near East. *Ecography* 42: 887-898.
- Lin, H. Y., Schuster, R., Wilson, S., Cooke, S. J., Rodewald, A. D., & Bennett, J. R. (2020). Integrating season-specific needs of migratory and resident birds in conservation planning. *Biological Conservation* 252: 108826.
- Lindenmayer, D. B., Lavery, T., & Scheele, B. C. (2022). Why We Need to Invest in Large-Scale, Long-Term Monitoring Programs in Landscape Ecology and Conservation Biology. *Current Landscape Ecology Reports*.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Andrew Royle, J., Langtimm, C.A., 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83: 2248–2255.
- 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: 909–920.

- Meehan, T. D., Saunders, S. P., DeLuca, W. V., Michel, N. L., Grand, J., Deppe, J. L., ... & Wilsey, C. B. (2022). Integrating data types to estimate spatial patterns of avian migration across the Western Hemisphere. *Ecological applications* 32: e2679.
- Michelot, T., Blackwell, P. G., & Matthiopoulos, J. (2019). Linking resource selection and step selection models for habitat preferences in animals. *Ecology* 100: e02452.
- 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 Ecol. Evol.* 10: 22–37.
- Miller, J., (2010). Species Distribution Modeling. *Geogr. Compass* 4: 490–509.
- Nathan, R., Monk, C. T., Arlinghaus, R., Adam, T., Alós, J., Assaf, M., Baktoft, H., Beardsworth, C. E., Bertram, M. G., Bijleveld, A. I., Brodin, T., Brooks, J. L., Campos-Candela, A., Cooke, S. J., Gjelland, K. Ø., Gupte, P. R., Harel, R., Hellström, G., Jeltsch, F., ... Jarić, I. (2022). Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science* 375: eabg1780.
- Pacifici, K., Reich, B. J., Miller, D. A., & Pease, B. S. (2019). Resolving misaligned spatial data with integrated species distribution models. *Ecology* 100: e02709.
- Panuccio, M., Dell'Omo, G., Bogliani, G., Catoni, C. and Sapir, N. (2019). Migrating birds avoid flying through fog and low clouds. *Int J Biometeorol* 63: 231–239.
- Péron, G., Calabrese, J. M., Duriez, O., Fleming, C. H., García-Jiménez, R., Johnston, A., Lambertucci, S. A., Safi, K. and Shepard, E. L. C. (2020). The challenges of estimating the distribution of flight heights from telemetry or altimetry data. *Anim Biotelemetry* 8: 5.
- Rakhimberdiev, E., Senner, N. R., Verhoeven, M. A., Winkler, D. W., Bouten, W., & Piersma, T. (2016). Comparing inferences of solar geolocation data against high-precision GPS data: annual movements of a double-tagged black-tailed godwit. *Journal of Avian Biology* 47: 589-596.
- Renner, I. W., Louvrier, J., & Gimenez, O. (2019). Combining multiple data sources in species distribution models while accounting for spatial dependence and overfitting with combined penalized likelihood maximization. *Methods in Ecology and Evolution* 10: 2118-2128.
- Riecke, T. V., Williams, P. J., Behnke, T. L., Gibson, D., Leach, A. G., Sedinger, B. S., Street, P. A. and Sedinger, J. S. (2019). Integrated population models: Model assumptions and inference (R Freckleton, Ed.). *Methods Ecol Evol* 10: 1072–1082.
- Ross-Smith, V. H., Thaxter, C. B., Masden, E. A., Shamoun-Baranes, J., Burton, N. H., Wright, L. J., ... & Johnston, A. (2016). Modelling flight heights of lesser black-backed gulls and great skuas from GPS: A Bayesian approach. *Journal of Applied Ecology* 53: 1676-1685.
- Royle, J. A., & Nichols, J. D. (2003). Estimating abundance from repeated presence-absence data or point counts. *Ecology* 84: 777–790.
- Schmidt, M., Aschwanden, J., Liechti, F., Wichmann, G., & Nemeth, E. (2017). Comparison of visual bird migration counts with radar estimates. *Ibis* 159: 491-497.
- Schuster, R., Wilson, S., Rodewald, A. D., Arcese, P., Fink, D., Auer, T., & Bennett, J. R. (2019). Optimizing the conservation of migratory species over their full annual cycle. *Nature communications* 10: 1754
- Shamoun-Baranes, J., Liechti, F. and Vansteelant, W. M. G. (2017). Atmospheric conditions create freeways, detours and tailbacks for migrating birds. *J Comp Physiol A* 203: 509–529.

- Shiple, J. R., Kelly, J. F., & Frick, W. F. (2018). Toward integrating citizen science and radar data for migrant bird conservation. *Remote Sensing in Ecology and Conservation* 4: 127-136.
- Strebel, N., Kéry, M., Guélat, J., & Sattler, T. (2022). Spatiotemporal modelling of abundance from multiple data sources in an integrated spatial distribution model. *Journal of Biogeography* jbi.14335.
- Thompson, P. R., Harrington, P. D., Mallory, C. D., Lele, S. R., Bayne, E. M., Derocher, A. E., Edwards, M. A., Campbell, M. and Lewis, M. A. (2024). Simultaneous estimation of the temporal and spatial extent of animal migration using step lengths and turning angles. *Mov Ecol* 12: 1.
- Thurfjell, H., Ciuti, S., & Boyce, M. S. (2014). Applications of step-selection functions in ecology and conservation. *Movement Ecology* 2: 4.
- Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K., Possingham, H.P., (2003). Improving Precision and Reducing Bias in Biological Surveys: Estimating False-Negative Error Rates. *Ecol. Appl.*
- Watanabe, Sumio. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research* 11.
- Weisshaupt, N., Lehtiniemi, T., & Koistinen, J. (2021). Combining citizen science and weather radar data to study large-scale bird movements.
- Welcker, J., Liesenjohann, M., Blew, J., Nehls, G. and Grünkorn, T. (2017). Nocturnal migrants do not incur higher collision risk at wind turbines than diurnally active species. *Ibis* 159: 366–373.
- Wills (2017). Monitoring Avian Migration with Dedicated Vertical-looking Radar.
- Zipkin, E. F. and Saunders, S. P. (2018). Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation* 217: 240–250.
- Zipkin, E. F., Inouye, B. D., & Beissinger, S. R. (2019). Innovations in data integration for modeling populations. *Ecology*, e02713.
- Zipkin, E. F., Zylstra, E. R., Wright, A. D., Saunders, S. P., Finley, A. O., Dietze, M. C., Itter, S., & Tingley, M. W. (2021). Addressing data integration challenges to link ecological processes across scales. *Frontiers in Ecology and the Environment* 9.
- Zhu, Q., Hobson, K. A., Zhao, Q., Zhou, Y., Damba, I., Batbayar, N., ... & David Fox, A. (2020). Migratory connectivity of Swan Geese based on species' distribution models, feather stable isotope assignment and satellite tracking. *Diversity and Distributions* 26: 944-957.

