|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Contents**  **1 Introduction** | | | | **3** |
| **2** | **Methods** | | | **6** |
|  | 2.1 Study area and data collection . . . . . . . . . . . . . . . . . . . . . . . . . . . | | | 6 |
|  | 2.1.1 Study area . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | | 6 |
|  | 2.1.2 Seabirds at-sea counts . . . . . . . . . . . . . . . . . . . . . . . . . . . | | | 7 |
|  | 2.1.3 Telemetry data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | | 8 |
|  | 2.1.4 Environmental data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | | 8 |
|  | 2.2 Statistical analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | | 9 |
|  | 2.2.1 Seabirds occupancy models with detection/non-detection data . . . . . | | | 9 |
|  | 2.2.2 Seabirds abundance modeling with count data . . . . . . . . . . . . . . | | | 10 |
|  | 2.2.3 Resource Selection Function space-use modeling based on Telemetry  GPS data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | | 11 |
|  | 2.2.4 Relative space use models integrating N-mixture and RSF models . . . | | | 12 |
|  | 2.2.5 Implementation and model fitting . . . . . . . . . . . . . . . . . . . . . | | | 13 |
|  | 2.2.6 Model selection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | | 14 |
|  | 2.3 Mapping seabirds’ vulnerability to wind farms . . . . . . . . . . . . . . . . . . | | | 15 |
|  | 2.3.1 Vulnerability indexes . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | | 15 |
|  | 2.3.2 Vulnerability maps . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | | 16 |
| **3** | **Results** | | | **16** |
|  | 3.1 Comparison of seabird occurrence and abundance modeling . . . . . . . . . . . | | | 17 |
|  | 3.2 Comparison and integration of movement data with count . . . . . . . . . . . | | | 20 |
|  | 3.3 Seabird spatial vulnerability to wind farm . . . . . . . . . . . . . . . . . . . . | | | 21 |
| **4** | **Discussion** | | | **22** |
|  | 4.1 Seabird vulnerability to wind farms . . . . . . . . . . . . . . . . . . . . . . . . | | | 22 |
|  | 4.2 Comparing detection / non-detection and count data . . . . . . . . . . . . . . | | | 24 |
|  | 4.3 Comparing and integrating telemetry and count data . . . . . . . . . . . . . . | | | 25 |
| **5** | **Conclusion** | | | **27** |
| **I** | **Appendix A** | | | **37** |
| **II Appendix B** | | |  | **38** |
|  | |

**Résumé** (cette page sera en 4e de couverture, donc à la fin du doc)

La mise en place de parcs éoliens en mer est en cours de développement dans le golfe du Lion, zone située au sud de la France, au nord-ouest de la Méditerranée. Cette région maritime revêt une grande importance pour les oiseaux marins, déjà confrontés à de multiples menaces. La préservation des espèces d’oiseaux marins méditerranéens est capitale en raison de leur statut de conservation et de leur endémisme. Il est donc impératif de procéder à une évaluation exhaustive de la vulnérabilité des oiseaux marins aux parcs éoliens dans cette zone afin d’en atténuer les impacts futurs.

Pour ce faire, nous avons estimé la distribution des espèces d’oiseaux marins dans le golfe du Lion en combinant différentes sources de données inexploitées de la région. Ces sources comprennent trois évaluations par bateau, une évaluation aérienne et des données de télémétrie. En combinant ces données à l’aide de modèles bayésiens intégrés, nous avons pu cartographier la distribution des oiseaux marins en ciblant un large éventail d’espèces. Nous avons comparé plusieurs approches de modélisation, en exploitant des données de détection/non-détection, des données de comptage et des données de télémétrie. Nous avons modélisé les différentes sources de données séparément et en combinaison pour fournir des recommandations en vue des prochains programmes de collecte. Nos résultats indiquent que bien que les données de détection/non-détection puissent fournir des recommandations précises pour la planification des parcs éoliens, les données de comptage demeurent l’approche à privilégier. Cependant, l’importance des données de télémétrie est croissante avec leur disponibilité accrue. L’utilisation conjointe de la télémétrie et des données de comptage représente une nouvelle approche prometteuse qui a permis d’obtenir des estimations plus précises que l’utilisation isolée de chaque source de données.

Notre analyse révèle une grande vulnérabilité des oiseaux marins dans les zones côtières et près de l’estuaire de la Camargue. Par conséquent, les zones actuellement envisagées pour les premiers parcs éoliens se trouvent dans des endroits très vulnérables.

# Summary

Offshore wind farms construction is planned in the Gulf of Lion, located in the South of France, in the northwestern Mediterranean. This area is highly important for seabirds,

which are already impacted by numerous threats. Protecting Mediterranean seabirds is crucial due to their conservation status and endemism. Therefore, conducting a comprehensive assessment of seabird vulnerability to wind farms in this area become essential for mitigating future impacts. To achieve this, we estimated seabird species distribution within the Gulf of Lion using an integrated approach that combines various unanalysed data sources from the area. These sources include three boat surveys, one aerial survey, and GPS telemetry data. By integrating these data using a model-based approach, we were able to estimate seabird distribution with great accurately and target a broad range of species. We employed several modeling approaches, utilizing detection/non-detection data, count data, and telemetry data. We modeled different data sources both separately and in combination. Our findings indicate that while detection/non-detection data can provide accurate recommendations for wind farm planning, count data remain the gold standard. However, telemetry data are becoming increasingly relevant as their availability grows. Combining telemetry with count data in this study was a novel and promising approach that provided more precise estimates than using each data source alone. Our analysis reveals high seabird vulnerability in coastal areas and near the Camargue estuarine. Therefore, the zones currently planned for the first wind farms fall in highly vulnerable locations.

# 1 Introduction

The Gulf of Lion is a hotspot of productivity in the Mediterranean Sea, crucial for wildlife like seabirds [1, 2, 3]. The high primary production is due to the vastness of the continental shelf, the freshwater and nutrient inputs from the Rhone River, and wind-driven upwelling [4, 5, 6, 7]. This productivity supports significant fish resources [8] that seabirds rely on for food [9]. Seabirds travel from colonies as far as 500 km away to forage in these waters [6, 9], highlighting the conservation importance of this area, even for seabirds living outside the Gulf of Lion.

The Mediterranean seabird community is characterized by a low species richness but a high degree of endemism [10]. Furthermore, some species and subspecies that use the Gulf of Lion are considered threatened either at the national or global level. For instance, this is the case of the Yelkouan shearwater (*Puffinus yelkouan*), the Balearic shearwater (*Puffinus mauretanicus*), and the European storm petrel (*Hydrobates pelagicus melitensis*). Mediterranean seabirds face multiple threats both on land and at sea. On land, predation by feral cats and black rats on eggs, chicks, and adults constitute the main threat [11, 12]. At sea, they are vulnerable to accidental bycatch by fisheries [9, 13], food shortages due to fishing activities [8, 14], various pollutants [15, 16, 17], and disturbances caused by offshore infrastructure [18]. The Gulf of Lion is a hub for commercial fishing [8], tourism, and recreational activities, making it an area where anthropogenic impacts accumulate and affect seabird population dynamics and viability [2].

In addition to existing threats, new human activities are planned in the Gulf of Lion, including offshore wind farm projects [19]. These projects are part of the international effort to develop renewable energy to mitigate climate change [20]. While offshore wind energy is expected to significantly contribute to renewable energy development, its impact on seabird populations is a major environmental concern [18, 21, 22, 23]. Offshore wind farms affect seabirds through direct collision mortality [21, 23] and through population displacement due to disturbance created by the wind turbines and associated activities [22, 23]. Furthermore,

wind farms alter the local ecosystem by changing wind and water circulation [24], transforming seabed communities [25], and restricting fishing activities [26]. These changes often make the area less attractive to seabirds, but in some cases, resources can become more abundant than they were before construction, increasing the area attractiveness [27, 28]. However, this increase in wind farm use could provide new feeding opportunities for seabirds and in turn improve their population dynamics, it also increases the risk of collisions [28].

The best way to mitigate the future impact of wind farms on seabird populations in the Gulf of Lion is to integrate biodiversity consideration into project design alongside technical and financial elements. The primary focus should be to avoid important areas for seabirds [29]. This initial phase is indispensable and precedes all other measures aimed at minimizing wind farm impacts, or ultimately compensating for residual impacts. That is why a spatial assessment of seabirds’ vulnerability to these installations is essential. This requires understanding the spatial distribution of seabird species within the assessment area. However, a comprehensive assessment of seabird distribution at the scale of the Gulf of Lion has yet not been conducted. Previous studies have estimated seabird distribution in the French Mediterranean Exclusive Economic Zone (EEZ) [3, 30]. Still, these studies were either not conducted at the species level [3] or focused on only a few seabird species [30]. To be comprehensive, the assessment of species distribution should encompass a broader range of species than previous studies.

In the Gulf of Lion, several programs have collected data on seabirds from 2011 to 2023 [3, 31]. Most of the data consists of counts collected during boat or aerial surveys, each conducted with different protocols. These protocols vary, for example, in the number of observers and the counting platforms used. The probability of detecting species and individuals of a species at each site depends on the protocol used and differs for each survey [32, 33, 34]. Besides, for some species GPS-tracking data are now also available. GPS-tracking involves remotely recording bird locations using a small device worn by the bird that utilizes the Global Positioning System (GPS). This technology provides data on the bird’s longitude, latitude, speed, and direction. Current assessments rely on single dataset approaches, and datasets are analyzed separately even though they measure the same process. Taking only one dataset offer limited information on a given species. Furthermore, data sources build on different technologies give access to complementary information concerning the same ecological process. Combining all these data in a single analysis will allow us to have a comprehensive and accurate assessment of seabird distribution.

The most reliable method to combine several data sources is to develop so-called ”integrated species distribution models” [35, 36, 37]. Integrated species distribution models allow to take detection and protocol differences into account [35, 36, 37]. Integrated models are often hierarchical models [36], i.e. a sequence of related models ordered by their conditional probability structure [34]. In hierarchical models, there are one or more intermediate levels involving latent variables. In our case, the models have a component for the observations and a component for the ecological state of our system [36, 34]. The component for the observation is different for each data sources, while the component for the ecological state is the same across all data sources. The idea is that each observation protocol gives a different representation of the same reality.

Compared to models based on a single data source, integrated models present several major advantages, with the primary one being enhanced accuracy in estimating ecological parameters of interest [38, 34]. Sometimes, they enable to estimate additional parameters not attainable through separate analyses of each data source or to have a broadened spatial and temporal coverage of the studied population [39, 37].

Depending on the available data and the specific ecological question, different classes of species distribution models can be employed. Occupancy models [34, 32, 40] use detection/non-detection data to estimate occurrence probability of a species at a site on a defined area. When count data are available and the ecological focus is on the abundance of a species at each site, N-mixture models can be utilized [34, 33, 41]. To guide public policy or understand species distribution, occupancy models are more widely used than N-mixture models [42], mainly because detection/non-detection data are more frequently available, due to their lower cost. Count data contain more information than detection / non-detection data, thus one would expect the power to detect spatial changes to be stronger for N-mixture models that can provide more detailed information on spatial process than occupancy models. But, in some cases occupancy models might be sufficient and could even serve as a surrogate for abundance estimation [43], particularly for rare species with low density [44]. However, this is less likely to be the case for seabirds, that are neither rare or at-low density. Additionally, occupancy models are often easier to fit than

N-mixture models [44], as the fundamental assumptions of the latter are frequently not met. Few studies have compared these two methods directly [44]. In the context of wind farm planning, it would be interesting to know if the use of detection/non-detection data could be sufficient to determine the spatial vulnerability of seabirds to wind farm.

Apart from count data, spatial distribution can also be determined from positions acquired by GPS devices [45]. GPS-tracking data is traditionally used to study habitat selection through a used-availability design, which compares the proportion of observations in different habitat types to the availability of those habitats across the landscape [45, 46]. Although not its primary focus, habitat selection analysis can also provide insights into species distribution [45].

As the quantity and variety of data increase, there’s a growing trend of using diverse data types to model species distribution [37], but employing a model-based approach to combine disparate data is not yet common practice. Count and GPS data capture complementary aspects of animal space-use [1]. Count data provide broad-scale population-level information,

while GPS data offer detailed, fine-scale information for a few individuals. Several studies have attempted to combine information from GPS and count data to determine species distribution, particularly in seabird research [1, 47, 48]. However, none of these studies utilized a model-based approach. Lauret et al. (under review) [49] developed a method to combine telemetry and count data into a single model. They used a hierarchical model hierarchical, incorporating a component for count data using an N-mixture model and a component for movement data through a Resource Selection Function (RSF), the most commonly used method in habitat selection analysis [46]. In this study, we used this model to formally integrate count and telemetry data to estimate seabird distribution.

Several and distinct distribution maps describing fine patterns of risk for each seabird species are not easy to interpret for public policy makers, as they may largely differ between species and do not contain information on species-specific vulnerability to wind farm. The construction of vulnerability indexes is generally favoured by public policy decision-makers. Yet producing such a synthetic map is complex because the final indicator must account for both species distribution and the risks posed by future wind farm installations, which is strongly species-dependent [28, 21, 50]. To address species-specific differences, Furness et al. (2013) [23] and Bradbury et al. (2014) [50] developed vulnerability index for each species. This index considers the conservation status of the species and metrics determining collision and habitat loss risk [23]. The final vulnerability map is created by taking a weighted average of all species distribution maps, with each species weighted by its respective vulnerability index [48, 50]. This approach enables prioritizing species that are particularly vulnerable to wind farm development by assigning them greater importance in the final map.

