

Master 2 de Modélisation en écologie  
2023-2024

---

# Integrating Multiple Data Sources to Model Seabird Distribution to Inform Wind Farm Policies

---



**Louis SCHROLL**

Supervised by Valentin LAURET, Coline CANONNE  
and Aurélien BESNARD

Centre d'Ecologie Fonctionnelle et Evolutive  
1919 Route de Mende, Montpellier

## Acknowledgements

First, I want to thank Valentin Lauret and Coline Canonne for their guidance, help, and great advice throughout this internship. I also want to thank Aurélien Besnard for his supervision and advice, and Sébastien Roques for carefully proofreading the report and sharing his knowledge about seabirds. Thanks to Maud Queroue for her help and advice, which was much appreciated. I'm also grateful to Jean-Pierre Elise for his assistance with L<sup>A</sup>T<sub>E</sub>X, which made writing this report much easier.

Finally, I want to thank my colleagues on the HAIR team for their warm welcome and constant support. Their camaraderie made this experience even more enjoyable.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Methods</b>	<b>4</b>
2.1	Study area and data collection . . . . .	4
2.1.1	Study area . . . . .	4
2.1.2	Seabirds at-sea counts . . . . .	5
2.1.3	Telemetry data . . . . .	6
2.1.4	Environmental data . . . . .	6
2.2	Statistical analysis . . . . .	7
2.2.1	Seabirds occupancy models with detection/non-detection data . . . . .	7
2.2.2	Seabirds relative abundance modeling with count data . . . . .	8
2.2.3	Resource Selection Function using Telemetry data . . . . .	9
2.2.4	Relative space-use models integrating N-mixture and RSF models . . . . .	10
2.2.5	Implementation and model fitting . . . . .	10
2.2.6	Model selection . . . . .	10
2.2.7	Interpretation of models outputs . . . . .	12
2.3	Mapping seabirds' vulnerability to wind farms . . . . .	12
2.3.1	Vulnerability indexes . . . . .	12
2.3.2	Vulnerability maps . . . . .	13
<b>3</b>	<b>Results</b>	<b>14</b>
3.1	Comparison of seabird space-use modeled with detection/non-detection and count data . . . . .	14
3.2	Comparison and integration of GPS-tracking data with counts . . . . .	15
3.3	Seabird spatial vulnerability to wind farm . . . . .	17
<b>4</b>	<b>Discussion</b>	<b>19</b>
4.1	Comparison of seabird space-use estimated with detection / non-detection and count data . . . . .	19
4.2	Comparison of seabird space-use estimated with telemetry and count data . . . . .	20
4.3	Seabird vulnerability to wind farms . . . . .	22
<b>5</b>	<b>Conclusion</b>	<b>24</b>
<b>A</b>	<b>Appendix A</b>	<b>31</b>
<b>B</b>	<b>Appendix B</b>	<b>32</b>
<b>C</b>	<b>Appendix C</b>	<b>33</b>
<b>D</b>	<b>Appendix D</b>	<b>38</b>

# Internship Report

Louis SCHROLL

---

## 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 attributed to several factors, including the extensive continental shelf, the influx of freshwater and nutrients 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 are known to 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 that do not reside in the Gulf of Lion.

The Mediterranean seabird community is characterized by a relatively low species richness yet a high degree of endemism [10]. Furthermore, some species and subspecies that use the Gulf of Lion are categorized as threatened 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. The primary threat to these species on land is predation by feral cats and black rats on eggs, chicks, and adults [11, 12]. At sea, they are vulnerable to accidental bycatch by fisheries [9, 13], food shortages resulting from fishing activities [8, 14], various pollutants