Annexes

Annexe 1 : Article en cours de rédaction concernant le modèle intégrant télémétrie et données de comptages/d'observation en mer : **Integrating telemetry and count data to improve estimation of animal spatial distributon**

Valentin Lauret^{1,2*}, Nicolas Courbin¹, Aurélien Besnard¹

(1) CEFE, Univ Montpellier, CNRS, EPHE PSL University, IRD, Montpellier, France

(2) France Energies Marines

(*) corresponding author: Valentin Lauret, valentin.lauret@cefe.cnrs.fr

Abstract: Both telemetry data and population count contribute to understand how animals select the components of the landscape despite capturing complementary aspects of habitat selection, i.e. animal tracking provide detailed insights on few individuals while counts theoretically sample all the population but resulting in fewer data. Following the statistical framework of Inhomogeneous Poisson Point processes, we suggested that it is possible to integrate telemetry and count to estimate habitat selection.

Based on simulations and Sandwich tern (*Thalasseus sandvicensis*) data collected in the French Mediterranean Sea, we compared habitat selection estimates obtained by Resource Selection Function (RSF) using telemetry data, and by Poisson generalized linear models (GLM) using count data. We developed an integrated model that combine RSF and Poisson GLM.

Our simulations and the sandwich terns case study illustrated that RSF, Poisson GLM, and the integrated model correctly estimated habitat selection while the integrated model achieved better accuracy and precision than models using the data sources separately. Our study represented a novel attempt towards the formal integration of telemetry and population data to estimate habitat selection. We hope to contribute to a promising research avenue as telemetry and count monitoring are abundant in many ecological contexts.

Keywords: animal ecology, habitat selection, data integration, occupancy models, resource selection function, telemetry.

• 1 Introduction

Telemetry tracking and population counts provide insightful spatial information to understand the ecological drivers shaping population distributions within the landscape, hence thriving from animal cognition ([Robira et al., 2021](#)) to conservation application ([Courbin et al., 2022](#); [Thurfjell et al., 2014](#)). Within the last two decades, the increasing accessibility of animal tracking devices and of population monitoring make telemetry data and counts widely available to estimate habitat selection, i.e. the process by which an individual disproportionately selects some environmental features ([Boyce et al., 2002](#); [Matthiopoulos et al., 2022](#); [Muff et al., 2020](#); [Northrup et al., 2022](#); [Thurfjell et al., 2014](#)). And now, such data coexist in multiple ecological contexts (e.g. management

of wild ungulates, Gamelon et al. (2021), Fradin & Chamaillé-Jammes (2023), Chandler et al. (2022); estimating seabirds habitat selection, Yamamoto et al. (2015)). The availability of both count and telemetry data for the same study area allows to capture complementary aspects of habitat selection. Individual animal telemetry data and population counts focus on different ecological scales due to fundamental differences in the nature of their recorded data (Lindenmayer et al., 2022; Michelot et al., 2019; Nathan et al., 2022). Individual tracking typically provided a large amount of data but on a restricted subset of the population while count monitoring programs theoretically sample all individuals of the population despite collecting fewer data. Thus, combining telemetry and count data sources to estimate habitat selection has not yet been carried out, although it is appealing (Matthiopoulos et al., 2022).

The statistical framework of Inhomogeneous Poisson point processes (IPP) describes the intensity function of a spatial pattern of points as a log-linear function of environmental covariates. IPPs are widely used tools to model spatial ecological processes including using telemetry and count data to estimate habitat selection (Illian, 2019; Johnson et al., 2013; Renner et al., 2015).

Telemetry data are classically fitted with a Resource Selection Function (RSF) that contrasts environmental predictors at observed locations (e.g. GPS locations) with those at available locations using logistic regression, and estimates the exponential of the linear predictor (Boyce et al., 1999; Manly et al., 2002; Muff et al., 2020; Northrup et al., 2022). RSF estimates the relative density of observed locations in space depending on environmental predictors for individuals (Aarts et al., 2012; Fieberg et al., 2021; Johnson et al., 2013; Muff et al., 2020). When RSF is specified in line with the weighted distribution theory, i.e. when approximating an infinite number of available locations, RSF and IPP likelihood converge and regression coefficient are asymptotically equivalent (Aarts et al., 2012; Fieberg et al., 2021; Fithian & Hastie, 2013; D. I. Warton & Shepherd, 2010).

Count data can also be analysed using IPP via log-linear regression of environmental covariates and a Poisson draw for count data, hereafter Poisson GLM (Aarts et al., 2012). In this situation, the point pattern represents the density of individuals from the studied population (Matthiopoulos et al., 2022; Renner et al., 2015).

Under certain conditions, IPPs would reconcile the analytical ground of telemetry and count models (Aarts et al., 2012; D. Warton & Aarts, 2013). Specifying RSF and count Poisson GLM as the intensity function of an IPP gives a similar significance to the regression coefficients, also called habitat selection coefficients (Aarts et al., 2012; Fieberg et al., 2021). In this paper, we contributed to compare animal distribution methods that use telemetry and count data. The equivalence of habitat selection coefficients of weighted RSF for telemetry data and of Poisson GLM for count data open perspectives to develop an integrated model that combine both data sources simultaneously.

Integrating (or combining) multiple datasets is an intense research avenue of current statistical ecology (Kéry & Royle, 2020; Zipkin et al., 2021). The root principle of most integrated models is that multiple datasets described a common underlying ecological processes in which at least one parameter is shared between several datasets, which are linked via a joint likelihood to each single dataset analysis (Fletcher et al., 2019; Kéry & Royle, 2020). The two main advantages of integrated models are i) increased accuracy of parameter of interest, and ii) in some situations, the ability to estimate parameters that could not be inferred from each data source alone (Kéry & Royle, 2020; Zipkin et al., 2019). Methodological developments on data integration have made it possible to combine many different data types (see Kéry & Royle (2020) for a review). Recently, telemetry data have been integrated in spatial capture-recapture models to better inform activity centers process (Converse et al., 2022). Using count, presence, and/or absence data collected from multiple monitoring programs, have also been used

to fit integrated species distribution models estimating a joint abundance or distribution process ([Matthiopoulos et al., 2022](#); [Miller et al., 2019](#); [Strebel et al., 2022](#)).

Building on the work of Aarts et al. ([2012](#)) and Matthiopoulos et al. ([2022](#)) on the IPP formulation of RSF and count models, we created an integrated model that estimate habitat selection combining count and telemetry data. We are not aware of any attempt to formally integrate telemetry and count data to estimate habitat selection coefficient, and follow the recent call of Matthiopoulos et al. ([2022](#)) in their review that stress for a such attempt.