The primary question of this study was: where should wind farms be placed to minimize harm to seabird populations? This led to the sub-question: which areas are most and least utilized by each seabird species? To answer these questions, we: (1) make the most of existing data to predict reliable distribution of seabirds in the Gulf of Lion and then (2) use these predictions to assess their spatial vulnerability to future wind farm installations. We modeled species distributions using various techniques and data types, including integrated occupancy and integrated N-mixture models. These models combined at-sea observations from four different programs. For species with available GPS-tracking data, we performed habitat selection analysis using Resource Selection Functions (RSF) to derive species distributions. Finally, we integrated count and GPS-tracking data into a single model combining N-mixture and RSF. We then compared the outcomes generated by each method to provide insights into

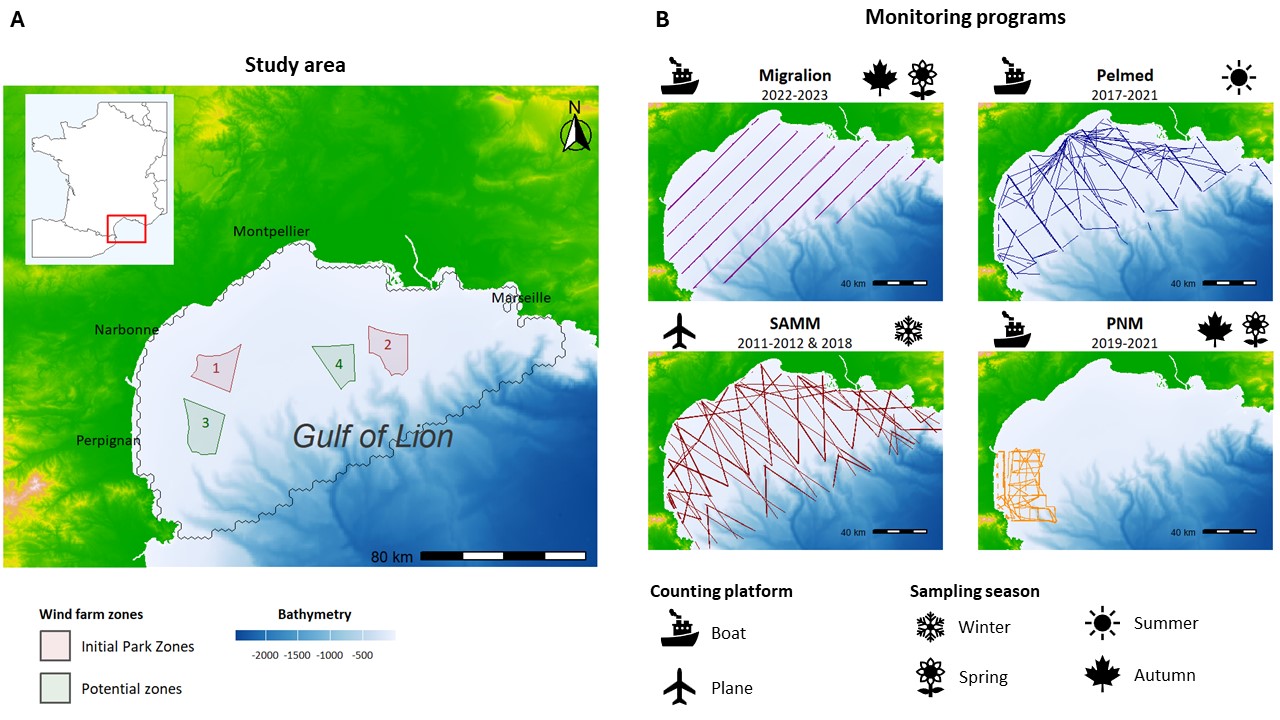
which type of data future data collection efforts should prioritize.

# 2 Methods

## 2.1 Study area and data collection

### 2.1.1 Study area

The Gulf of Lion is located in the north-western Mediterranean Sea (Fig.1). It is one of the most productive areas in the Mediterranean Sea, due to a combination of freshwater input by the Rhone River [4, 5] and small-scale upwellings created by winds [51]. The study area was determined based on suitability to receive wind farms, which is mainly driven by the bathymetry. In consequence, the study area coincides with the continental shelf. The total area considered for the study had a surface of approximately 17,300 km². Within this area, assessment units were defined by 1380 hexagonal cells with a size of 12,6 km². All the ecological parameters are estimated at the scale of the grid cells, thus each grid cell defines a ’site’. Grid resolution was chosen to have a ratio of approximately 1 cell with detection for 10 cells without any detection on average for each species. This choice was made based on simulations showing a loss in results accuracy below this threshold.



**Fig.1: Study area and transect lines - (A)** Location of the Gulf of Lion. Future wind farm spots are represented by red zones (1, 2) for the area that will receive the first farms and their extension and green zones (3, 4) for potential zones for future development. The black line delimits the study area and corresponds to the continental shelf. **(B)** The transects followed by each monitoring program to visually count seabirds. Migralion program was performed every year during April and September from 2022. PELMED (PELagic MEDiterranean) was conducted every year in July from 2017 to 2021. SAMM (Survol Aerien de la Megafaune Marine, Aerial Sensus of Marine Megafauna) was done during winter between 2011-2012 and 2018-2019. PNM (Parc Naturel Marin) designates data collected by the Marine Natural Parc of the Gulf of Lion, every year between 2019 and 2021, in late spring and autumn. It’s important to note that not every transect was sampled during each campaign.

### 2.1.2 Seabirds at-sea counts

Count data were collected from four different programs, as summarized in Figure 1. Observations were conducted using strip-transect methodology (Fig.1) [52, 53, 54], employing ships or aircraft as counting platforms. This methodology involves reporting and counting every seabird encountered within a band, typically spanning a few hundred meters[52].

SAMM (Survol Aerien de la Megafaune Marine, Aerial Census of Marine Megafauna) [30, 55], was the only aerial survey. It was conducted from November to February 2011-2012 and from January to February 2019, covering the entirety of the French Exclusive Economic Zone. Although this aerial survey also occurred during the summer of 2012, it was not replicated and thus not used in our study, as our models require temporally replicated data (see model description in 2.2 Statistical analyses for more details).

PELMED is an annual survey primarily conducted to assess small pelagic fish resources

[31], but collecting seabirds data with an ornithologist on board. This campaign occurred every summer from late June to early August since ?.

We also utilized data from the Marine National Park (PNM) of the Gulf of Lion, a large marine protected area. Surveys have been conducted twice a year since 2019, in autumn and late spring.

Data from Migralion was collected in 2022 and 2023, twice a year around April and September.

All these programs collected data at the species level, but in some cases, species identification was too difficult, leading to aggregation into the same category. Then, we labeled Balearic shearwater (*Puffinus mauretanicus*) and Yelkouan shearwater (*Puffinus yelkouan*), as small shearwaters, and we labeled Pomarine skua (*Stercorarius pomarinus*) and Arctic skua (*Stercorarius parasiticus*), as small skuas. The distribution of twelve different species were modeled (see species list in Table 3).

Seabirds’ distribution could change between seasons and some species are only present in the study area during a part of the year [30, 55]. For these reasons, our analysis was split into two periods: breeding season (from April to August) and non-breeding season (from September to March).

### 2.1.3 Telemetry data

GPS-tracking data within the Gulf of Lion was available for four species: Scopoli’s shearwater,

Yelkouan shearwater, Yellow-legged gull, and Sandwich tern. For the Migralion project in

2021 and 2022, 51 Scopoli’s shearwaters, 39 Yelkouan shearwaters, and 22 Sandwich terns were equipped with GPS at their nesting colonies on Riou Island near Marseille, in Port-Cros, and at the Etang de Thau near Montpellier, respectively, all located in the South of France. Additionally, 106 Yellow-legged gulls were equipped with GPS in 2021 in the Natural Park of Camargue.

We sampled recorded locations within the study area, and retained one location per hour to avoid spatial autocorrelation between successive points. These sampled locations were then assigned to two seasons, based on periods defined by count data.

### 2.1.4 Environmental data

We extracted environmental variables from the E.U. Copernicus Marine database [56, 57]

(https://data.marine.copernicus.eu) and MARSPEC database [58] (http://www.marspec.org).

A total of 8 geophysical and biotic covariates were included and are listed in Tab.1 (see also Appendix I Fig.6 for maps). Covariates choice was based on knowledge about seabirds ecology and frequently used covariates for seabird distribution modeling [1, 30, 59, 60, 61].

The dynamic variables (dynamic and SST covariates in Table 1) were obtained with monthly temporal resolution and averaged values were calculated from 2011 to 2023, covering the oldest to most recent seabird data available. We also computed the standard deviation for each covariate from the monthly temporal resolution and considered it as a measure of temporal variability of the covariate value. Sea Surface Temperature (SST) showed great seasonal

|  |  |  |
| --- | --- | --- |
| Variable | Description | Block |
| Bathymetry | Seabed depth  (in m) | Static covariate |
| Distance | Distance to the coast (in km) | Static covariate |
| Slope | Slope of the seabed  (in degree) | Static covariate |
| Seasonal SST | Sea Surface Temperature for each season (in °C) | SST covariates |
| SSS | Sea Surface Salinity | Dynamic covariate |
| SSH | Sea Surface Height above geoid (in m) | Dynamic covariate |
| Chlorophyll A | Chlorophyll-A concentration (in mg/m3) | Dynamic covariate |
| Current velocity | Sea water velocity  (in m/s) | Dynamic covariate |

**Table 1: Environmental covariates used in the study.**

variability, thus average values were calculated for each season. We extracted covariates at the spatial scale of each grid cell.

## 2.2 Statistical analysis

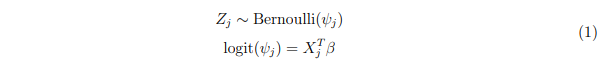
### 2.2.1 Seabirds occupancy models with detection/non-detection data

We first fitted occupancy models [32] to estimate seabird distribution in the Gulf of Lion for each species. The fundamental concept of this modeling approach is to consider the occurrence of a species as a Bernoulli random variable where the success is the species presence, and the success probability is the occurrence probability *ψ* [34, 32].

### Latent process

Let’s consider J sites (i.e. hexagonal cells of the grid), each surveyed K times. We denote *Zj* the true occupancy state at site j, where *j* ∈ 1*,...,J*. *Zj* = 0 if the species is absent and *Zj* = 1 if it is present. This latent occurrence variable is assumed to follow a Bernoulli process

[32]:

Here, *ψj* represents the probability of species occurrence at site j. The occurrence probability is modeled by logistic regression of site-specific environmental covariates *Xj*, with *β* being a vector of regression coefficients [34, 32].

### Observation process

The non-detection of a species at a site does not necessarily mean it is not using that site,

i.e. the observation process can yield false absence (*Y* = 0) while the species is occupying a site (*Z* = 1) (false negative errors). Species can be missed by the observers or absent at the time of survey while it is using this site [62] (an individual of a mobile species with a large home range may regularly use a site but be absent from it at the time of survey because it is visiting part of its home range outside of the sample unit). Not accounting for such imperfect detection when modeling ecological data could lead to wrong inferences [34, 63, 64, 65].

Let *Yj,k,r* be the observed detection (1) or non-detection (0) at site j during sampling occasion k for dataset r. The observation process is modeled separately for each data source to account for the differences between each protocol, which results in different sources of bias and detection probabilities. However, the observation process is conditional on the latent process described in Equation (1), which is shared for all data sources. For a data source r, the detection/non-detection data arises from a Bernoulli process conditional on the true latent occurrence process:



Where *pj,k,r* is the detection probability at site j for sampling occasion k and dataset r. In our model, the detection probability depends on one covariate: the length of the transect crossing the cells j at replicate k for dataset r. This covariate is a measure of the sampling effort.

To ensure that detection probability could be accurately identified, multiple sampling occasions are needed. Counts were repeated for each dataset over different years. We considered these counts as identical replicates despite potential inter-annual variability. This decision was made to obtain a static and averaged representation of seabird distribution, without accounting for year-to-year fluctuations. Additionally, seabirds are long-lived species with colonies locations that remain relatively fixed over time. Their population sizes are not expected to vary significantly between years, except for occasional migration variability. For all these reasons, this approach was deemed appropriate.

### 2.2.2 Seabirds abundance modeling with count data

As a second step, we inferred seabird at-sea abundance using N-mixture models [41, 33]. Similar to occupancy models, N-mixture models have a hierarchical structure where the observation is defined conditionally on the latent process describing the true abundance [41,

33].

### Latent process

Let *Nj* represent the true abundance at site j. This latent variable is assumed to follow a negative binomial distribution:

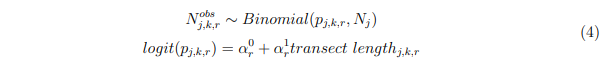
*Nj* ∼ *NegBin*(*λj, κ*)

(3) *log*(*λj*) = *XjT β*

Where *λj*, the mean of the distribution, was modeled as a function of site-specific covariates, and *κ* was the overdispersion parameter. A negative binomial distribution was preferred to a Poisson distribution to take overdispersion of the count data into account [66, 62], the negative binomial allowing the standard deviation and the mean to be different (the variance is *λ* + *λ*2*/κ*). Overdispersion is frequent with seabirds counts as individuals aggregate in great numbers at the same place for many species. This phenomenon occurs because seabirds are colonial birds that often displace and forage in groups.

### Observation process

We employed an integrated model with a specific detection process to accommodate differences between each dataset. The real number of individuals at a site was consistent across all datasets (latent abundance), but the number of observed individuals varied for each dataset. The observed number of individuals *Nj,k,robs* is assumed to follow a Binomial distribution conditional on the true number of individual *Nj*:

In other words, each individual at site j has a probability *pj,k,r* of being detected. Detection is considered independent for each individual. The detection probability *pj,k,r* is a function of the sampling effort, measured by the length of the transect crossing the cells as for occupancy modelling.

### 2.2.3 Resource Selection Function using telemetry data

Here, we used Resource Selection Function (RSF), a particular case of habitat selection function, where we fitted a logistic regression of used and available locations simultaneously on the entire dataset.

For each recorded location on the GPS track, we randomly selected 10 available points, or ”pseudo-absences,” uniformly distributed across the study area. We chose a uniform distribution because we considered the entire area to be equally accessible to seabirds, given its small size compared to the distances seabirds can travel [6, 67].

Then, our dataset take the form of a set of locations, divided between used and available locations. Let’s denote *Kj* the variable describing the location j. *Kj* = 1 for used locations and 0 for the available locations. *Kj* is considered to result from a Bernoulli draw:

*Kj* ∼ *Bernoulli*(*λj*)

(5) *logit*(*λj*) = beta0 + *XjT β*

Where *λj* is the probability for a location being used knowing its environmental covariates values (*λj* = *P*(*Kj* = 1|*Xj*)). *λj* is modeled as a logistic regression of location-specific covariates and regression coefficients *β*. Coefficients *β* represent the strength of the selection on the associated covariates.

To model potential population heterogeneity, we added a random effect on slope coefficients

*β*. Let’s I denote the total number of individuals, and take an individual *i* ∈ [1;*I*], we have:

*βi* ∼ *Normal*(*β,τ*) (6)

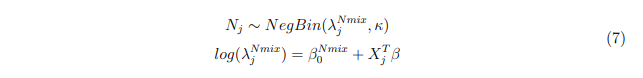
With *β* being the slope coefficient and *τ* a parameter quantifying the inter-individual variability regarding habitat selection on the associated covariate.

### 2.2.4 Relative space use models integrating N-mixture and RSF models

Lastly, we integrated GPS data with count data into a single model by combining an N-mixture model with a Resource Selection Function.

N-mixture and RSF both quantify habitat selection through the use of generalized linear regression on covariates X. Based on previous work [45, 46] showing that these models are similar to Inhomogeneous Poisson Point process (IPP), Lauret et al. (under review) [49] proposed to integrate these two data sources in a single model using IPP as a unifying framework. IPP describes the stochastic distribution of points in space when the expected density of points is correlated to environmental covariates, as in occupancy, N-mixture, and RSF models. An IPP reflects the intensity *λ* of a spatial point process over a defined study area. N-mixture and RSF do not estimate the same IPP absolute intensity *λ*. In RSF, the intercept beta0 has no biological meaning and only the relative habitat selection is estimated [68]. Relative IPP intensity *λ* being determined by slope coefficients *β*, coefficient estimates should be consistent between RSF and N-mixture. The integrated model keeps the hierarchical structure of the N-mixture and the logistic regression of the RSF and jointly estimates *β* parameters.