[15, 16, 17], and disturbances caused by off-shore infrastructures [18]. The Gulf of Lion is a region of significant anthropogenic activity, including commercial fishing [8], tourism, and recreational activities. This accumulation of anthropogenic impacts affects 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 form part of an international initiative to develop renewable energy sources 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 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. Still, in some cases, resources can become more abundant than before construction, increasing the area's attractiveness for birds [27, 28]. This could provide new feeding opportunities for seabirds and improve their population dynamics, but it could also increase 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 considerations into project design alongside technical and financial elements. The primary focus should be avoiding important seabird areas [29]. This initial phase is indispensable and precedes all other measures to minimize wind farm impacts or ultimately compensate for residual impacts. That is why a spatial assessment of seabirds' vulnerability to these installations is essential. Such a spatial assessment requires understanding the spatial distribution of seabird species within the area of potential wind farm development. While the location of wind farms is currently under decision, a comprehensive assessment of seabird distribution at the scale of the Gulf of Lion has not yet 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 for informed decision-making, the assessment of species distribution should encompass a broader range of species than previous studies.

Several programs have collected data on seabirds in the Gulf of Lion from 2011 to 2023 [3, 31]. Most 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 sur-

vey [32, 33, 34]. Besides, for some species, telemetry data is also available. Telemetry 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, altitude, and direction.

Current assessments of seabird distribution rely on single-dataset approaches, with each dataset analyzed separately even though it measures the same process. Taking only one dataset offers limited information on a given species. Furthermore, data sources collected using different technologies provide complementary information concerning the same ecological process. Combining all these data in a single analysis should allow us to comprehensively and accurately assess seabird distribution.

The most reliable method to combine data from multiple sources is to develop so-called "integrated species distribution models" [35, 36, 37]. Integrated species distribution models allow to take detection process and protocol differences into account [35, 36, 37]. Integrated models frequently take the form of 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 comprise two components: one for the ecological state of the system and another for the observation process [36, 34]. The component for the observation process is different for each data source, while the component for the ecological state is the same for all data sources. The un-

derlying assumption is that each observation protocol provides a different representation of the same reality.

Integrated models offer several major advantages over models based on a single data source. The primary one is 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].

The choice of the species distribution model will depend on the available data and the specific ecological question. While occupancy models [34, 32, 40] use detection/non-detection data to estimate the presence probability of a species at a site on a defined area, N-mixture models [34, 33, 41] rely on individual counts and give information on the abundance of a species at each site. For this reason, N-mixture models can provide more detailed information on spatial processes than occupancy models and one would expect the power to detect spatial variations to be stronger for count data. However, the gains associated with this method depend on the characteristics of the species studied, notably their rarity and density [42]. So far, occupancy models are more widely used than N-mixture models to guide public policy or understand species distribution [43]. There are two main reasons behind that choice. First, detection/non-detection data are more frequently available due to their lower cost. Second, occupancy models are often easier to fit than N-mixture models [42], as the latter's fundamental assumptions are frequently not

met [34]. Therefore, in the context of wind farm planning, it appears necessary to evaluate whether detection/non-detection data may be sufficient and can be used as a surrogate for abundance estimation to assess the spatial vulnerability of seabirds to wind farms.

Apart from count data, spatial distribution can also be determined from positions acquired by GPS devices [44]. 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 [44, 45]. Although not its primary focus, habitat selection analysis can provide insights into species distribution [44].

As the quantity and variety of data increase, there is 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 GPS and count data information to determine species distribution, particularly in seabird research [1, 46]. However, none of these studies used a model-based approach. Instead, they modeled the two data types separately with different objectives for each. Lauret et al. (under review) [47] developed a method to combine telemetry and count data into a single model. They used a hierarchical model, 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 [45]. In this study, we used this model to formally integrate count and telemetry to estimate seabird distribution.

Interpreting several distinct distribution maps describing fine patterns of risk for each seabird species is challenging for public policymakers, as they may largely differ between species and do not contain information on species-specific vulnerability to wind farms. The construction of vulnerability indexes is generally favored 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, 48]. To address species-specific differences, Furness et al. (2013) [23] and Bradbury et al. (2014) [48] developed a 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 [49, 48]. This approach enables to put more emphasis on 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