In this paper, we first detailed how RSF using telemetry data and Poisson GLM using count data estimate habitat selection coefficient when designed as the intensity function of an IPP, which underlined the possibilities of integrating telemetry and count data to infer the habitat selection process and ultimately the relative intensity of use of animals in a study area. Second, we described a formulation of our integrated model made of two main components, a Poisson GLM using count data, and a weighted RSF using telemetry data. Then, we assessed the performances of the integrated model using simulated data. In addition, we applied our integrated model to estimate habitat selection of Sandwich Terns (*Thalasseus sandvicensis*) in the Northwestern Mediterranean Sea. Finally, we discussed the conditions of application of our modelling approach while suggesting future extensions. We put emphasis on the opportunity of integrating telemetry data and population count data to infer habitat selection process and animal distribution.

- **Methods**

- **Inhomogeneous Poisson Point process as a unifying framework**

An Inhomogeneous Poisson Point process (IPP) is a model that describe the stochastic distribution of random points in space where the expected density of points is correlated to spatial covariates. Hence, an IPP reflects the intensity of a spatial point process over a defined study area S with an intensity function λ_s that returns the expected number of points at every infinitesimal location s in S . The expected intensity λ is written as a log-linear function of environmental predictors X :

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

{#eq-test}

where parameters to be estimated are β_0 is the density intercept and β 's are the regression coefficients of the environmental predictors X .

The spatial ecological processes describing habitat selection can be model via an IPP using telemetry or count data ([Aarts et al., 2012](#); [Johnson et al., 2013](#); [Renner et al., 2015](#)). RSF estimate the relative selection of habitat components analysing the spatial point pattern of telemetry locations ([Aarts et al., 2012](#); [Fieberg et al., 2021](#); [Muff et al., 2020](#); [Northrup et al., 2022](#)). Density surface modelling estimate abundance and density from count data or mark-recapture data (e.g. Spatial Distance Sampling, or Spatial Capture-Recapture ([Farr et al., 2020](#); [Kéry & Royle, 2020](#); [D. L. Miller et al., 2013](#))). Aarts et al. ([2012](#)) demonstrated the convergence of RSF models using telemetry data with Poisson GLM used to analyse count data meaning that regression coefficients (intercept removed) of both RSF and Poisson GLM describe the relative intensity of locations in space, and are habitat selection coefficients ([Aarts et al., 2012](#); [Fieberg et al., 2021](#)). Since both methods estimate habitat selection as

a linear combination of environmental (biotic or abiotic) predictors, one can compare the relative habitat selection strength estimated through the regression coefficients of covariates, the β 's.

- **Count data - Poisson generalized linear model**

To estimate habitat selection from count data, we fitted a generalized linear model with a Poisson link function over a discrete geographical space where the expected number of individuals in a grid-cell follow the λ_s intensity function of an IPP, named the Poisson GLM ([Aarts et al., 2012](#)). We considered count data collected over multiple grid-cells as a count dataset N . N_s represents the expected counts in grid-cell s and is estimated as a Poisson draw in λ_s the IPP intensity at s .

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

{#eq-two}

One can extend the Poisson GLM to account for imperfect detection via a hierarchical model formulation ([Royle & Nichols, 2003](#)). Let us name Y_s the observed counts at s . We estimated N_s the latent number of points in the grid-cell s through a binomial draw with individual detection probability p . Then, Y_s links with the IPP intensity through {#eq-one} and {#eq-two}.

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

{#eq-three}

where p can be grid-cell specific and estimated as a logistic regression of covariates (see Sandwich tern application for an example).

- **Telemetry data - Resource Selection Function**

To estimate habitat selection from telemetry data, Resource Selection Functions (RSFs) are classically modelled using a logistic regression to estimate the exponential of the linear predictor based on used-available locations ([Boyce et al., 1999](#); [Manly et al., 2002](#); [Muff et al., 2020](#)). Animal locations being coded as 1 and available locations as 0, let us name K the dataset that stores the use-available data. RSF convergence with IPP holds only if available locations are 'infinitely' weighted ([Aarts et al., 2012](#); [Fieberg et al., 2021](#); [Fithian & Hastie, 2013](#); [D. I. Warton & Shepherd, 2010](#)). In practice, the weight w must be at least equal to 1000 for locations coded 0, and w is 1 for locations coded 1 ([Fieberg et al., 2021](#); [Muff et al., 2020](#)). RSFs operate in continuous geographical space, and the K_s of each point s in S is modelled as a Poisson draw in the IPP intensity λ from equation (1).

$$K_s \sim \text{Binomial}(\lambda_s, w_s)$$

The regression coefficients of the RSF indicate the relative selection for covariate X . Overall, predictions of a RSF written as $\omega = \exp(\beta X)$ is equivalent to the relative (not absolute) intensity of an IPP without its intercept. Following recommendations of Muff et al. ([2020](#)), we accounted for variation in habitat selection among I individuals. We included an individual random effect on regression parameters β 's and on the intercept with

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

$$\beta_{0,i} \sim \text{Normal}(\beta_0, 1e6)$$

for $i \in [1, I]$

With β being the parameter of interest describing the relative habitat selection strength of covariate X for the population, τ represents the between-individuals variance, i.e. population heterogeneity regarding habitat selection of covariate X , and β_0 representing the population intercept that is biologically meaningless in the RSF.

- **Data integration - RSF-Poisson model**

Combining an RSF and a Poisson GLM, we built an integrated model that jointly estimates regression coefficients β 's whose reflect the relative selection intensity in the IPP. The integrated model keeps the structure of the Poisson GLM and of the RSF, hence including two IPP-like models with specific intercepts and shared habitat selection coefficients.

Poisson GLM submodel:

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

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

RSF submodel:

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

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

For details, see Rcodes. Note that intercepts β_0^c of the Poisson GLM submodel and β_0^r of the RSF submodel are different. While IPP intensities λ_{counts} and λ_{RSF} are not equal, β s reflecting the relative selection for habitat covariates are shared among the two submodels. Thus, we obtained maps of relative habitat selection intensity by representing $\omega = \exp(\beta X)$ for RSF, Poisson GLM, and integrated model (Figure 2A).

- **2 Applications**

- **2.1 Simulations**

To test the validity and robustness of our models, we fitted it to several simulations. First, we generated a discrete landscape of two random habitat types A and B equally available with low spatial autocorrelation on a study area of 1000 x 1000 cells. We simulated a population of $N = 1000$ individuals who moved in this landscape during 300 steps, and set the parameters to obtain a higher use of habitat type A than B . We simulated tracks based on a biased correlated random walk to mimic a central place forager behaviour, as observed in colonial seabirds (see Rcodes in supplementary materials, Benhamou (2014)). Due to central place forager constraints, residual landscape autocorrelation and stochasticity in movement, the observed preference of habitat type A over B was $\hat{\Delta} = \frac{\text{number of steps in habitat A}}{\text{number of steps in habitat B}} = 3.22$, with $sd(\hat{\Delta}) = 0.16$ between the 1000 individuals (see Supplementary material for details about the simplest simulations). $\hat{\Delta}$ approximates the relative selection of habitat type A over B (i.e., ratio of selection ratio) in a context of equal availability of both habitat types and low spatial correlation. Therefore, the regression coefficient of the habitat type covariate to be estimated was $\beta = -\log(\hat{\Delta}) = -1.17$. To be more realistic, we included individual heterogeneity in habitat selection in the main batch of simulations. We assigned to each individual n in N , $\Delta_n \sim \text{Normal}(\Delta, \tau)$, with $\tau = 0.5$ the individual heterogeneity in habitat selection.