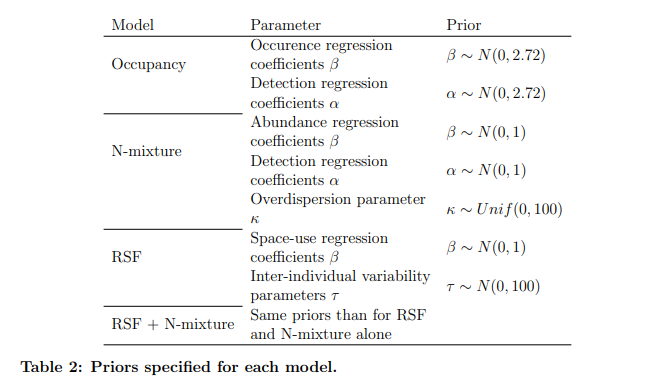
The model is divided into the N-mixture submodel:

And the RSF submodel:



### 2.2.5 Implementation and model fitting

All models were implemented in R (ref). Occupancy models were fitted with the package spOccupancy [69], allowing fast computation, and easy model selection. With this package, models are fit in a Bayesian framework. We ran 3 MCMC chains with 30,000 iterations each and a burn-in of 3,000 iterations. N-mixture, RSF, and the model integrating N-mixture and RSF were implemented with Nimble [70]. We ran 3 MCMC chains with 100,000 iterations each and a burn-in of 10,000 iterations. We used uninformative priors for all parameters (see Table 2). For all the models, convergence was checked by verifying that *R <*ˆ 1*.*1 [71] and by visual checks on trace plot.



Each method provides slightly different insights into seabird distribution. Occupancy models give the probability of species presence predicting the occurrence of the species for each grid cell. N-mixture models aim to model abundance predicting the number of individuals in each grid cell. However, in our case, model outputs interpretation is slightly different as the site closure assumption is not met. This assumption, crucial to the interpretation of occupancy and N-mixture models’ parameters, imply that the latent ecological state of a site (i.e. a grid cell) remains unchanged between the repeated visits [72]. For seabirds, this assumption is unrealistic, as they can easily travel from one site to another and occupied locations are used only temporarily by individuals [73]. The size of a sampling grid cell is much lower than their range of activity. In this situation, the occupancy estimator *ψj* represents the probability that grid cell j is used by the target species [74], being interpreted as space use by the seabird species. Similarly, N-mixture estimation does not capture the expected number of individuals at a site, but reflect the relative number of individuals using the site when it is surveyed [34, 75]. Consequently, occupancy and N-mixture models do not directly measure occurrence and abundance but rather provide a relative intensity of space use by seabirds. Besides, the intercept of the RSF is not meaningful (see section on RSF), and the output reflects relative habitat selection, indicating areas of higher or lower use. Overall, we interpreted our model outputs as relative intensity of space use, which is consistent across different data sources and modelling approaches. Thus, we will use the term ”relative space-use”, despite subtle differences in what each method specifically measures.

### 2.2.6 Model selection

The selection procedure was carried out exclusively using the occupancy model. For each species, the set of covariates used in the N-mixture, RSF, and the integrated N-mixture and

RSF models was the same as that obtained through model selection on the occupancy models. This ensured that the results produced by each model were comparable. It was also due to computation time that were too long to perform model selection with the other models during the time of the internship.

To have the most accurate prediction on seabird distribution as possible to inform wind farm policies, the selection procedure aimed to maximize the predictive power of our models. For this purpose, we evaluated model performance using two criteria: the Widely Applicable Information Criterion (WAIC) [76] and cross-validation [77]. In cross-validation, the model is fitted q times, each time excluding a fraction J/q of the data, with J being the total number of sites. Then, predictions are made on the q sites excluded from model fitting. Predictive performance is assessed by computing the difference between real site values and predicted values using the deviance [77]. When predictive scores were too close (*difference <* 3), we chose the model with fewer covariates.

When the correlation between two covariates was too high (*R*2 *>* 0*.*8), one of the covariates was excluded from the model selection procedure. Were excluded in priority covariates highly correlated with more than one variable, and covariates based on standard deviation before mean values.

Testing all possible models was computationally intensive, so we implemented a selection procedure in four steps:

* First, for each covariate, we tested whether adding a quadratic effect or applying a log transformation improved the predictive performance of a one-covariate model. Quadratic effects allow to take into account the existence of optimal value rather than linear effect for some covariates.
* Second, we divided the covariates into three thematic blocks (static, dynamic and SST covariates, see Table 1) and tested all possible combinations within each block, selecting the one with the best predictive performance. The three blocks were: the static covariates, the covariates linked to sea surface temperature, and the other dynamic covariates.

Then, with the best covariates of each of the three blocks resulting from step two, we tested whether combining all three blocks improved performance, or if two or even one block yielded the same predictive performance.

* In the final step, we evaluated whether incorporating a spatial autocorrelation structure would enhance predictive performance, as it is noted in many situations [78]. The spOccupancy package offers various spatial covariance models (spherical, exponential, and Gaussian) [69],

which we all tested.

Vulnerability index

Scientific name Common name

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Breeding season | Non-breeding season |
| *Calonectris diomedea* | Scopoli’s shearwater | 2.2 | 1.7 |
| *Chroicocephalus ridibundus* | Black-headed gull | 7.2 | 2.7 |
| *Fratercula arctica* | Atlantic puffin |  | 2.6 |
| *Hydrocoloeus minutus* | Little gull |  | 10 |
| *Hydrobates pelagicus melitensis* | European storm-petrel | 3.5 |  |
| *Larus melanocephalus* | Mediterranean gull | 6.3 | 7.3 |
| *Larus michahellis* | Yellow-legged gull | 10 | 7.6 |
| *Morus bassanus* | Northern gannet |  | 2.3 |
| *Puffinus mauretanicus &*  *Puffinus yelkouan* | Small shearwaters | 2.7 | 2.7 |
| *Stercorarius pomarinus &*  *Stercorarius parasiticus* | Small skuas |  | 3.3 |
| *Sterna hirundo* | Common tern | 2.9 |  |
| *Sterna sandvicensis* | Sandwich tern | 4.8 | 3.7 |

**Table 3: Species and their associated vulnerability index.** Blank space indicates no score for a species at a certain period due to its absence from the area.

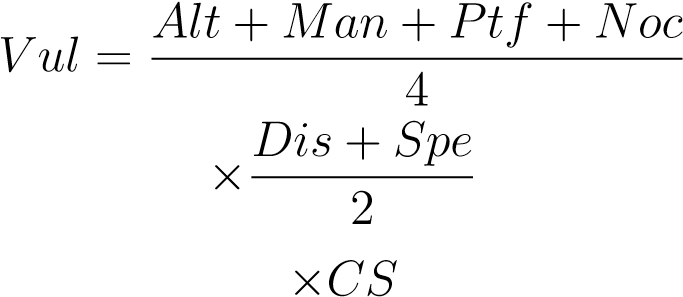
## 2.3 Mapping seabirds’ vulnerability to wind farms

### 2.3.1 Vulnerability indexes

Based on the results obtained with occupancy and N-mixture models, we constructed a synthetic index to evaluate species risk to offshore wind farms integrating three main aspects of seabird vulnerability [50, 23]:

1. **Conservation Status (CS):** Defined in écrire les auteurs [79], this score combines species status and representativeness of French populations. Species status is based on the worst IUCN Red List rankings or population trends. Representativeness is measured by the percentage of the global population in France and on the Mediterranean coast. Different representativeness values are defined for the breeding and non-breeding periods as seabird distribution can vary greatly between these periods [79, 3]. Representativeness and species are scored out of 10, and CS is the mean between these two scores.
2. **Risk of Collision** was evaluated with four criteria: the estimated percentage of time flying at blade height (Alt); flight maneuverability (Man); percentage of time spent flying (Ptf); and a nocturnal Activity Score (Noc).
3. **Risk of Displacement** was calculated using two criteria: the sensibility to disturbance by wind turbines, boats, and helicopters (Dis) and habitat specialization score (Spe).

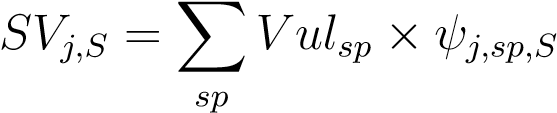
The vulnerability index was computed as follows [23]:



Vulnerability indexes for every species are indicated in Tab.3

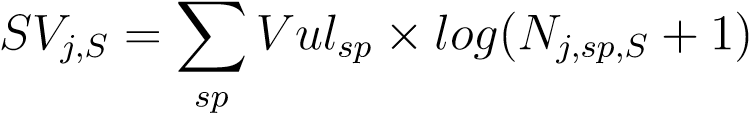
### 2.3.2 Vulnerability maps

To create a single vulnerability map, all the maps were combined, using the vulnerability index as a weighting factor. With occupancy maps, the seasonal seabird vulnerability *SV* at site *j* and season *S* was defined as:



where *ψj,sp,S* represents the space-use probability of site *j* by species *sp* during season *S*.

With abundance maps, the seasonal seabird vulnerability *SV* at site *j* and season *S* was defined following Bradbury et al. (2014) [50]:



These calculations were performed for the breeding (B) and non-breeding (NB) seasons.

The final vulnerability map was obtained by averaging the two seasonal maps at each site:

*SVtot* = mean(*SVB,SVNB*)

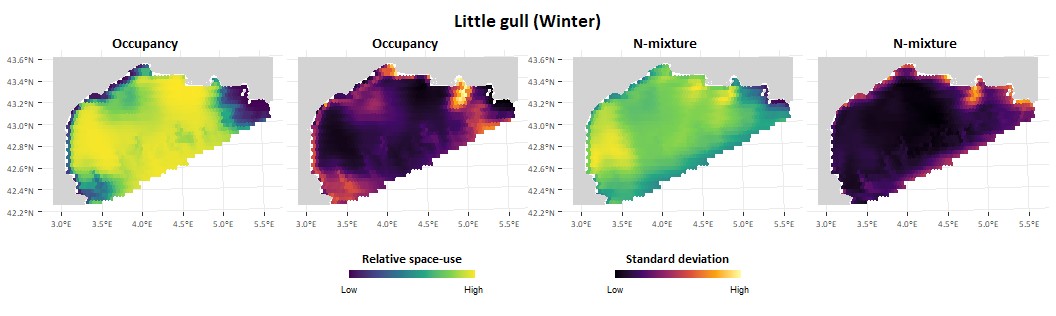
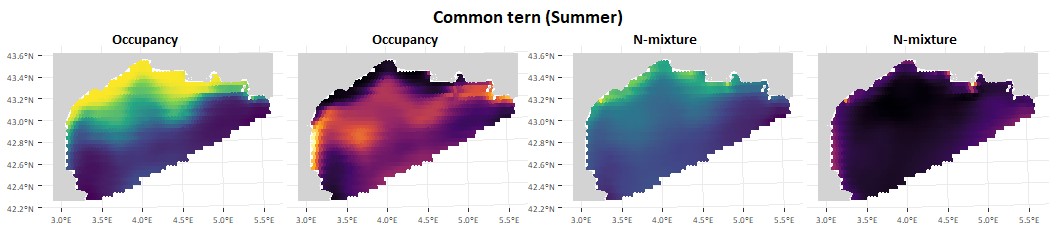
This approach ensured that both breeding and non-breeding seasons contribute to the overall assessment of seabird sensitivity. *SVtot* was then centered-reduced to obtain a relative vulnerability score over the entire study area. The mean of relative vulnerability was 0, positive values indicated areas with higher vulnerability and negative values indicated locations with lower risk.

# 3 Results

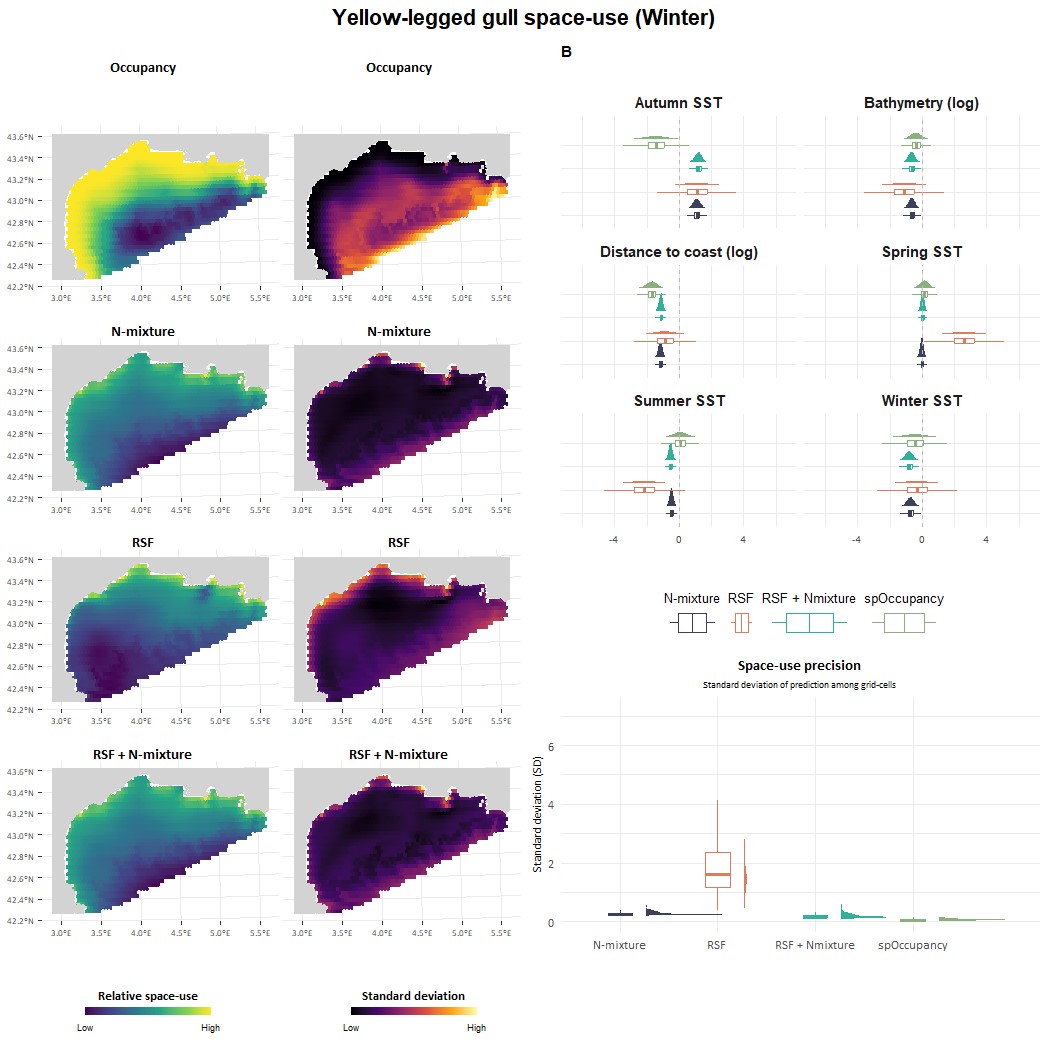
All models were able to effectively differentiate between less used and most used areas, showing good concordance with observable patterns in the data. The models selected after the model selection process are summarized in Appendix X, Table X. Models that included spatial autocorrelation resulted in low predictive performance and overfitting, none of the final model incorporate a spatial autocorrelation structure.

The results for occupancy and N-mixture models are presented for four representative species representative of the pattern variations: Common Tern and Little Gull in Table 2, and Yellow-legged Gull and Small Shearwaters in Table 3. Results for other species are provided in Appendix B and C.