To mimic tracking data, we randomly selected I individual among the 1000 to fit the RSF model described in the previous section. Then, to mimic the monitoring of count data, we simulated zigzag line transect length using `{dsims}` R package (Marshall, 2023). To build our grid for Poisson GLM analyses, we segmented the line transects in cells of the same size as habitat grid-cells. We summed the number of locations of the 1000 individuals that fell into the grid-cell to constitute our count dataset N . We did not account for any imperfect detection and considered $p = 1$ in the simulations. Based on these count data, we fitted Poisson GLM to estimate habitat selection of A over B.

We explored the effect of the quantity of each data source to be included in the model on the estimation of habitat selection coefficient β . Among the 1000 individuals, we kept subsets of $I = 2, 5, \text{ and } 20$ individuals (i.e. 0.2%, 0.5%, and 2% of the population) to fit RSFs without accounting for individual heterogeneity in habitat selection. We fitted Poisson GLMs on line transect length ranging from 2000 to 15000 cell-units sampling effort (i.e. sampling 0.2% to 1.5% of the study area). We fitted the integrated RSF-Poisson model using increasing amount of data from count and RSF data (Figure 1). We showed the results for a single random scenario of simulations to have the RSF, the Poisson GLM, and the integrated model fitted to the same datasets, hence enabling to observe the effect of data integration without the “blur” of multiple data simulations.

- **2.2 Sandwich tern case study**

We applied our modelling development to estimate at-sea habitat selection of Sandwich tern (*Thalasseus sandvicensis*) during the breeding period (June to August) in the Gulf of Lion, French Mediterranean Sea.

Twenty-two sandwich terns were equipped with GPS-GSM in Etang de Thau nesting colony (43°23'13"N; 3°37'17"E) near Montpellier (Southern France) in late May 2021 and 2022 (i.e. representing 0.9% of nesting population of 2269 individuals, GISOM breeding monitoring 2022). We kept one location every hour between June and August 2022 and removed all locations at land. Filtering our dataset, we obtained 8490 GPS locations from 16 individuals, other individuals being dead, out of the study area, or the tracking device stopped functioning. To model the RSF, we simulated 10 available points for each used location. We accounted for the central place forager behavior of terns during the breeding period by and drawing available locations with a negative exponential bivariate distribution centered on the breeding colony (see Appendix for Rcodes). We restricted our study area to the continental shelf of the Gulf of Lion considering seafloors above 200m depth. We divided our study area in 711 contiguous hexagonal grid-cells of 21.65 km² (Figure 2, and supplementary materials). We considered linear and quadratic effect of bathymetry, and distance to coastline as environmental covariates (zotero-4644?). We found low correlation between distance to coastline and bathymetry ($r = -0.51$). Covariates were scaled before the analysis. We fitted an RSF accounting for individual heterogeneity in Sandwich tern habitat selection.

Then for each individual i and each location s in the study area, RSF model was:

$$\text{logit}(\lambda_{i,s}^{RSF}) = \beta_{0,i} + \beta_{1,i}\text{Bathymetry}_s + \beta_{2,i}\text{Bathymetry}_s^2 + \beta_{3,i}\text{Distance to coastline}_s$$

with

$$\beta_{0,i} \sim \text{Normal}(\beta_0, 1e6)$$

$$\beta_{1,i} \sim \text{Normal}(\beta_1, \tau_1)$$

$$\beta_{2,i} \sim \text{Normal}(\beta_2, \tau_2)$$

$$\beta_{n,3} \sim \text{Normal}(\beta_3, \tau_3)$$

where $\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2, \tau_3$ are to be estimated.

For the count model, we benefited from a yearly large-scale line transects monitoring program targeting small pelagic fisheries in the Gulf of Lion in July but also seabirds (PELMED, Laran et al. (2017); see Supplementary materials for details about monitoring programs and datasets). We fitted a Poisson GLM using count data collected during 2017-2021 summers, hence considering $J = 5$ sampling occasions assuming no temporal variation in the expected number of individuals per grid-cell. We obtained 281 detections for a total count of 584 sandwich terns between 2017 and 2021. We included sampling effort calculated as the transect length (in km) prospected in each grid-cell s during each sampling occasion j as an observation covariate affecting detection probability $p_{s,j}$.

$$\text{logit}(p_{s,j}) = \rho_0 + \rho_1 \text{Effort}_{s,j}$$

We estimated the IPP intensity λ^{count} using the same environmental covariates as the RSF model and the Poisson GLM submodel was:

$$\log(\lambda_{\text{counts}_s}) = \beta_0 + \beta_1 \text{Bathymetry} + \beta_2 \text{Bathymetry}^2 + \beta_3 \text{Distance to coastline}$$

Finally, we implemented the integrated model joining the RSF and the Poisson GLM as described in the previous section *Integrated RSF-Poisson model*.

- **2.3 Bayesian implementation**

We fitted all models with NIMBLE R-package (de Valpine et al., 2017). We ran two MCMC chains with 110 000 iterations each, 10 000 iterations burnin, and we assessed convergence ensuring that Gelman test provide $\hat{R} < 1.1$ (Gelman et al., 2013).

We displayed the posterior distribution of β for the RSF, the Poisson GLM, and the integrated model in Figures 1 and 2B for the simulation case and empirical Sandwich tern case study respectively. We predicted the relative intensity of use plotting the linear predictor

$$\theta = \exp(\beta_1 \text{Bathymetry} + \beta_2 \text{Bathymetry}^2 + \beta_3 \text{Distance to coastline})$$

in each grid-cell of the study area, according to the estimates of RSF, Poisson GLM and the integrated model (Figure 2A). We assessed the precision associated with the prediction of habitat selection intensity θ calculating the coefficient of variation $CV(\theta) = \frac{sd(\theta)}{\bar{\theta}}$ in each grid-cell for each model (Figure 2A) and representing the distribution of the precision over all grid-cells (Figure 2C). For additional details about the Sandwich tern case study, refer to Supplementary materials.

- **3 Results**

- **3.1 Simulations**

RSF, Poisson GLM, and integrated models were able to estimate β , the coefficient of relative selection of habitat type A over B, with varying bias and precision depending on amount of data used to fit the models (Figures 1). Poisson GLM produced better precision and lower bias of β with increasing number of cell-units sampled along the line transect (Figure 1A). Similarly, the increasing number of individuals in the RSF also improved the precision and reduced the bias associated with the habitat selection coefficient (Figure 1B). Overall, integrated models combining RSF and Poisson GLM increased precision and reduced bias in the estimated habitat selection coefficients β compared to models using each dataset alone (Figures 1C and 1D). Integrating increasing count data to RSF fitted with 5 individuals significantly reduced bias and increased precision of the β estimate (Figure 1C). Similarly, integrating increasing number of individuals in a RSF to a Poisson GLM with a sampling effort of 5000 cell-unit line transect increased the precision and reduced the bias of β compared to the Poisson GLM fitted with 5000 cell-units in isolation (Figure 1D). However, integrated model with RSF fitted to 5 individuals resulted in more biased estimates when integrating with Poisson GLM with 5000 cell-units than with 2000 cell-units. Indeed, 2000 cell-units Poisson GLM in isolation overestimated regression coefficient β (Figure 1A), hence counterbalancing the underestimation of RSF fitted with 5 individuals (Figure 1B-C). As 5000 cell-units Poisson GLM and RSF fitted with 5 individuals both underestimated β , hence these two datasets did not counterbalance in the integrated model and produced more biased estimates.

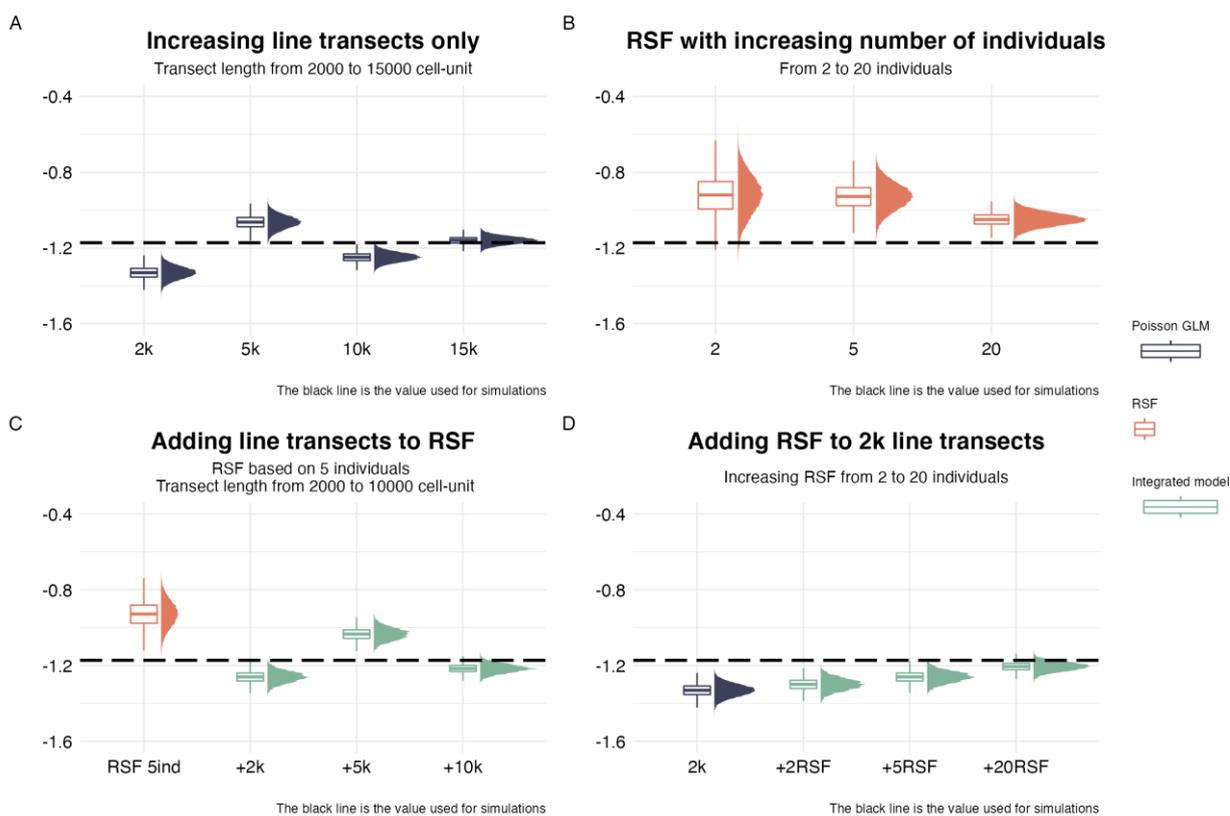
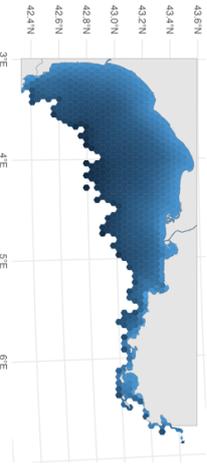


Figure 1: Estimation of habitat selection coefficient for habitat type from simulated data. Coefficients of habitat selection estimated with (A) Poisson GLM models with an increasing amount of data, (B) RSF with an increasing number of individuals, (C) with integrated RSF-Poisson GLM model, based on a five individuals for the RSF submodel and an increasing amount of line transect data (2000, 5000 and 15000) for the Poisson GLM submodel, and (D) with integrated RSF-Poisson GLM model, based on a 5000 cell-units line transects dataset for the Poisson GLM submodel and an increasing amount of individuals (2, 5, and 20) for the RSF submodel.

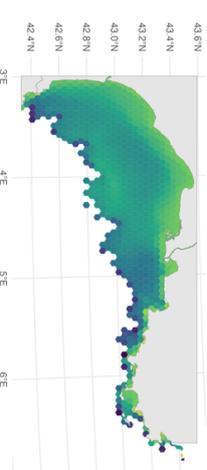
- **3.2 Sandwich terns in the French Mediterranean Sea**

RSF, Poisson GLM, and integrated models showed that Sandwich terns were more likely to select areas close than far from the coast (Figure 2B). The linear effect of bathymetry was more similar between Poisson GLM and the integrated model than with RSF. The RSF and integrated model estimated a significant quadratic effect of bathymetry while Poisson GLM estimated a quadratic effect of bathymetry with low precision (Figure 2A). RSF found a strong negative effect of distance to coastline, e.g. terns selected habitat close the shore, but with low precision (Figure 2B). In contrast, Poisson GLM and integrated model supported a weak negative effect with high precision (Figure 2B). Then, the integrated model benefit from both datasets to produce precise estimates of all covariates, closer to the estimates of the Poisson GLM than to those of the RSF, while Poisson GLM and RSF in isolation produced low precision for one covariate. In terms of relative intensity of spatial use (Figure 2A), the integrated model predicted a high use of coastal and mid-range depths seas, while Poisson GLM and RSF predicted a high use of shallower seas compared to mid-range depth and pelagic seas. When comparing coefficient of variation associated with relative space-use selection (Figure 2C), the integrated model displayed a better precision than Poisson GLM model, RSF resulting in the lower precision (i.e. higher standard deviation).