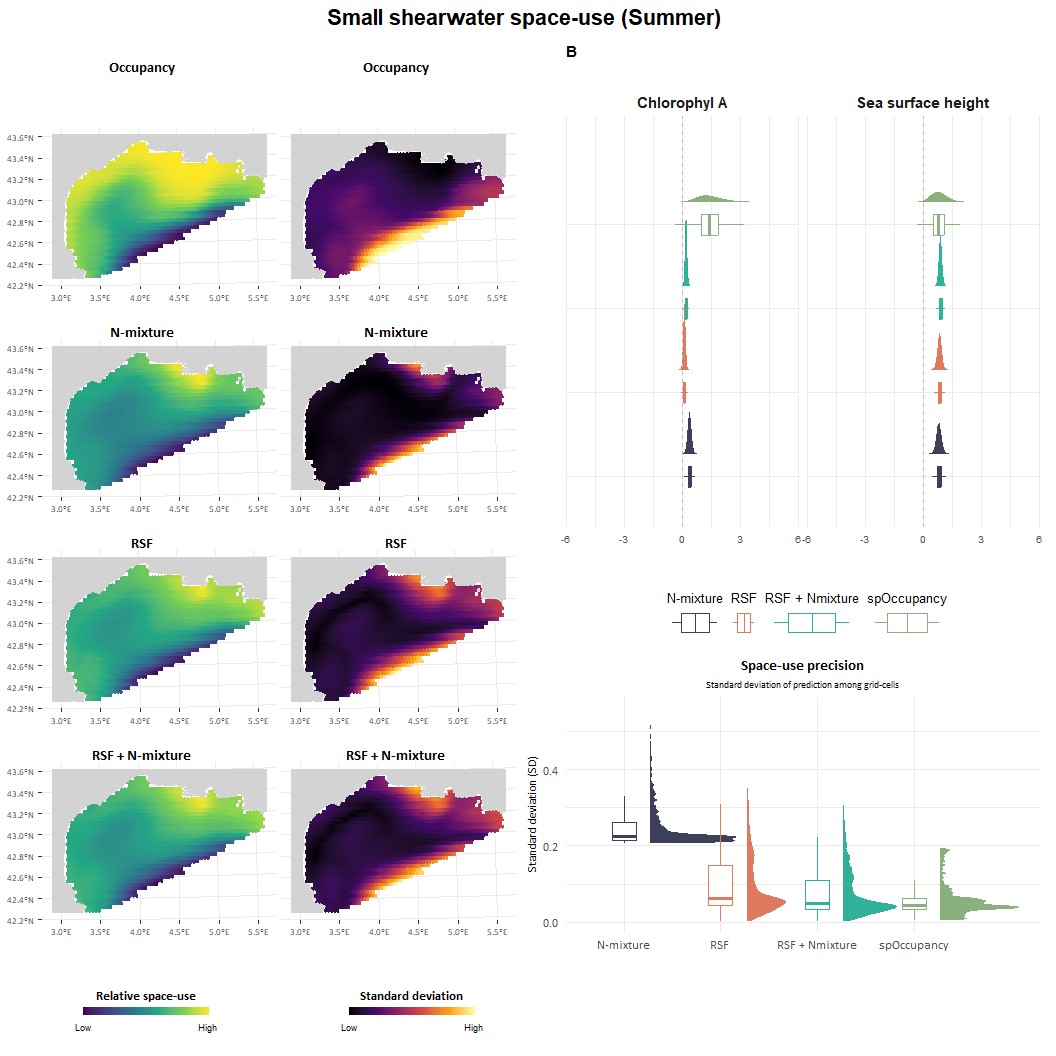
## 3.1 Comparison of seabird occurrence and abundance modeling



**Fig.2: Common tern (Sterna hirundo) and Little gull (Hydrocoloeutus minutus) distribution in the Gulf of Lion.** At-sea distribution estimated using an occupancy model with detection/non-detection data and an N-mixture model with count data. The figure shows the distribution maps and the associated standard deviation on the prediction.



**Fig.3: Yellow-legged gull (***Larus michaellis***) distribution in the Gulf of Lion.** Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining both movement and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty on the results, computed as the standard deviation on the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Prediction on the predicted space use, computed as the standard deviation of the grid cells.



**Fig.4: Small shearwaters (***Puffinus sp***) distribution in the Gulf of Lion.** Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining both movement and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty on the results, computed as the standard deviation on the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Prediction on the predicted space use, computed as the standard deviation of the grid cells.

**Common Tern.** Occupancy and N-mixture show similar distribution patterns. However, space-use estimated with N-mixture indicates a much more coastal distribution. While Common Terns can be found slightly offshore, the majority are concentrated along the coast and in the lagoons behind the coastline.

**Little Gull.** The distribution patterns from each method are quite similar, with highly used pelagic areas on the west and east sides of the Gulf. The main difference arises in the center of the Gulf and the slope area, where species space use is predicted to be contrasted between pelagic and coastal seas with occupancy, while space-use estimated from N-mixture is more homogeneous.

**Yellow-legged Gull.** Similar to the Common Tern, the distribution based on N-mixture shifts towards the coast compared to occupancy. Yellow-legged Gulls are the most abundant species in the study area and can be found almost everywhere. However, the N-mixture indicates that they remain very coastal, with most individuals staying near the coast.

**Small Shearwaters (Yelkouan and Balearic Shearwaters).** With both methods, space use intensity is predicted to be high near the Rhone mouth (the Camargue estuarine). Similarly, high space-use intensity is predicted to cover a wider area with the occupancy model than with the N-mixture model.

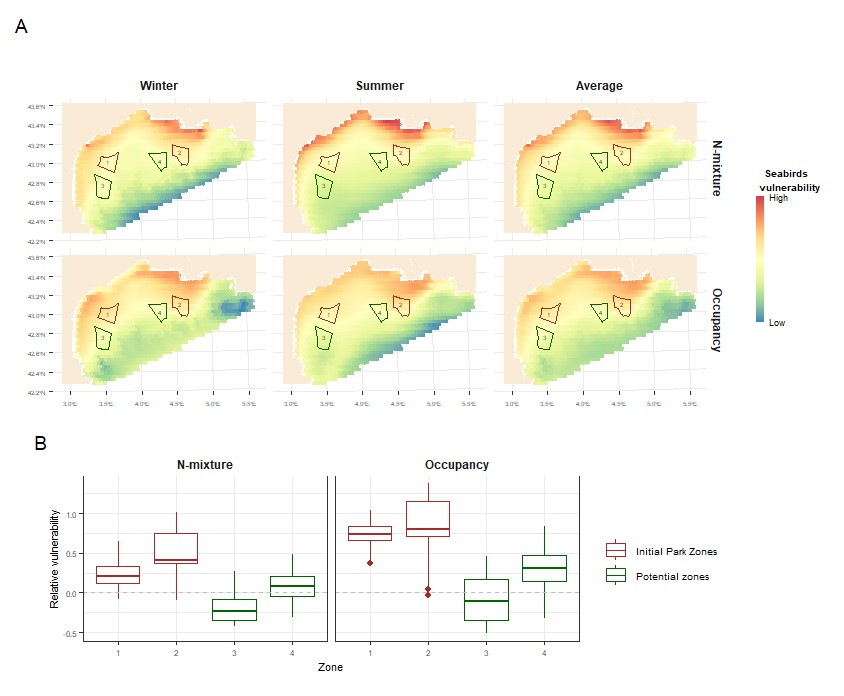
Differences frequently appear when comparing regression coefficient values between occupancy and N-mixture models (Figure 3, 4). Generally, the values remain close, but sometimes the slope direction are different. For example, for the Yellow-legged gull (Figure 3), the *β* coefficient associated with autumn SST shows a negative value in the occupancy model and a positive value in the N-mixture model.

## 3.2 Comparison and integration of movement data with count

The results obtained from RSF using telemetry data are consistent with those derived from N-mixture using count data (Figure 3, 4) (refer to the appendix for other species). For small shearwaters, the maps produced are similar. However, a notable difference is observed for the Yellow-legged Gull: the RSF map displays a West-East pattern with minimal predicted space use in the western part of the Gulf, whereas the N-mixture model predicts similar space use for both the western and eastern parts of the Gulf (Figure 3). In other words, higher space-use is predicted on the eastern side with RSF, corresponding to the area where Yellow-legged gulls

were equipped.

Integrating GPS-tracking data with count data yields contrasting results depending on the species. For small shearwaters (Fig. 4), coefficient estimates are similar between RSF and N-mixture models, and integration results in more precise estimates. Conversely, for Yellow-legged gulls, the coefficient estimates differ significantly, and RSF coefficients show considerable uncertainty. The final map and coefficient estimate for Yellow-legged gulls are identical to those obtained using the N-mixture model alone. For some species, such as the Sandwich tern in summer, the coefficient estimates from the integrated model are the same as those from the RSF alone, resulting in identical maps.



**Fig.5: Wind farm vulnerability maps.** (A) Maps illustrating the relative spatial vulnerability of seabirds to wind farms during two seasons and on average. The vulnerability maps were derived from species distribution maps using two types of models: Occupancy and N-mixture. Each row shows the results for each model. (B) Average vulnerability in areas proposed for wind farms, again using both the N-mixture and Occupancy models for species distribution. The grey line represents the mean vulnerability across the entire study area. Caution should be taken when comparing the maps and values between each method. The vulnerability index is relative, so only spatial differences and patterns can be compared.

## 3.3 Seabird spatial vulnerability to wind farm

The species with the highest vulnerability scores are the Little Gull, the Mediterranean Gull, and the Yellow-legged Gull in winter (non-breeding season), and the Yellow-legged Gull, Black-headed Gull, and Mediterranean Gull in summer (breeding season) (Table 3).

Seasonal variation in the spatial distribution of relative vulnerability is minimal, regardless of the methods used for species distribution (Fig. 5). The main difference appears in the western part of the Gulf, where there is an increased risk during winter due to the presence of the Little Gull during this period.

Across the entire area, higher vulnerability values are observed near the coast. On the maps based on abundance, vulnerability rapidly decreases with increasing distance from the coast, whereas it remains high over a larger area with occupancy modeling. A high vulnerability at the Rhone River mouth, near the Camargue region, is visible for both methods. Apart from the distance-to-coast effect, the maps obtained with the two modeling approaches seem to be in agreement.

The results on the maps also reflect when looking at the relative vulnerability of the different areas considered for wind farm locations (Fig. 5). With Nmixture modeling, vulnerability rapidly decreases with distance from the coast, so Zones 3 and 4 show a relative vulnerability equivalent to that of the whole study area. In contrast, Zones 1 and 2 have a much higher risk than the rest of the study area. With occupancy modelling, results are really similar except for zone 4, which has a vulnerability above the average vulnerability of the entire study area.

# 4 Discussion

## 4.1 Seabird vulnerability to wind farms

Despite strong variations in seabird distribution among species, averaging the spatial vulnerability across species reveals two distinct and consistent patterns. The first pattern is a strong inshore-offshore gradient, with very high vulnerability near the coast and lower vulnerability offshore (Figure 5?). The second pattern is the presence of a high vulnerability zone near the Camargue estuarine, corresponding to a gradient from high to low primary productivity.

The inshore-offshore gradient has also been reported in studies conducted in the United Kingdom [50], the east [80] and west [81] coasts of the United States, and Norway [48]. This gradient is likely partly explained by seabird behavior during the breeding period. During this time, seabirds act as central place foragers from their colonies, which are located along the coast. This explanation is supported by the stronger inshore-offshore gradient visible in summer (see Figure 5), and remind that wind farm could also indirectly affect offspring during this period. Another contributing factor is the presence of a high number of highly coastal species in our analysis, such as terns (Sandwich and Common terns) and gulls (Black-headed, Mediterranean, and Yellow-legged gulls) (see Figure 2 for distribution map).

The productivity gradient was not directly highlighted in other studies, although Goodale et al. (2014) [80] emphasized the importance of excluding seabird hotspots from wind farm infrastructure projects. Highly productive areas are known to be seabird hotspot, this includes upwelling regions and river mouths [82, 83]. In these areas, habitat loss would greatly impact seabirds’ populations as it corresponds to important area for them. Collision mortality is also expected to be higher due to high seabird density. Therefore, highly productive areas, such as the Camargue estuarine, should be avoided when considering wind farm projects. This gradient also explain why zone 2, one of the zones that should receive the first wind farm, exhibited a really high relative vulnerability (Figure 5).

Avoiding coastal and productive area is good general advice, supported by studies on seabird vulnerability to wind farm [80, 48], but this simplification could hide variation in spatial vulnerability between species. Even if moving wind farm away from coast decrease the impact on seabirds as a group, when looking at the different seabird guilds, the positive impact is more nuanced [80]. For instance, wind farm constructed far from shore would pose a greater risk on pelagic species [80].

In our study, pelagic species includes Yelkouan shearwater (*Puffinus yelkouan*), Balearic shearwater (*Puffinus mauretanicus*), Scopoli’s shearwater (*Calonectris diomeda*), Little gull (*Hydrocoloeus minutus*), European Storm petrel (*Hydrobates pelagicus melitensis*) and Atlantic puffin (*Fratercula arctica*). Shearwaters, storm-petrels and Atlantic puffin flight close to the water surface, thus spending a high proportions of their time below the rotor blade height [84, 85], thus having a low collision risk. In addition, shearwaters and storm-petrels exhibit low displacement risk [23, 50]. In comparison, Little gulls face a low risk of displacement but a high risk of collision [50]. Little gull is among the most abundant seabird species during winter and its distribution overlap with the future wind farm zones (Figure 2). The potential additional mortality caused by wind farms raises significant concern for this species.

The risk from the currently projected wind farm areas in zones 1 and 2 should be lower for coastal than for pelagic species. However, the risk persists, as many coastal species still use the projected wind farm areas, albeit to a lesser extent. Furthermore, post-construction studies in the North Sea show that some species do not avoid wind farms and even find them attractive [28, 27]. This attraction stems from opportunities to roost on above-water structures [28] and the higher abundance of prey. Vanermen et al. (2015) [28] reported significantly higher numbers of black-legged kittiwakes (not included in our analysis) foraging within wind farm areas compared to control areas. The turbine foundations offer hard substrates in typically soft-bottom areas [25], altering the seabed community and attracting more fish. This phenomenon is known as the ’reef effect’. Furthermore, benefits from the exclusion of trawling, allowing the soft-bottom community to recover [86], should also act positively on fish abundance. Whether these changes result in long-term beneficial effects remains unknown. However, they strongly suggest a potential increase in collision risk. Studies on Mediterranean seabirds and wind farms are scarce, and the behavioral responses of endemic species remain unexplored. However, based on the behavior of other gulls [28, 27], one could hypothesize that the wide-spread Yellow-legged gulls would find wind farms attractive, thus increasing their collision risk.

Contrary to birds attracted by wind farms, some species show strong avoidance behavior [27], resulting in a low collision risk. Therefore, species with a high displacement risk often exhibit a low collision risk and vice versa [81]. This reasoning led Furness et al. (2013) [23] to propose scoring displacement and collision risks independently. Recent studies have followed this method, computing these two scores separately and using the higher score for each species in wind farm risk assessments [48]. In our case, using two scores and taking the highest one did not significantly alter the relative vulnerability and the results.