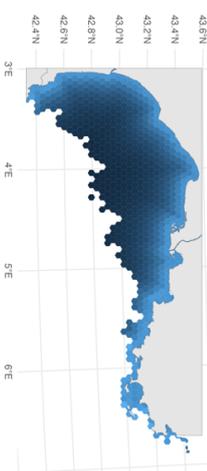
Relative space-use
Low High



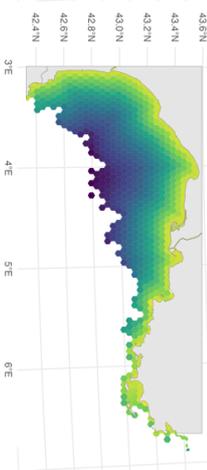
Coefficient of variation
Low High



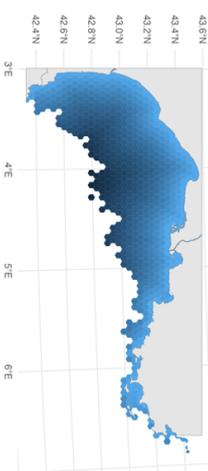
Relative space-use
Low High



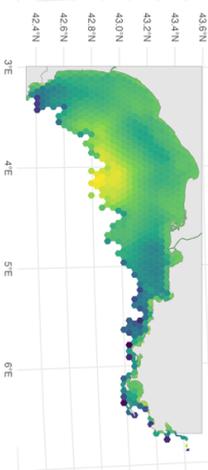
Coefficient of variation
Low High



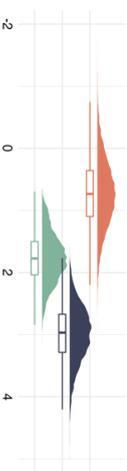
Relative space-use
Low High



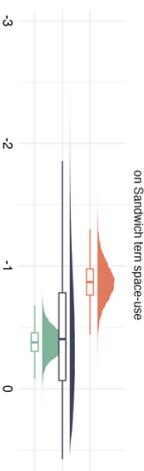
Coefficient of variation
Low High



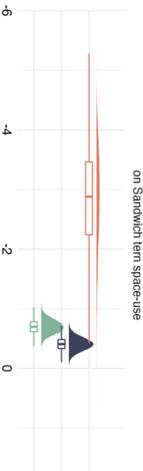
Linear bathymetry effect
on Sandwich tern space-use



Quadratic bathymetry effect
on Sandwich tern space-use



Distance to coast effect
on Sandwich tern space-use



Space-use precision
Predicted CV among grid-cells



RSF Integrated model Poisson GLM

Figure 2: Habitat selection of Sandwich tern (*Thalasseus sandvicencis*) habitat selection in the Gulf of Lion, French Mediterranean Sea, fitted with Resource Selection Function (RSF), Poisson GLM, and an integrated model combining telemetry and count data. We estimated habitat selection as a function of a linear and quadratic effect of bathymetry, and a linear effect of distance to coastline as environmental covariates. A. Spatial relative intensity of use for Sandwich tern and associated coefficient of variation predicted by RSF, Poisson GLM, and the integrated model. B. Effect of covariates estimated by RSF, Poisson GLM, and the integrated model. C. Distribution of coefficient of variation for Sandwich terns space-use predictions among the grid-cells of the study area for RSF, Poisson GLM, and the integrated model.

• 4 Discussion

We showcased the integration of telemetry and count data to estimate relative habitat selection. Our simulated data and the empirical Sandwich tern case study supported the benefit of the integrated model.

• 4.1 Simulations & Sandwich terns

The integrated model correctly retrieved the expected value of relative selection of habitat type A over B when the data are abundant and unbiased but also when the data quantity is limited or when individual heterogeneity in habitat selection occurred (Figure 1A-B). Simulations outlined integrated models displayed an increased or equivalent precision and a reduced bias compared to single data models (Figure 1C-D). Although we highlighted the ability of the integrated model to produce correct estimation of habitat selection, we did not explore every possible scenario. We support that investigating the relevant conditions of applications is beyond the scope of our article and is worth another dedicated study. Caution has to be taken when performing data integration as adding more data without explicitly accounting for a potential bias would lead to erroneous or unprecise ecological estimations compared to using an unbiased dataset alone ([Simmonds et al., 2020](#)). Integrating two datasets that are biased in the same direction (i.e. both under or overestimating regression coefficient) would not necessarily improve the accuracy of the estimates as shown when comparing the integrated models with 5-individuals RSF and -5000 cell-units Poisson GLM vs -2000 cell-units Poisson GLM (see *Results* section).

The Sandwich terns case study also exhibited that integrating telemetry and count data improved the precision estimation of relative habitat selection (Figure 2C). The integrated model benefit from both datasets and produce precise regression coefficient estimates for all covariates when Poisson GLM or RSF models did not (Figure 2). Quadratic effect of bathymetry are reported for similar seabirds species ([Pratte et al., 2021](#)), while a significant preference for shallow and close to colony waters is also documented on sandwich terns ([van Bemmelen et al., 2023](#)). Bathymetry is known to affect sandwich terns behavior and spatial use, however its role remains unclear ([Fijn et al., 2022](#)) and seems to depend on the local context and the spatial scale studied. Besides, most studies found a negative effect of increasing distance to coastline or distance from the colony for nesting sandwich terns ([Fauchald et al., 2021](#); [Fijn et al., 2022](#); [van Bemmelen et al., 2023](#)), which support the results of the three models. Furthermore, we suggested that data integration have potential conservation implications for the sandwich terns study as the integrated model predicted a more - although limited - pelagic habitat selection compared to the Poisson GLM and to RSF model (Figure 2A).

• 4.2 Condition of use and possible extensions

Our approach relies on the hypothesis that our animal tracks sample is representative of what occur at the population level regarding the environmental covariates. Scaling-up habitat selection from tracked individual to

the population level remain challenging ([Hooten et al., 2016](#); [Northrup et al., 2022](#); [Potts & Börger, 2023](#); [Signer et al., 2017](#)). For our sandwich terns case study, we must assume that relative habitat selection regarding bathymetry and distance to coastline span to all individuals nesting in the different colonies of the study area and remain consistent throughout the time period when the data were collected. Although we supported that such hypothesis was respected for sandwich tern in the Gulf of Lion, it is the duty of the researcher to verify the applicability of our modelling approach.

The convergence of Poisson GLM and RSF under the IPP framework allows estimating relative (and not absolute) habitat selection, the exponential predictor $\exp(\beta X)$ aligns without intercepts that are biologically specific to each data sources ([Northrup et al., 2022](#)). However, although RSF estimate relative habitat selection conditional on availability, the Poisson GLM produces an absolute estimation λ^{count} that reflects the expected number of individual per area unit and. One should obtain absolute estimation of density reconstructing the Equation (6) using shared slopes and specific Poisson GLM intercept. Ability to estimate absolute habitat selection from telemetry data while being integrated with count data constitutes a major asset of our modelling approach.

Besides, possibility to combine telemetry and count data should not return in a systematic application of integrated modelling when both datasets are available. One should wonder whether it is worth integrating multiple datasets. When one dataset is abundant and produced precise estimates, engaging in data integration might not significantly improve the outputs (Figure 1). However, if interested in a large-scale habitat selection assessment where by both telemetry data and population counts covered distinct spatial areas, integrated models can benefit from both data sources and capture complementary aspects of the targeted population (Figure 2, [Lauret et al. \(2022\)](#)). We acknowledged that further investigation of our modelling development would help to better understand to what extent and in which situations combining RSF and Poisson GLM is beneficial, hence providing guidance for ecological applications.

While we built our integrated model based on RSF and Poisson GLM, many possible extensions are worth considering for models using regression methods to estimate habitat selection. For telemetry data, SSFs and integrated SSFs (iSSFs) constitute widely used methods to estimate habitat selection, hence representing promising candidates for integration with population data ([Avqar et al., 2015](#); [McLoughlin et al., 2010](#)). It should be possible to integrate SSF with Poisson GLM to obtain joint inference of relative habitat selection coefficient when estimations converge between SSFs and RSFs fitted to the same data. However, in practice conditions to extrapolate SSF estimates to population-level predictions would be rarely fulfilled and require additional extension that are under current development ([Michelot et al., 2019](#)). When considering population-level data, species distribution models widely used presence-absence (or presence-only) data to estimate habitat selection (e.g. occupancy models, [Mackenzie et al. \(2002\)](#), [Tyre et al. \(2003\)](#)). A high resolution grid and the use of complementary loglog link function facilitate the convergence of logistic regression fitted to presence-absence data with the slope parameters estimated with an IPP fitted to count data ([Aarts et al., 2012](#)). Thus, occupancy models represent possible candidates to integrate with RSF to estimate relative habitat selection. Among species distribution models, one can consider to adapt the RSF-Poisson model to any niche model, presence-only model, occupancy models, distance sampling, or integrated species distribution models ([Farr et al., 2020](#); [Mackenzie et al., 2002](#); [Miller et al., 2019](#); [Royle, 2004](#); [Tyre et al., 2003](#)).

• 4.3 Conclusion

Overall, our integrated model reflected the complementary of telemetry and count data for habitat selection studies. For the recent years, animal tracking and count monitoring programs resulted in increasingly available

data although often targeting different ecological questions and being supported by different people and motivation ([Lindenmayer et al., 2022](#); [Nathan et al., 2022](#)). We demonstrated that, under some conditions, integration of telemetry and count data constitutes a relevant opportunity to investigate habitat selection, which can be a promising tool that would echo many ecological contexts.

- **References**

van Bemmelen, R. S. A., Leemans, J. J., Collier, M. P., Green, R. M. W., Middelveld, R. P., Thaxter, C. B., & Fijn, R. C. (2023). Avoidance of offshore wind farms by Sandwich Terns in the North Sea increases with turbine density. *Ornithological Applications*, duad055. <https://doi.org/10.1093/ornithapp/duad055>

Fradin, G., & Chamaillé-Jammes, S. (2023). Hogs sleep like logs: Wild boars reduce the risk of anthropic disturbance by adjusting where they rest. *Ecology and Evolution*, 13(7), e10336. <https://doi.org/10.1002/ece3.10336>

Marshall, L. (2023). *Dsims: Distance Sampling Simulations*. Distance development.

Potts, J. R., & Börger, L. (2023). How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection. *Journal of Animal Ecology*, 92(1), 16–29. <https://doi.org/10.1111/1365-2656.13832>

Courbin, N., Dortel, E., Grémillet, D., Lebreton, J.-D., & Besnard, A. (2022). Note sur la démographie pour une aide à la gestion et à la conservation des populations d'Oiseaux marins nicheurs du littoral français métropolitain. *Naturae*, 16. <https://doi.org/10.5852/naturae2022a16>

Lindenmayer, D. B., Lavery, T., & Scheele, B. C. (2022). Why We Need to Invest in Large-Scale, Long-Term Monitoring Programs in Landscape Ecology and Conservation Biology. *Current Landscape Ecology Reports*. <https://doi.org/10.1007/s40823-022-00079-2>

Fijn, R., Thaxter, C., Aarts, G., Adema, J., Middelveld, R., & Van Bemmelen, R. (2022). Relative effects of static and dynamic abiotic conditions on foraging behaviour in breeding Sandwich terns. *Marine Ecology Progress Series*, 692, 137–150. <https://doi.org/10.3354/meps14076>

Nathan, R., Monk, C. T., Arlinghaus, R., Adam, T., Alós, J., Assaf, M., Baktoft, H., Beardsworth, C. E., Bertram, M. G., Bijleveld, A. I., Brodin, T., Brooks, J. L., Campos-Candela, A., Cooke, S. J., Gjelland, K. Ø., Gupte, P. R., Harel, R., Hellström, G., Jeltsch, F., ... Jarić, I. (2022). Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science*, 375(6582), eabg1780. <https://doi.org/10.1126/science.abq1780>

Strebel, N., Kéry, M., Guélat, J., & Sattler, T. (2022). Spatiotemporal modelling of abundance from multiple data sources in an integrated spatial distribution model. *Journal of Biogeography*, jbi.14335. <https://doi.org/10.1111/jbi.14335>

Chandler, R. B., Crawford, D. A., Garrison, E. P., Miller, K. V., & Cherry, M. J. (2022). Modeling abundance, distribution, movement and space use with camera and telemetry data. *Ecology*, 103(10), e3583. <https://doi.org/10.1002/ecy.3583>