Synthesizing seabird distribution with vulnerability scores involves substantial uncertainty

and biases [87]. Uncertainty arises from both the distribution maps and the vulnerability scores. Our integrated approach to combining multiple data sources should minimize uncertainty in the distribution maps [38, 34]. Though, predicted maps come with associated uncertainty not transferred to the spatial vulnerability. The second source of variability lies in the vulnerability scores [87]. The reliability of these scores depends on the methods used to determine the values of each criterion and the amount of supporting literature [87]. For instance, uncertainty about the time spent flying at turbine height remains very high when based on anecdotal observation but much lower when determined using direct studies, such as GPS or radar tracking. Therefore, it is important to remember that vulnerability scores and maps could evolve as more knowledge emerges.

In addition to wind farms, seabirds face numerous other threats at sea, such as episodic pollution incidents, fisheries bycatch, depletion of fish stocks, and disturbances from recreational activities. These threats can exert additive or compensatory effects when occurring in the same area [88]. Incorporating all these diverse threats into a cumulative impact assessment [89, 90] would improve recommendations for wind farm planning. The impact of wind farms may differ between two areas with similar seabird vulnerability if other threats already exist. In addition to local threats, climate change currently alters marine ecosystems, which in turn affects seabird distribution and abundance. Therefore, current assessments remain valid for only a limited time.

## 4.2 Comparing detection / non-detection and count data

Recent studies assessing seabird risk related to wind farm construction primarily focus on bird abundance or density [50, 81, 80], without using detection/non-detection data (or presence-absence data). However, detection/non-detection data are easier and cheaper to collect, and are often more abundant [91, 92]. Some studies suggest that occupancy estimates derived from detection/non-detection data sometimes serve as a surrogate for abundance [91], with little additional information gained from using abundance data [44]. This approach works particularly well for rare and elusive species [92, 44], and when budgets constrain research efforts [91]. In such cases, detection/non-detection data enable surveying more sites because detecting the presence of a species requires less time than counting all individuals, although this advantage might be less pronounced in open environments like the sea. Seabirds, however, are not particularly hard to detect and nor rare, a priori limiting the relevance of detection/non-detection data for these species.

For each species studied, we compared space-use estimates from two methods: N-mixture models using count data and occupancy models using detection / non-detection data. The estimates generally aligned, with the main difference occurring in areas of high space use. Occupancy models did not distinguish between areas with different levels of use when the species presence in the area is regular, whereas N-mixture models identified areas with both regular occurrences of few individuals and areas with high numbers of individuals. Maps of species space-use reflected differences that make sense considering the focus of each method. While detection/no-detection data used in occupancy models have the same weight in the model no matter how many individuals have been detected, count data used by Nmixture models allow to distinguish between levels of observed abundance, hence resulting in more detailed space-use intensity maps. For species-specific distribution, N-mixture models based on count data provided a more detailed view of seabird space use. Nevertheless, occupancy models based on detection/non-detection data still offered an accurate, albeit less detailed, picture of seabird distribution.

When synthesizing distribution maps for all species into a single map, both methods yield very similar results. The relative vulnerability of future wind farm development areas appears higher in the same zones for both methods. However, the areas with the highest vulnerability appear more spread out on the maps obtained using occupancy compared to those derived from N-mixture models. This difference stems directly from the variations observed in individual species maps. In this case, using detection/non-detection data rather than count data tends to provide more conservative recommendations for wind farm planning. However, this might not always occur, as detection/non-detection data could potentially obscure areas with very high risk.

The primary focus here involves assessing seabird vulnerability before wind farm construction. If the data also aim to evaluate the wind farm’s impact post-construction by comparing the state before and after, detection/non-detection data may not offer as much utility as count data. This happens because while abundance might decrease, species presence could remain unchanged (e.g. [93, 94]). Therefore, detection/non-detection data might be relevant for initial wind farm planning if more detailed data remain unavailable. However, count data should receive preference when collecting new data.

## 4.3 Comparing and integrating telemetry and count data

Using telemetry data from GPS devices and count data from at-sea observations to model species distribution produces similar and consistent results, especially for Scopoli and small shearwaters (Figure 4, Appendix B Figure 7). Both data types, however, are subjected to different bias [95, 96]. For telemetry data the main bias arises from the fact that only a few numbers of individuals, all from the same colony, are equipped with GPS [96].

When the sample size is too small, it limits the ability to make accurate population-level inferences, resulting in low statistical power [97] and high bias due to individual-specific traits [98, 99]. These traits include behavioral characteristics, such as more adventurous individuals being more likely to be captured and equipped [99, 100], as well as age and sex, with most equipped individuals being breeding adults [96], thus excluding non-breeding and younger birds. Although we took into account individual heterogeneity in the RSF model, limited sample size can affect the accuracy of inferences at the population level, as younger seabirds, for example, are likely to have different distributions and foraging behaviors compared to adults [101, 102].

For Sandwich terns, the distribution maps and relationships between distribution and environmental covariates were less consistent than for shearwaters (appendix B, Figure 8, 9). One reason for this could be the small sample size of only 22 individuals. Additionally, collar failures, individual deaths, and emigration outside the study area further reduced the available telemetry data for this species.

Furthermore, seabirds are generally equipped with GPS devices at a single breeding colony [96] (e.g. [102]). While this is the simplest method to equip seabirds, it can lead to highly biased estimates because space use may vary among different colonies of the same species [103]. This could explain the discrepancy between the distribution of Yellow-legged gulls modeled with RSF and N-mixture models. The distribution based on telemetry data showed a bias, with higher space use predicted on the east side of the Gulf, where the birds were equipped.

In contrast, for Scopoli’s shearwaters, there is only one large colony in the northwestern part of the Mediterranean Sea, so both birds observed from at-sea surveys and equipped with GPS devices came from this same colony, resulting in a strong match in the final distribution map (Figure 8?). This is also true for Yelkouan shearwaters (Figure X?). Although count data did not differentiate between Balearic and Yelkouan shearwaters, Yelkouan shearwaters are much more abundant than Balearic shearwaters.

Count data from at-sea surveys have the advantage of recording all individuals regardless of their behavioral traits, age, breeding status, or colony of origin. However, these surveys are conducted on a limited time scale, only during good weather conditions and daylight [3, 96]. In contrast, telemetry data provide information continuously, regardless of time or weather. Since distribution can depend on weather and time of day [104, 105], these factors can explain differences in distribution predicted by telemetry and count data. Carroll et al. (2019) [96] explored the differences between distributions obtained from GPS devices and boat-based surveys for two species and found a stronger match between the two data types when considering only daytime data for one of the species.

Other disadvantages of count data include limited taxonomic resolution for closely related species and observation errors. However, the latter issue can be addressed through modeling approaches [32, 33], as we did in our study. Despite these drawbacks, count data provide the advantage of indicating both the presence and absence of a species in an area (though non-detection does not always equate to absence), unlike telemetry data, which only show locations visited by sampled individuals and does not inform on their absence [45].

Telemetry data offer additional insights beyond species distribution. They can be used to study migration and dispersal events, estimate survival, reproduction, and recruitment rates, or understand the effects of human disturbances [100, 97, 95]. This can be particularly useful in the context of wind farm development where such data have been used to assess the behavioral response of new wind farms on seabirds [106, 107, 108], or to assess flight height [85], a key determinant of collision risk [23].

The choice between data types is often determined by financial resources. While telemetry data require fewer human resources compared to at-sea surveys, the cost of telemetry devices remains high, though it is declining [97]. The cost of GPS-tracking is not only financial but also ecological, as the fitness of tagged birds can be reduced, and include ethical consideration of animal wellbeing [109]. Conducting a comprehensive assessment using only telemetry data is challenging because typically only a few well-studied species are equipped [110], and small species, like Storm petrels, cannot be fitted with GPS tags. Equipping every species with GPS devices would demand significant financial resources, potentially reducing sample sizes and resulting in costly programs with limited statistical power [97]. Additionally, it would be crucial to equip birds from multiple colonies [96], which further increases both human and financial costs.

When we modeled species distribution combining telemetry data, count data in an integrated model (referred to hereafter as the integrated model), we expected the results from this integrated model to be less biased and more precise (ref Zipkin). We observed slightly improved precision when the regression coefficients were similar between the RSF and N-mixture models. When the coefficients differed, the values from the integrated model were intermediate, potentially indicating reduced bias. For Yellow-legged gulls, the maps from the integrated model closely matched those from count data alone, likely because the count data had abundant non-zero counts, suggesting it contained more information than the telemetry data. Conversely, the integrated model sometimes yielded results identical to those of the RSF suggesting that telemetry data had more weight for this species (Figure x) .

Our study had a substantial amount of count data from four datasets, which already allowed for accurate estimation of seabird distribution. Similarly, telemetry data, although available only for very abundant species, provided significant information. The integrated approach would be particularly beneficial for less abundant species, where count data mainly consist of zeros. Nevertheless, when both data types are available, using the integrated approach is recommended as it should help to overcome the different bias associated with each data type.

# 5 Conclusion

Seabird vulnerability in the Gulf of Lion increases near the coast and the Camargue region, where the Rhone and other factors drive a high primary productivity. Consequently, the proposed zones for the firsts wind farms fall in highly vulnerable areas for seabirds. Further improvement of the work done here would be to include other bird species with the additional data collected in 2024 (e.g. Black-legged kittiwake (*Rissa tridactyla*), Razorbill (*Alca torda*)). Future work should focus on count data modeled with N-mixture models, potentially benefiting from new model selection and incorporating a zero-inflated structure. Integrating telemetry data offers promise, especially if other species like the Mediterranean gull receive equipment. Adding other anthropogenic threats in the spatial assessment would be the logical pursuit of this work and would help identify areas where seabird conservation conflicts most with human activities.

# References

1. M. Louzao et al. “Combining vessel-based surveys and tracking data to identify key marine areas for seabirds”. *Marine Ecology Progress Series* (Sept. 28, 2009). doi:

[10.3354/meps08124.](https://doi.org/10.3354/meps08124)

1. Marta Coll et al. “The Mediterranean Sea under siege: spatial overlap between marine biodiversity, cumulative threats and marine reserves”. *Global Ecology and Biogeography* (2012). doi: [10.1111/j.1466-8238.2011.00697.x.](https://doi.org/10.1111/j.1466-8238.2011.00697.x)
2. Emeline Pettex et al. “Using large scale surveys to investigate seasonal variations in seabird distribution and abundance. Part I: The North Western Mediterranean Sea”. *Deep Sea Research Part II: Topical Studies in Oceanography*. Abundance, distribution and habitats of Atlantic and Mediterranean marine megafauna (July 1, 2017). doi:

[10.1016/j.dsr2.2016.11.008.](https://doi.org/10.1016/j.dsr2.2016.11.008)

1. Julie Gatti et al. “The Rhone river dilution zone present in the northeastern shelf of the Gulf of Lion in December 2003”. *Continental Shelf Research* (Oct. 1, 2006). doi: [10.1016/j.csr.2006.05.012.](https://doi.org/10.1016/j.csr.2006.05.012)
2. Marion Fraysse et al. “Intrusion of Rhone River diluted water into the Bay of Marseille:

Generation processes and impacts on ecosystem functioning”. *Journal of Geophysical Research: Oceans* (2014). doi: [10.1002/2014JC010022.](https://doi.org/10.1002/2014JC010022)

1. Josep Maria Gili et al. “Description of the ecology of the Gulf of Lions shelf and slope area and identification of the areas that may deserve to be protected” (2013). doi: [10.13039/501100000780.](https://doi.org/10.13039/501100000780)
2. B. Espinasse et al. “Defining zooplankton habitats in the Gulf of Lion (NW Mediterranean Sea) using size structure and environmental conditions”. *Marine Ecology Progress Series* (June 23, 2014). doi: [10.3354/meps10803.](https://doi.org/10.3354/meps10803)
3. D. Ba˘naru et al. “Trophic structure in the Gulf of Lions marine ecosystem (north-western

Mediterranean Sea) and fishing impacts”. *Journal of Marine Systems* (Feb. 1, 2013). doi: [10.1016/j.jmarsys.2012.09.010.](https://doi.org/10.1016/j.jmarsys.2012.09.010)

1. Carles Carboneras. “Seabirds in the Gulf of Lions shelf and slope area” (2013).
2. Richard Zotier, Vincent Bretagnolle, and Jean-Claude Thibault. “Biogeography of the marine birds of a confined sea, the Mediterranean”. *Journal of Biogeography* (1999). doi: [10.1046/j.1365-2699.1999.00260.x.](https://doi.org/10.1046/j.1365-2699.1999.00260.x)
3. E. Bonnaud et al. “Feeding Ecology of a Feral Cat Population on a Small Mediterranean Island”. *Journal of Mammalogy* (Aug. 20, 2007). doi: [10.1644/06-MAMM-A-](https://doi.org/10.1644/06-MAMM-A-031R2.1)

[031R2.1.](https://doi.org/10.1644/06-MAMM-A-031R2.1)

1. L. Ruffino et al. “Invasive rats and seabirds after 2,000 years of an unwanted coexistence on Mediterranean islands”. *Biological Invasions* (Aug. 1, 2009). doi: [10.1007/ s10530-008-9394-z.](https://doi.org/10.1007/s10530-008-9394-z)
2. K. Laneri et al. “Trawling regime influences longline seabird bycatch in the Mediterranean: new insights from a small-scale fishery”. *Marine Ecology Progress Series* (2010).
3. Philippe M. Cury et al. “Global Seabird Response to Forage Fish Depletion—One-Third for the Birds”. *Science* (Dec. 23, 2011). doi: [10.1126/science.1212928.](https://doi.org/10.1126/science.1212928)
4. Carolina Sanpera et al. “Audouin’s gull chicks as bioindicators of mercury pollution at different breeding locations in the western Mediterranean”. *Marine Pollution Bulletin* (June 1, 2007). doi: [10.1016/j.marpolbul.2007.01.016.](https://doi.org/10.1016/j.marpolbul.2007.01.016)
5. Marina Codina-Garc´ıa et al. “Plastic debris in Mediterranean seabirds”. *Marine Pollution Bulletin* (Dec. 15, 2013). doi: [10.1016/j.marpolbul.2013.10.002.](https://doi.org/10.1016/j.marpolbul.2013.10.002)
6. Chris Wilcox, Erik Van Sebille, and Britta Denise Hardesty. “Threat of plastic pollution to seabirds is global, pervasive, and increasing”. *Proceedings of the National Academy*

*of Sciences* (Sept. 22, 2015). doi: [10.1073/pnas.1502108112.](https://doi.org/10.1073/pnas.1502108112)

1. Stefan Garthe and Ommo Hu¨ppop. “Scaling possible adverse effects of marine wind farms on seabirds: developing and applying a vulnerability index”. *Journal of Applied Ecology* (2004). doi: [10.1111/j.0021-8901.2004.00918.x.](https://doi.org/10.1111/j.0021-8901.2004.00918.x)
2. Minist`ere de la transition ´ecologique. *Projets de parcs ´eoliens flottants et de leurs raccordements en Mer M´editerran´ee. Synth`ese du dossier de saisine de la Commission nationale du d´ebat public*. 2020.
3. *Paris agreement*. Dec. 12, 2015.
4. Sue King. “Seabird collision (2019). In: Perrow, M.R. (ed.) Wildlife and Wind Farms, Conflicts and Solutions. Vol. 3 Offshore: Potential effects, Chapter: 9.” Pelagic Publishing, Exeter, UK, Sept. 5, 2019. isbn: 978-1-78427-131-2.
5. Nicolas Vanermen and Eric W.M. Stienen. “Seabird displacement (2019). In: Perrow, M.R. (ed.) Wildlife and Wind Farms, Conflicts and Solutions. Vol. 3 Offshore: Potential effects, Chapter: 8.” Pelagic Publishing, Exeter, UK, Sept. 5, 2019. isbn: 978-1-78427131-2.
6. Robert W. Furness, Helen M. Wade, and Elizabeth A. Masden. “Assessing vulnerability of marine bird populations to offshore wind farms”. *Journal of Environmental Management* (Apr. 15, 2013). doi: [10.1016/j.jenvman.2013.01.025.](https://doi.org/10.1016/j.jenvman.2013.01.025)
7. Go¨ran Brostro¨m et al. “Atmosphere and ocean dynamics (2019). In: Perrow, M.R. (ed.) Wildlife and Wind Farms, Conflicts and Solutions. Vol. 3 Offshore: Potential effects, Chapter: 3.” Pelagic Publishing, Exeter, UK, Sept. 5, 2019. isbn: 978-1-78427-131-2.
8. J. Dannheim et al. “Seabed communities (2019). In: Perrow, M.R. (ed.) Wildlife and Wind Farms, Conflicts and Solutions. Vol. 3 Offshore: Potential effects, Chapter: 4.” Pelagic Publishing, Exeter, UK, Sept. 5, 2019. isbn: 978-1-78427-131-2.
9. Delphine A. Coates et al. “Short-term effects of fishery exclusion in offshore wind farms on macrofaunal communities in the Belgian part of the North Sea”. *Fisheries Research* (July 1, 2016). doi: [10.1016/j.fishres.2016.02.019.](https://doi.org/10.1016/j.fishres.2016.02.019)
10. Volker Dierschke, Robert W. Furness, and Stefan Garthe. “Seabirds and offshore wind farms in European waters: Avoidance and attraction”. *Biological Conservation* (Oct. 1, 2016). doi: [10.1016/j.biocon.2016.08.016.](https://doi.org/10.1016/j.biocon.2016.08.016)
11. Nicolas Vanermen et al. “Seabird avoidance and attraction at an offshore wind farm in the Belgian part of the North Sea”. *Hydrobiologia* (Sept. 1, 2015). doi: [10.1007/s10750014-2088-x.](https://doi.org/10.1007/s10750-014-2088-x)
12. Martin R. Perrow. “Wildlife and Wind Farms - Conflicts and Solutions, Volume 4. Offshore: Monitoring and Mitigation.” Pelagic Publishing, Exeter, UK, Sept. 5, 2019.
13. C. Lambert et al. “How does ocean seasonality drive habitat preferences of highly mobile top predators? Part I: The north-western Mediterranean Sea”. *Deep Sea Research Part II:*

*Topical Studies in Oceanography* (July 2017). doi: [10.1016/j.dsr2.2016.06.012.](https://doi.org/10.1016/j.dsr2.2016.06.012)

1. BOURDEIX Jean-Herv´e. “PELMED - PELAGIQUES MEDITERRANEE” (1985).´ doi:

[10.18142/19.](https://doi.org/10.18142/19)

1. Darryl I. MacKenzie et al. “Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One”. *Ecology* (2002). doi: [10.1890/0012-9658(2002) 083[2248:ESORWD]2.0.CO;2.](https://doi.org/10.1890/0012-9658(2002)083%5B2248:ESORWD%5D2.0.CO;2)
2. J. Andrew Royle. “N-Mixture Models for Estimating Population Size from Spatially Replicated Counts”. *Biometrics* (2004). doi: [10.1111/j.0006-341X.2004.](https://doi.org/10.1111/j.0006-341X.2004.00142.x)

[00142.x.](https://doi.org/10.1111/j.0006-341X.2004.00142.x)

1. “Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and

Species Richness in R and BUGS”. Ed. by Marc K´ery and J. Andrew Royle. Academic Press, Jan. 1, 2020. isbn: 978-0-12-823768-7. doi: [10.1016/B978-0-12-809585-](https://doi.org/10.1016/B978-0-12-809585-0.12001-0)

[0.12001-0.](https://doi.org/10.1016/B978-0-12-809585-0.12001-0)

1. David A. W. Miller et al. “The recent past and promising future for data integration methods to estimate species’ distributions”. *Methods in Ecology and Evolution* (2019). doi: [10.1111/2041-210X.13110.](https://doi.org/10.1111/2041-210X.13110)
2. Robert J. Fletcher Jr. et al. “A practical guide for combining data to model species distributions”. *Ecology* (2019). doi: [10.1002/ecy.2710.](https://doi.org/10.1002/ecy.2710)
3. Nick J. B. Isaac et al. “Data Integration for Large-Scale Models of Species Distributions”. *Trends in Ecology & Evolution* (Jan. 1, 2020). doi: [10.1016/j.tree.2019.08.006.](https://doi.org/10.1016/j.tree.2019.08.006)
4. Elise F. Zipkin, Brian D. Inouye, and Steven R. Beissinger. “Innovations in data integration for modeling populations”. *Ecology* (2019). doi: [10.1002/ecy.2713.](https://doi.org/10.1002/ecy.2713)
5. Matthew T. Farr et al. “Integrating distance sampling and presence-only data to estimate species abundance”. *Ecology* (2021). doi: [10.1002/ecy.3204.](https://doi.org/10.1002/ecy.3204)
6. Alan H. Welsh, David B. Lindenmayer, and Christine F. Donnelly. “Fitting and Interpreting Occupancy Models”. *PLOS ONE* (Jan. 10, 2013). doi: [10.1371/journal. pone.0052015.](https://doi.org/10.1371/journal.pone.0052015)
7. Lisa Madsen and J. Andrew Royle. “A review of N-mixture models”. *WIREs Computational Statistics* (2023). doi: [10.1002/wics.1625.](https://doi.org/10.1002/wics.1625)
8. Laurentiu Rozylowicz et al. “Occupancy and N-mixture modeling applications in ecology: A bibliometric analysis”. *Global Ecology and Conservation* (Apr. 1, 2024). doi:

[10.1016/j.gecco.2024.e02838.](https://doi.org/10.1016/j.gecco.2024.e02838)

1. D.I. MacKenzie and J.D. Nichols. “Occupancy as a surrogate for abundance estimation — Animal Biodiversity and Conservation” (Mar. 29, 2007).
2. Robert J. Ward et al. “Optimising monitoring efforts for secretive snakes: a comparison of occupancy and N-mixture models for assessment of population status”. *Scientific Reports* (Dec. 22, 2017). doi: [10.1038/s41598-017-18343-5.](https://doi.org/10.1038/s41598-017-18343-5)
3. Geert Aarts, John Fieberg, and Jason Matthiopoulos. “Comparative interpretation of count, presence–absence and point methods for species distribution models”. *Methods*

*in Ecology and Evolution* (2012). doi: [10.1111/j.2041-210X.2011.00141.x.](https://doi.org/10.1111/j.2041-210X.2011.00141.x)

1. Joseph M. Northrup et al. “Conceptual and methodological advances in habitat-selection modeling: guidelines for ecology and evolution”. *Ecological Applications* (2022). doi:

[10.1002/eap.2470.](https://doi.org/10.1002/eap.2470)

1. Clara P´eron et al. “Importance of coastal Marine Protected Areas for the conservation of pelagic seabirds: The case of *Vulnerable* yelkouan shearwaters in the Mediterranean Sea”. *Biological Conservation* (Dec. 1, 2013). doi: [10.1016/j.biocon.2013.09.006.](https://doi.org/10.1016/j.biocon.2013.09.006)
2. Per Fauchald et al. “Mapping seabird vulnerability to offshore wind farms in Norwegian waters”. *Frontiers in Marine Science* (Mar. 13, 2024). doi: [10.3389/fmars.2024.](https://doi.org/10.3389/fmars.2024.1335224)

[1335224.](https://doi.org/10.3389/fmars.2024.1335224)

1. Valentin Lauret, Nicolas Courbin, and Aur´elien Besnard. “Comparing and integrating movement and counts data to estimate animal habitat selection” ().
2. Gareth Bradbury et al. “Mapping Seabird Sensitivity to Offshore Wind Farms”. *PLOS ONE* (Sept. 11, 2014). doi: [10.1371/journal.pone.0106366.](https://doi.org/10.1371/journal.pone.0106366)
3. Claude Millot. “Circulation in the Western Mediterranean Sea”. *Journal of Marine Systems* (Apr. 1, 1999). doi: [10.1016/S0924-7963(98)00078-5.](https://doi.org/10.1016/S0924-7963(98)00078-5)
4. Gr´egoire Certain and Vincent Bretagnolle. “Monitoring seabirds population in marine ecosystem: The use of strip-transect aerial surveys”. *Remote Sensing of Environment*.

Earth Observations for Marine and Coastal Biodiversity and Ecosystems Special Issue (Aug. 15, 2008). doi: [10.1016/j.rse.2008.01.019.](https://doi.org/10.1016/j.rse.2008.01.019)

1. L. L. Eberhardt. “Transect Methods for Population Studies”. *The Journal of Wildlife Management* (1978). doi: [10.2307/3800685.](https://doi.org/10.2307/3800685)
2. Mark L. Tasker et al. “Counting Seabirds at Sea from Ships: A Review of Methods Employed and a Suggestion for a Standardized Approach”. *The Auk* (July 1, 1984). doi: [10.1093/auk/101.3.567.](https://doi.org/10.1093/auk/101.3.567)
3. Sophie Laran et al. “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*. Abundance, distribution and habitats of Atlantic and Mediterranean marine megafauna (July 1, 2017). doi:

[10.1016/j.dsr2.2016.12.011.](https://doi.org/10.1016/j.dsr2.2016.12.011)

1. Gianpiero Cossarini et al. “High-Resolution Reanalysis of the Mediterranean Sea Biogeochemistry (1999–2019)”. *Frontiers in Marine Science* (Nov. 29, 2021). doi:

[10.3389/fmars.2021.741486.](https://doi.org/10.3389/fmars.2021.741486)

1. Romain Escudier et al. “A High Resolution Reanalysis for the Mediterranean Sea”. *Frontiers in Earth Science* (Nov. 24, 2021). doi: [10.3389/feart.2021.702285.](https://doi.org/10.3389/feart.2021.702285)
2. Elizabeth J. Sbrocco and Paul H. Barber. “MARSPEC: ocean climate layers for marine spatial ecology”. *Ecology* (2013). doi: [10.1890/12-1358.1.](https://doi.org/10.1890/12-1358.1)
3. Juan Serratosa et al. “Environmental Drivers of Seabird At-Sea Distribution in the Eastern South Pacific Ocean: Assemblage Composition Across a Longitudinal Productivity Gradient”. *Frontiers in Marine Science* (Feb. 5, 2020). doi: [10.3389/fmars.](https://doi.org/10.3389/fmars.2019.00838)

[2019.00838.](https://doi.org/10.3389/fmars.2019.00838)

1. Blanca Sarzo et al. “Modelling seabirds biodiversity through Bayesian Spatial Beta regression models: A proxy to inform marine protected areas in the Mediterranean Sea”. *Marine Environmental Research* (Mar. 1, 2023). doi: [10.1016/j.marenvres.](https://doi.org/10.1016/j.marenvres.2022.105860)

[2022.105860.](https://doi.org/10.1016/j.marenvres.2022.105860)

1. Steffen Oppel et al. “Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds”. *Biological Conservation*. Seabirds and Marine

Protected Areas planning (Nov. 1, 2012). doi: [10.1016/j.biocon.2011.11.013.](https://doi.org/10.1016/j.biocon.2011.11.013)

1. Francisco V. D´enes, Lu´ıs F´abio Silveira, and Steven R. Beissinger. “Estimating abundance of unmarked animal populations: accounting for imperfect detection and other sources of zero inflation”. *Methods in Ecology and Evolution* (2015). doi: [10.1111/](https://doi.org/10.1111/2041-210X.12333)

[2041-210X.12333.](https://doi.org/10.1111/2041-210X.12333)

1. Gurutzeta Guillera-Arroita et al. “Ignoring Imperfect Detection in Biological Surveys Is

Dangerous: A Response to ‘Fitting and Interpreting Occupancy Models’”. *PLOS ONE* (July 30, 2014). doi: [10.1371/journal.pone.0099571.](https://doi.org/10.1371/journal.pone.0099571)

1. Jos´e J. Lahoz-Monfort, Gurutzeta Guillera-Arroita, and Brendan A. Wintle. “Imperfect detection impacts the performance of species distribution models”. *Global Ecology and Biogeography* (2014). doi: [10.1111/geb.12138.](https://doi.org/10.1111/geb.12138)
2. Marc K´ery. “Towards the modelling of true species distributions”. *Journal of Biogeography* (2011). doi: [10.1111/j.1365-2699.2011.02487.x.](https://doi.org/10.1111/j.1365-2699.2011.02487.x)
3. Alain F. Zuur et al. *Mixed Effects Models and Extensions in Ecology with R*. Ed. by Alain F. Zuur et al. New York, NY: Springer, 2009. isbn: 978-0-387-87458-6. doi:

[10.1007/978-0-387-87458-6\_1.](https://doi.org/10.1007/978-0-387-87458-6_1)

1. Maria P. Dias, Jos´e P. Granadeiro, and Paulo Catry. “Do Seabirds Differ from Other Migrants in Their Travel Arrangements? On Route Strategies of Cory’s Shearwater during Its Trans-Equatorial Journey”. *PLOS ONE* (Nov. 7, 2012). doi: [10.1371/ journal.pone.0049376.](https://doi.org/10.1371/journal.pone.0049376)
2. John Fieberg et al. “A ‘How to’ guide for interpreting parameters in habitat-selection analyses”. *Journal of Animal Ecology* (2021). doi: [10.1111/1365-2656.13441.](https://doi.org/10.1111/1365-2656.13441)
3. Jeffrey W. Doser et al. “spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models”. *Methods in Ecology and Evolution* (2022). doi: [10.1111/2041-210X.13897.](https://doi.org/10.1111/2041-210X.13897)
4. Perry de Valpine et al. “Programming with models: writing statistical algorithms for general model structures with NIMBLE”. *Journal of Computational and Graphical Statistics* (Apr. 3, 2017). doi: [10.1080/10618600.2016.1172487.](https://doi.org/10.1080/10618600.2016.1172487) arXiv: [1505. 05093[stat].](https://arxiv.org/abs/1505.05093%20%5Bstat%5D)
5. Andrew Gelman and Donald B. Rubin. “Inference from Iterative Simulation Using Multiple Sequences”. *Statistical Science* (Nov. 1, 1992). doi: [10.1214/ss/1177011136.](https://doi.org/10.1214/ss/1177011136)
6. Darryl I. MacKenzie et al., eds. *Occupancy Estimation and Modeling*. Boston: Academic Press, Jan. 1, 2018. isbn: 978-0-12-407197-1. doi: [10.1016/B978-0-12-407197-](https://doi.org/10.1016/B978-0-12-407197-1.00023-5)