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>

- 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>
- 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>
- Northrup, J. M., Vander Wal, E., Bonar, M., Fieberg, J., Laforge, M. P., Leclerc, M., Prokopenko, C. M., & Gerber, B. D. (2022). Conceptual and methodological advances in habitat-selection modeling: Guidelines for ecology and evolution. *Ecological Applications*, 32(1), e02470. <https://doi.org/10.1002/eap.2470>
- Fauchald, P., Tarroux, A., Amélineau, F., Bråthen, V., Descamps, S., Ekker, M., Helgason, H., Johansen, M., Merkel, B., Moe, B., Åström, J., Anker-Nilssen, T., Bjørnstad, O., Chastel, O., Christensen-Dalsgaard, S., Danielsen, J., Daunt, F., Dehnhard, N., Erikstad, K., ... Strøm, H. (2021). Year-round distribution of Northeast Atlantic seabird populations: Applications for population management and marine spatial planning. *Marine Ecology Progress Series*, 676, 255–276. <https://doi.org/10.3354/meps13854>
- Robira, B., Benhamou, S., Masi, S., Llaurens, V., & Riotte-Lambert, L. (2021). Foraging efficiency in temporally predictable environments: Is a long-term temporal memory really advantageous? *Royal Society Open Science*, 8(9), 210809. <https://doi.org/10.1098/rsos.210809>
- Pratte, I., Ronconi, R., Craik, S., & McKnight, J. (2021). Spatial ecology of endangered roseate terns and foraging habitat suitability around a colony in the western North Atlantic. *Endangered Species Research*, 44, 339–350. <https://doi.org/10.3354/esr01108>
- 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>
- Gamelon, M., Nater, C. R., Baubet, É., Besnard, A., Touzot, L., Gaillard, J.-M., Lebreton, J.-D., & Gimenez, O. (2021). Efficient use of harvest data: A size-class-structured integrated population model for exploited populations. *Ecography*, 44(9), 1296–1310. <https://doi.org/10.1111/ecoq.05738>
- Zipkin, E. F., Zylstra, E. R., Wright, A. D., Saunders, S. P., Finley, A. O., Dietze, M. C., Itter, S., & Tingley, M. W. (2021). Addressing data integration challenges to link ecological processes across scales. *Frontiers in Ecology and the Environment*, 9.
- 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>
- Simmonds, E. G., Jarvis, S. G., Henrys, P. A., Isaac, N. J. B., & O’Hara, R. B. (2020). Is more data always better? A simulation study of benefits and limitations of integrated distribution models. *Ecography*, ecog.05146. <https://doi.org/10.1111/ecoq.05146>
- Muff, S., Signer, J., & Fieberg, J. (2020). Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation. *Journal of Animal Ecology*, 89(1), 80–92. <https://doi.org/10.1111/1365-2656.13087>

- 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>
- 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>
- 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>
- Miller, Pacifici, Sanderlin, & Reich. (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>
- Illian, J. B. (2019). Spatial and spatio-temporal point processes in ecological applications. In *Handbook of Environmental and Ecological Statistics*. Chapman and Hall/CRC.
- 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>
- 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>
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2), 403–413.
- Signer, J., Fieberg, J., & Avgar, T. (2017). Estimating utilization distributions from fitted step-selection functions. *Ecosphere*, 8(4), e01771. <https://doi.org/10.1002/ecs2.1771>
- Hooten, M. B., Buderman, F. E., Brost, B. M., Hanks, E. M., & Ivan, J. S. (2016). Hierarchical animal movement models for population-level inference. *Environmetrics*, 27(6), 322–333. <https://doi.org/10.1002/env.2402>
- 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.
- Renner, I. W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S. J., Popovic, G., & Warton, D. I. (2015). Point process models for presence-only analysis. *Methods in Ecology and Evolution*, 6(4), 366–379. <https://doi.org/10.1111/2041-210X.12352>
- Yamamoto, T., Watanuki, Y., Hazen, E. L., Nishizawa, B., Sasaki, H., & Takahashi, A. (2015). Statistical integration of tracking and vessel survey data to incorporate life history differences in habitat models. *Ecological Applications*, 25(8), 2394–2406. <https://doi.org/10.1890/15-0142.1>
- Thurfjell, H., Ciuti, S., & Boyce, M. S. (2014). Applications of step-selection functions in ecology and conservation. *Movement Ecology*, 2(1), 4. <https://doi.org/10.1186/2051-3933-2-4>

- Benhamou, S. (2014). Of scales and stationarity in animal movements. *Ecology Letters*, 17(3), 261–272. <https://doi.org/10.1111/ele.12225>
- Fithian, W., & Hastie, T. (2013). Finite-sample equivalence in statistical models for presence-only data. *The Annals of Applied Statistics*, 7(4), 1917–1939. <https://doi.org/10.1214/13-AOAS667>
- Miller, D. L., Burt, M. L., Rexstad, E. A., & Thomas, L. (2013). Spatial models for distance sampling data: Recent developments and future directions. *Methods in Ecology and Evolution*, 4(11), 1001–1010. <https://doi.org/10.1111/2041-210X.12105>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.
- Johnson, D. S., Hooten, M. B., & Kuhn, C. E. (2013). Estimating animal resource selection from telemetry data using point process models. *Journal of Animal Ecology*, 82(6), 1155–1164. <https://doi.org/10.1111/1365-2656.12087>
- Warton, D., & Aarts, G. (2013). Advancing our thinking in presence-only and used-available analysis. *Journal of Animal Ecology*, 82(6), 1125–1134. <https://doi.org/10.1111/1365-2656.12071>
- Aarts, G., Fieberg, J., & Matthiopoulos, J. (2012). Comparative interpretation of count, presence-absence and point methods for species distribution models: *Species Distribution as Spatial Point Process*. *Methods in Ecology and Evolution*, 3(1), 177–187. <https://doi.org/10.1111/j.2041-210X.2011.00141.x>
- 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>
- Warton, D. I., & Shepherd, L. C. (2010). Poisson Point Process Models Solve the "Pseudo-Absence Problem" for Presence-Only Data in Ecology. *The Annals of Applied Statistics*, 4(3), 1383–1402. <https://www.jstor.org/stable/29765559>
- Royle, J. A. (2004). N-Mixture Models for Estimating Population Size from Spatially Replicated Counts. *Biometrics*, 60(1), 108–115. <https://doi.org/10.1111/j.0006-341X.2004.00142.x>
- Royle, J. A., & Nichols, J. D. (2003). ESTIMATING ABUNDANCE FROM REPEATED PRESENCE–ABSENCE DATA OR POINT COUNTS. *Ecology*, 84(3), 777–790. [https://doi.org/10.1890/0012-9658\(2003\)084\[0777:EAFRPA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2)
- Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving Precision and Reducing Bias in Biological Surveys: Estimating False-Negative Error Rates. *Ecological Applications*, 13(6), 1790–1801. <https://doi.org/10.1890/02-5078>
- Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157(2), 281–300. [https://doi.org/10.1016/S0304-3800\(02\)00200-4](https://doi.org/10.1016/S0304-3800(02)00200-4)
- 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. *Ecology*, 83(8), 8.

Manly, B. F., McDonald, L., Thomas, D. L., McDonald, T. L., & Erickson, W. P. (2002). *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*. Springer Science & Business Media.

Boyce, M. S., McDonald, L. L., Boyce, M. S., & McDonald, L. L. (1999). Relating populations to habitats using resource selection functions. *Trends in Ecology & Evolution*, 14(7), 268–272. [https://doi.org/10.1016/S0169-5347\(99\)01593-1](https://doi.org/10.1016/S0169-5347(99)01593-1)