[1.00023-5.](https://doi.org/10.1016/B978-0-12-407197-1.00023-5)

1. Eric Neilson et al. “Animal movement affects interpretation of occupancy models from camera-trap surveys of unmarked animals”. *Ecosphere* (Jan. 29, 2018). doi: [10.1002/ecs2.2092.](https://doi.org/10.1002/ecs2.2092)
2. William L. Kendall et al. “Relaxing the closure assumption in occupancy models: staggered arrival and departure times”. *Ecology* (2013). doi: [10.1890/12-1720.1.](https://doi.org/10.1890/12-1720.1)
3. Frank A. Fogarty and Erica Fleishman. “Bias in estimated breeding-bird abundance from closure-assumption violations”. *Ecological Indicators* (Nov. 1, 2021). doi: [10.](https://doi.org/10.1016/j.ecolind.2021.108170)

[1016/j.ecolind.2021.108170.](https://doi.org/10.1016/j.ecolind.2021.108170)

1. Sumio Watanabe. “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory” (2010).
2. M. B. Hooten and N. T. Hobbs. “A guide to Bayesian model selection for ecologists”. *Ecological Monographs* (2015). doi: [10.1890/14-0661.1.](https://doi.org/10.1890/14-0661.1)
3. J´eroˆme Gu´elat and Marc K´ery. “Effects of spatial autocorrelation and imperfect detection on species distribution models”. *Methods in Ecology and Evolution* (2018). doi: [10.](https://doi.org/10.1111/2041-210X.12983)

[1111/2041-210X.12983.](https://doi.org/10.1111/2041-210X.12983)

1. OFB and GISOM. *Identification et priorisation de la responsabilit´e de chaque sous r´egion marine pour les enjeux ornithologiques.* 2020.
2. M. Wing Goodale, Anita Milman, and Curtice R. Griffin. “Assessing the cumulative adverse effects of offshore wind energy development on seabird foraging guilds along the East Coast of the United States”. *Environmental Research Letters* (July 2019). doi:

[10.1088/1748-9326/ab205b.](https://doi.org/10.1088/1748-9326/ab205b)

1. Emma C. Kelsey et al. “Collision and displacement vulnerability to offshore wind energy infrastructure among marine birds of the Pacific Outer Continental Shelf”. *Journal of Environmental Management* (Dec. 1, 2018). doi: [10.1016/j.jenvman.2018.08.](https://doi.org/10.1016/j.jenvman.2018.08.051)

[051.](https://doi.org/10.1016/j.jenvman.2018.08.051)

1. Robert M. Suryan, Jarrod A. Santora, and William J. Sydeman. “New approach for using remotely sensed chlorophyll a to identify seabird hotspots”. *Marine Ecology Progress Series* (Apr. 11, 2012). doi: [10.3354/meps09597.](https://doi.org/10.3354/meps09597)
2. W. James Grecian et al. “Seabird diversity hotspot linked to ocean productivity in the Canary Current Large Marine Ecosystem”. *Biology Letters* (Aug. 2016). doi:

[10.1098/rsbl.2016.0024.](https://doi.org/10.1098/rsbl.2016.0024)

1. Alison Johnston et al. “Modelling flight heights of marine birds to more accurately assess collision risk with offshore wind turbines”. *Journal of Applied Ecology* (2014). doi: [10.1111/1365-2664.12191.](https://doi.org/10.1111/1365-2664.12191)
2. Nicolas Courbin et al. “Flight heights of Scopoli’s shearwaters Calonectris diomedea in the context of offshore wind farm developments” (May 2023). doi: [10.1101/2023.](https://doi.org/10.1101/2023.05.14.540698)

[05.14.540698.](https://doi.org/10.1101/2023.05.14.540698)

1. Gerard C. A. Duineveld, Magda J. N. Bergman, and Marc S. S. Lavaleye. “Effects of an area closed to fisheries on the composition of the benthic fauna in the southern North Sea”. *ICES Journal of Marine Science* (July 1, 2007). doi: [10.1093/icesjms/ fsm029.](https://doi.org/10.1093/icesjms/fsm029)
2. H. M. Wade et al. “Incorporating data uncertainty when estimating potential vulnerability of Scottish seabirds to marine renewable energy developments”. *Marine Policy* (Aug. 1, 2016). doi: [10.1016/j.marpol.2016.04.045.](https://doi.org/10.1016/j.marpol.2016.04.045)
3. Michael Elliott, Angel Borja, and Roland Cormier. “Activity-footprints, pressuresfootprints and effects-footprints – Walking the pathway to determining and managing human impacts in the sea”. *Marine Pollution Bulletin* (June 1, 2020). doi: [10.1016/](https://doi.org/10.1016/j.marpolbul.2020.111201)

[j.marpolbul.2020.111201.](https://doi.org/10.1016/j.marpolbul.2020.111201)

1. Gerjan J. Piet et al. “A roadmap towards quantitative cumulative impact assessments: Every step of the way”. *Science of The Total Environment* (Aug. 25, 2021). doi:

[10.1016/j.scitotenv.2021.146847.](https://doi.org/10.1016/j.scitotenv.2021.146847)

1. Benjamin S. Halpern and Rod Fujita. “Assumptions, challenges, and future directions in cumulative impact analysis”. *Ecosphere* (2013). doi: [10.1890/ES13-00181.1.](https://doi.org/10.1890/ES13-00181.1)
2. Liana N. Joseph et al. “Presence–Absence versus Abundance Data for Monitoring Threatened Species”. *Conservation Biology* (2006). doi: [10.1111/j.1523-1739. 2006.00529.x.](https://doi.org/10.1111/j.1523-1739.2006.00529.x)
3. Jacob F. Pollock. “Detecting Population Declines over Large Areas with PresenceAbsence, Time-to-Encounter, and Count Survey Methods”. *Conservation Biology* (2006). doi: [10.1111/j.1523-1739.2006.00342.x.](https://doi.org/10.1111/j.1523-1739.2006.00342.x)
4. Verena Peschko et al. “Effects of offshore windfarms on seabird abundance: Strong effects in spring and in the breeding season”. *Marine Environmental Research* (Dec. 1, 2020). doi: [10.1016/j.marenvres.2020.105157.](https://doi.org/10.1016/j.marenvres.2020.105157)
5. Nicolas Vanermen et al. “Assessing seabird displacement at offshore wind farms: power ranges of a monitoring and data handling protocol”. *Hydrobiologia* (Sept. 1, 2015). doi:

[10.1007/s10750-014-2156-2.](https://doi.org/10.1007/s10750-014-2156-2)

1. A. David M. Latham et al. “The GPS craze: six questions to address before deciding to deploy GPS technology on wildlife”. *New Zealand Journal of Ecology* (2015).
2. Matthew J. Carroll et al. “Matches and Mismatches Between Seabird Distributions Estimated From At-Sea Surveys and Concurrent Individual-Level Tracking”. *Frontiers in Ecology and Evolution* (Sept. 24, 2019). doi: [10.3389/fevo.2019.00333.](https://doi.org/10.3389/fevo.2019.00333)
3. Mark Hebblewhite and Daniel T. Haydon. “Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology”. *Philosophical Transactions*

*of the Royal Society B: Biological Sciences* (July 27, 2010). doi: [10.1098/rstb.](https://doi.org/10.1098/rstb.2010.0087)

[2010.0087.](https://doi.org/10.1098/rstb.2010.0087)

1. Peter A. Biro. “Are most samples of animals systematically biased? Consistent individual trait differences bias samples despite random sampling”. *Oecologia* (Feb. 1, 2013). doi:

[10.1007/s00442-012-2426-5.](https://doi.org/10.1007/s00442-012-2426-5)

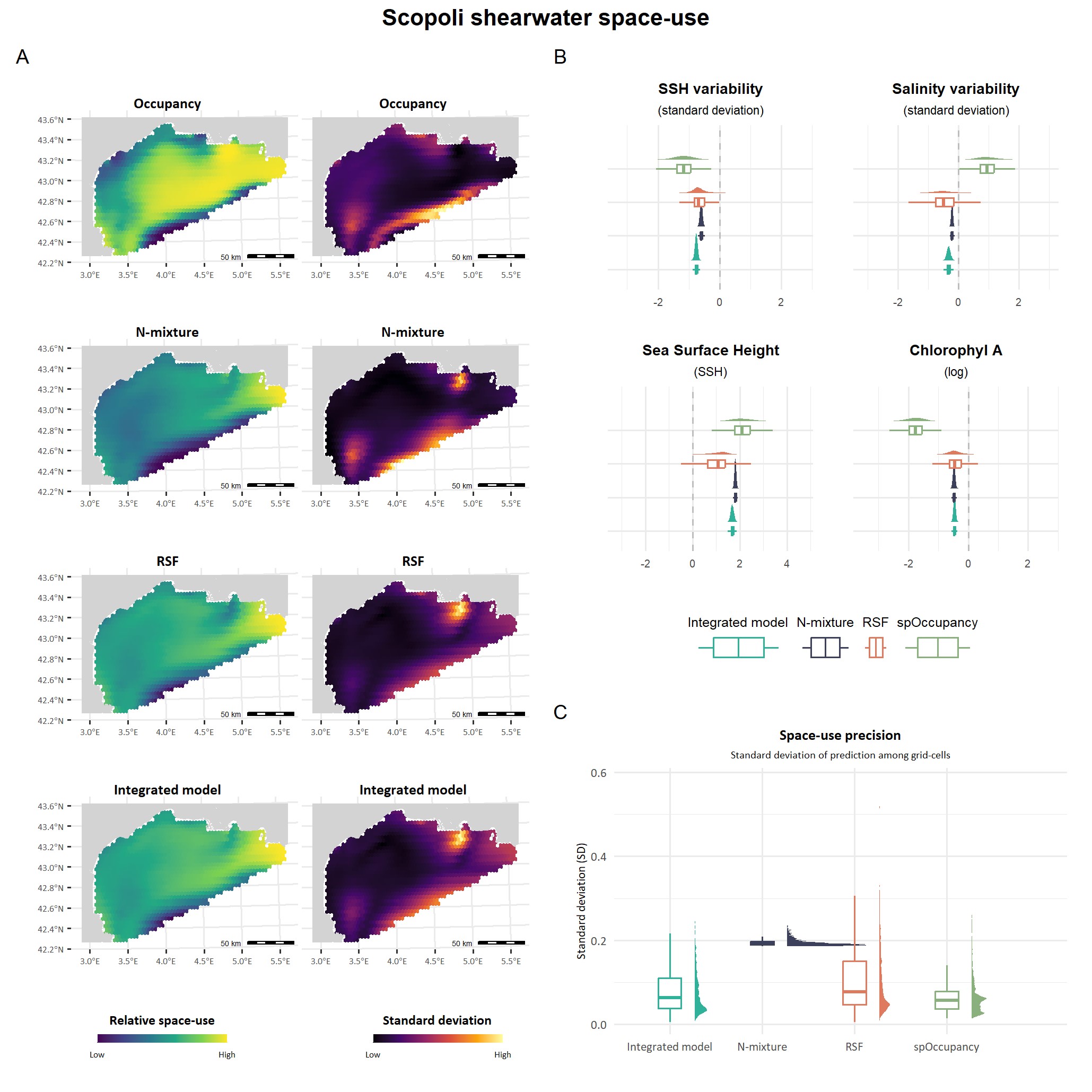
1. Erica F. Stuber et al. “Slow explorers take less risk: a problem of sampling bias in ecological studies”. *Behavioral Ecology* (Sept. 1, 2013). doi: [10.1093/beheco/ art035.](https://doi.org/10.1093/beheco/art035)
2. Corinne J. Kendall et al. “Value of combining transect counts and telemetry data to determine short-term population trends in a globally threatened species”. *Conservation Biology* (2023). doi: [10.1111/cobi.14146.](https://doi.org/10.1111/cobi.14146)
3. Annette L. Fayet et al. “Lower foraging efficiency in immatures drives spatial segregation with breeding adults in a long-lived pelagic seabird”. *Animal Behaviour* (Dec. 1, 2015). doi: [10.1016/j.anbehav.2015.09.008.](https://doi.org/10.1016/j.anbehav.2015.09.008)
4. Stephen C. Votier et al. “Inter-colony movements, at-sea behaviour and foraging in an immature seabird: results from GPS-PPT tracking, radio-tracking and stable isotope analysis”. *Marine Biology* (Feb. 1, 2011). doi: [10.1007/s00227-010-1563-9.](https://doi.org/10.1007/s00227-010-1563-9)
5. Ewan D. Wakefield et al. “Breeding density, fine-scale tracking, and large-scale modeling reveal the regional distribution of four seabird species”. *Ecological Applications* (2017). doi: [10.1002/eap.1591.](https://doi.org/10.1002/eap.1591)
6. Alexandra Egunez et al. “Weather effect on widely distributed seabirds through migration flyways: a case study with the Northern Gannet Morus bassanus in the Bay of Biscay”. *Bird Study* (July 3, 2018). doi: [10.1080/00063657.2018.1515177.](https://doi.org/10.1080/00063657.2018.1515177)
7. Stefan Garthe et al. “Effects of hydrographic and meteorological factors on seasonal seabird abundance in the southern North Sea”. *Marine Ecology Progress Series* (Sept. 28, 2009). doi: [10.3354/meps08170.](https://doi.org/10.3354/meps08170)
8. Nicolas Vanermen et al. “Attracted to the outside: a meso-scale response pattern of lesser black-backed gulls at an offshore wind farm revealed by GPS telemetry”. *ICES Journal of Marine Science* (Mar. 1, 2020). doi: [10.1093/icesjms/fsz199.](https://doi.org/10.1093/icesjms/fsz199)
9. R. C. Fijn et al. “GPS-tracking and colony observations reveal variation in offshore habitat use and foraging ecology of breeding Sandwich Terns”. *Journal of Sea Research*.

SI: North Sea Coastal System (Sept. 1, 2017). doi: [10.1016/j.seares.2016.11.](https://doi.org/10.1016/j.seares.2016.11.005)

[005.](https://doi.org/10.1016/j.seares.2016.11.005)

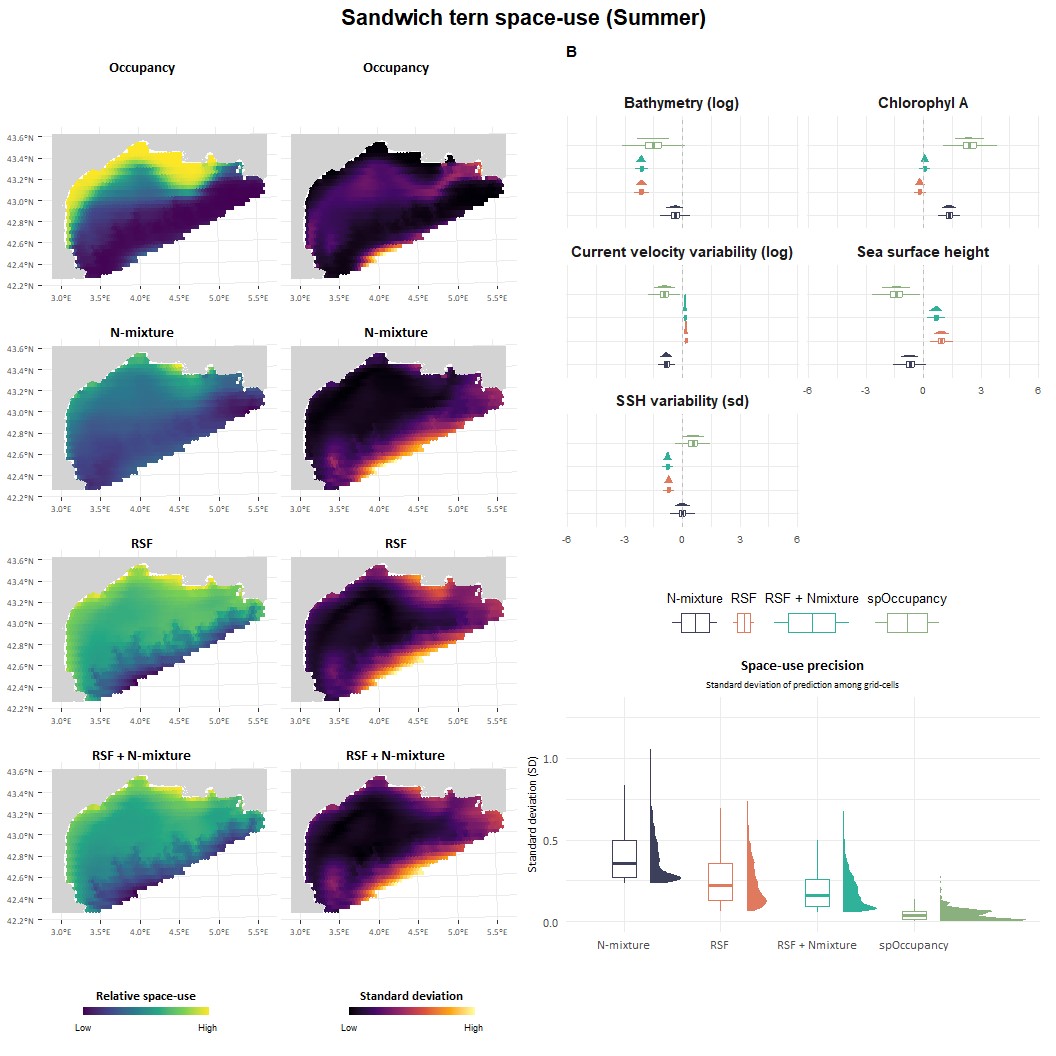
1. Philipp Schwemmer et al. “Behavioral responses to offshore windfarms during migration of a declining shorebird species revealed by GPS-telemetry”. *Journal of Environmental Management* (Sept. 15, 2023). doi: [10.1016/j.jenvman.2023.118131.](https://doi.org/10.1016/j.jenvman.2023.118131)
2. Alan E. Burger and Scott A. Shaffer. “Perspectives in Ornithology Application of Tracking and Data-Logging Technology in Research and Conservation of Seabirds”. *The Auk* (Apr. 1, 2008). doi: [10.1525/auk.2008.1408.](https://doi.org/10.1525/auk.2008.1408)
3. Alice Bernard et al. “Toward a global strategy for seabird tracking”. *Conservation Letters* (2021). doi: [10.1111/conl.12804.](https://doi.org/10.1111/conl.12804)

# II Appendix B

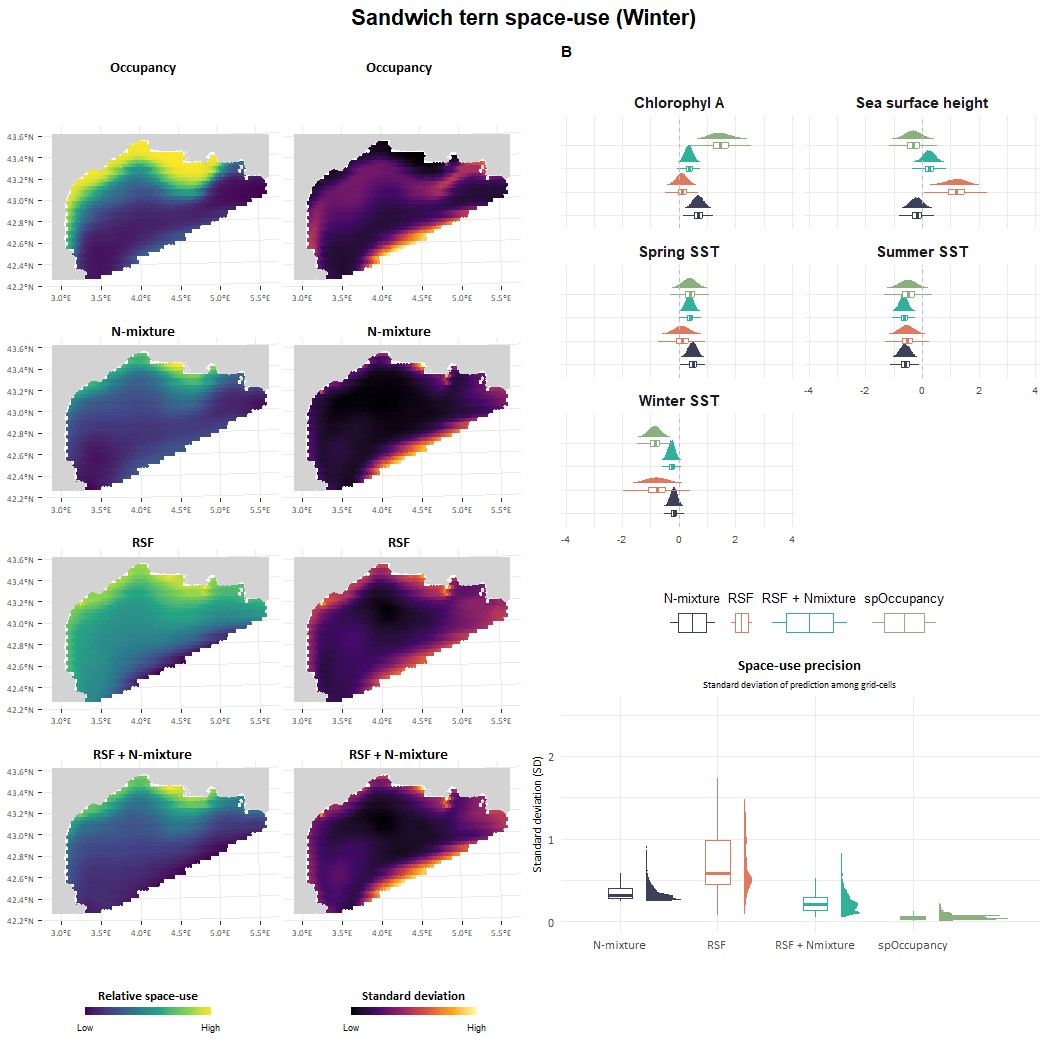


**Fig.7: Scopoli’s shearwater (***Calonectris diomedea***) distribution in the Gulf of Lion.**

Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining both movement and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty on the results, computed as the standard deviation on the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Prediction on the predicted space use, computed as the standard deviation of the grid cells.

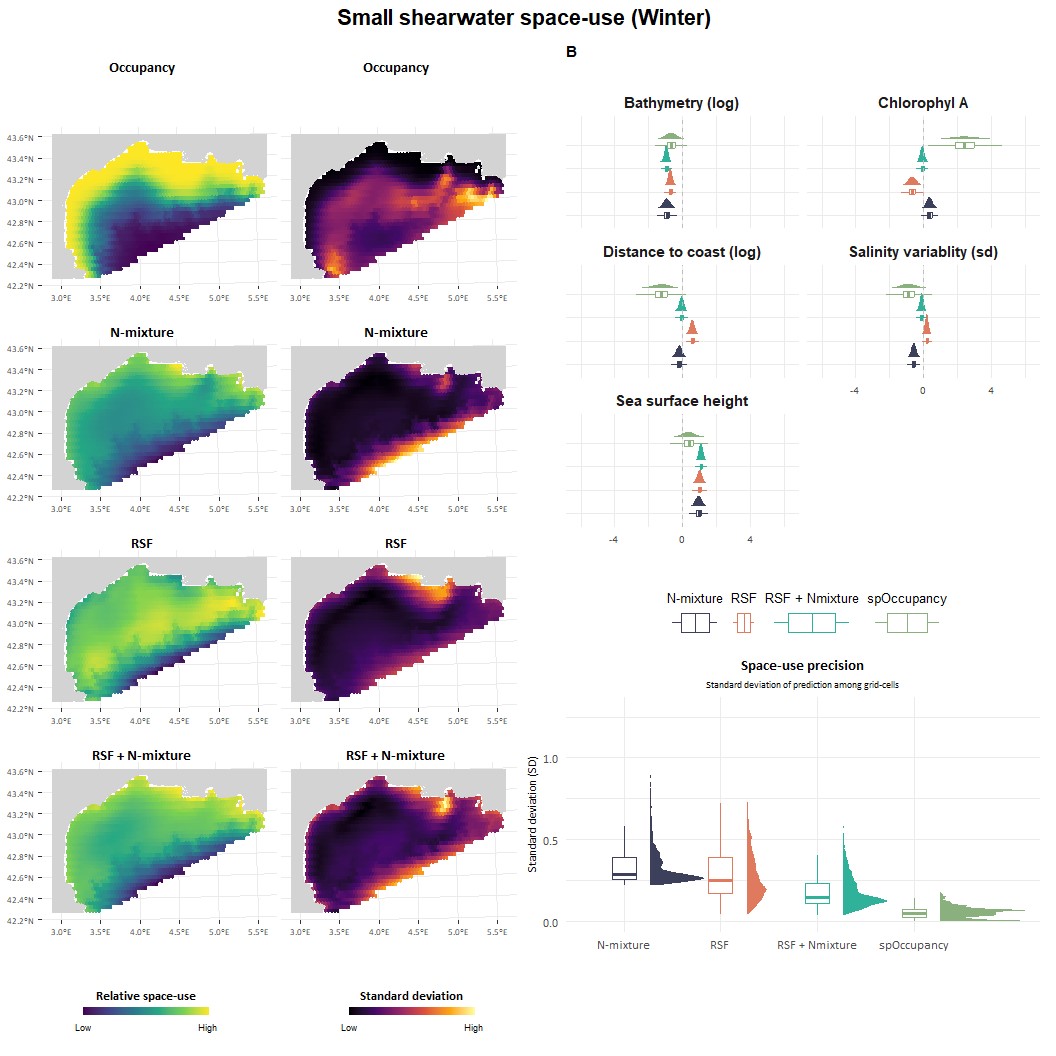


**Fig.8: Sandwich tern (***Sterna sandvicensis***) distribution in the Gulf of Lion.** Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining both movement and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty on the results, computed as the standard deviation on the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Prediction on the predicted space use, computed as the standard deviation of the grid cells.

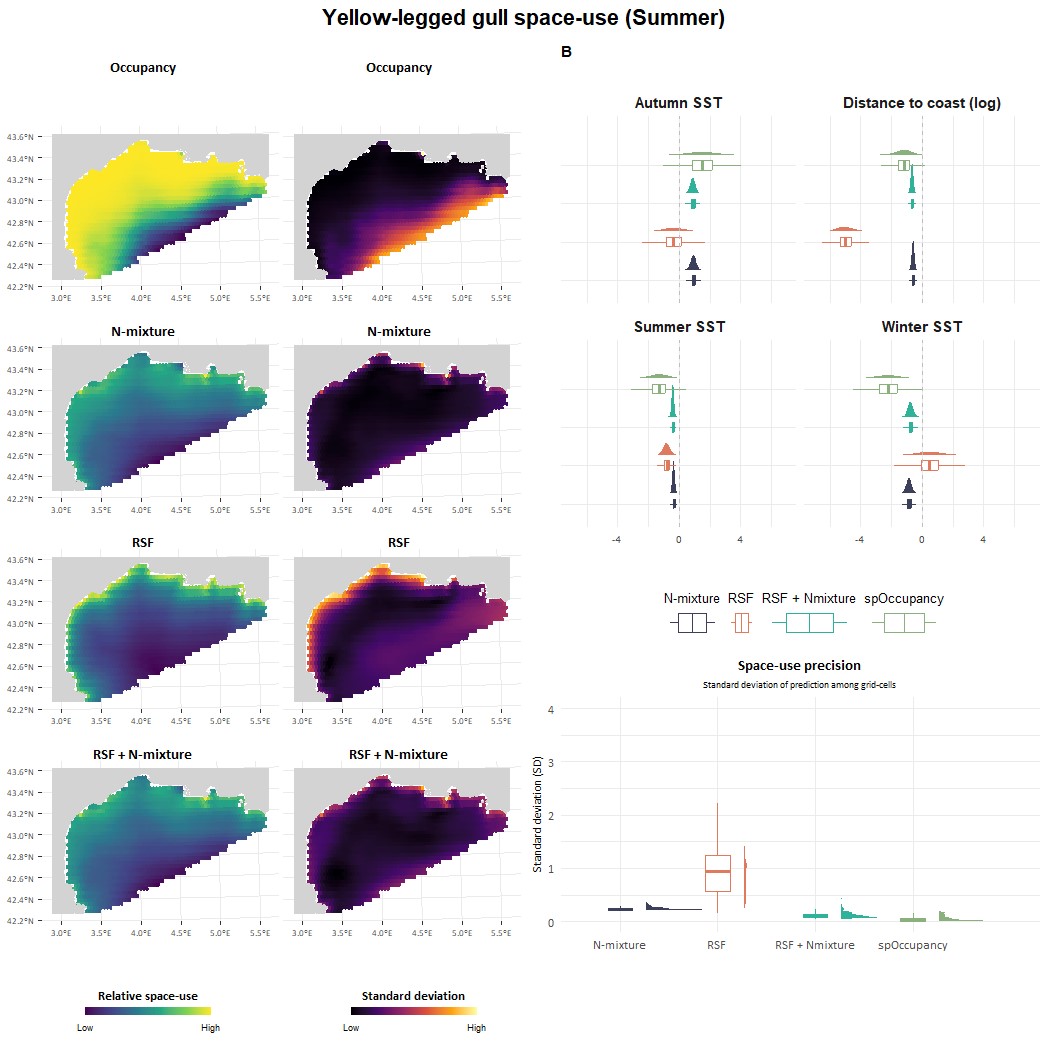


**Fig.9: Sandwich tern (***Sterna sandvicensis***) winter distribution in the Gulf of Lion.**

Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining both movement and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty on the results, computed as the standard deviation on the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Prediction on the predicted space use, computed as the standard deviation of the grid cells.



**Fig.10: Small shearwaters (***Puffinus sp***) winter distribution in the Gulf of Lion.** Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining both movement and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty on the results, computed as the standard deviation on the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Prediction on the predicted space use, computed as the standard deviation of the grid cells.



**Fig.11: Yellow-legged gull (***Larus michaellis***) summer distribution in the Gulf of Lion.**

Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining both movement and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty on the results, computed as the standard deviation on the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Prediction on the predicted space use, computed as the standard deviation of the grid cells.

# III Appendix C