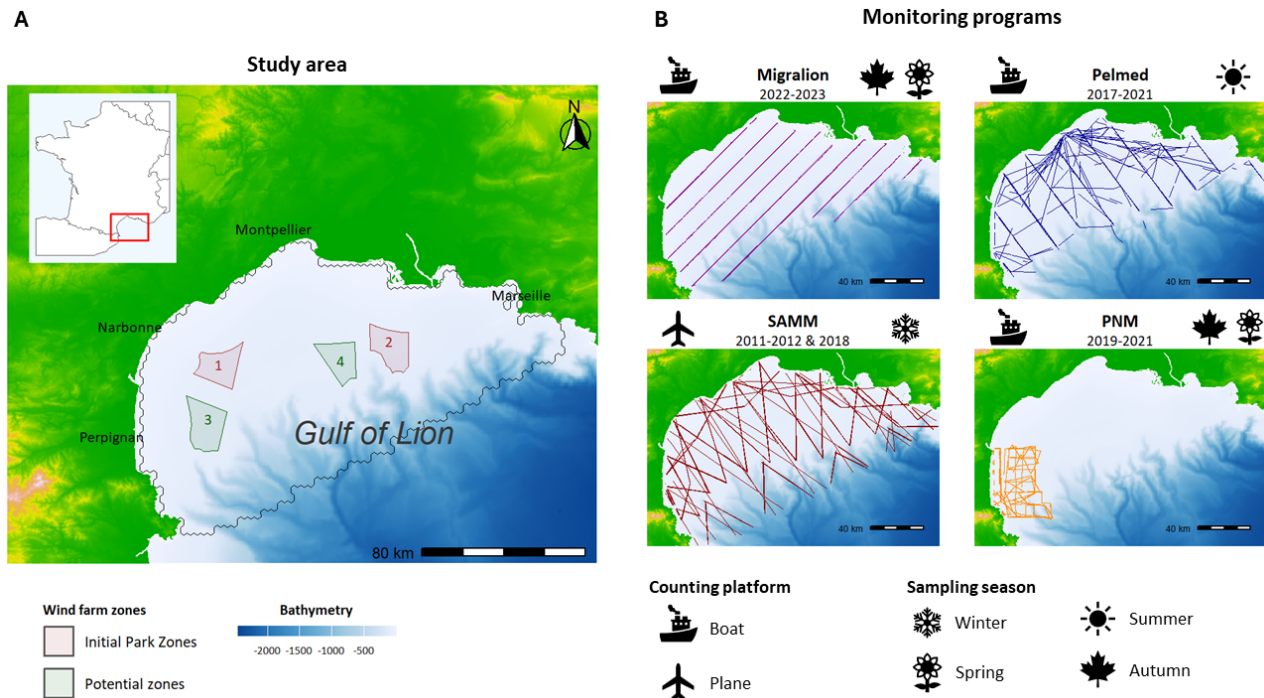
answer these questions, we (1) make the most of existing data to predict reliable distribution of seabirds in the Gulf of Lion and (2) use these predictions to assess their spatial vulnerability to future wind farm installations. We modeled species distributions using various techniques and data types, including integrated occupancy and integrated N-mixture models. These models combined at-sea observations from four different programs. We performed habitat selection analysis for species with available GPS-tracking data 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 (Figure 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 [50]. The study area was determined based on suitability to receive wind farms, mainly driven by the bathymetry. As a consequence, the study area coincides with the continental shelf. The total area considered for the study had a surface of approximately 17,300



**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. The Migration program was performed every year from April to September 2022. PELMED (PELagic MEDiterranean) was conducted every year from July 2017 to 2021. SMM (Survol Aérien de la Mégafaune 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.

km<sup>2</sup>. Within this area, assessment units were defined by 1380 hexagonal cells with a size of 12,6 km<sup>2</sup>. 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.

### 2.1.2 Seabirds at-sea counts

Count data were collected from four programs, as summarized in Figure 1. Observations were conducted using strip-transect

methodology (Figure 1) [51, 52, 53], 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 [51].

SMM (Survol Aérien de la Mégafaune Marine, Aerial Census of Marine Megafauna) [30, 54], was the only aerial survey. It was conducted over the course of two periods: November to February 2011-2012 and from January to February 2018. This program covered 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 (PELagic MEDiterranean) is an annual survey primarily conducted to assess small pelagic fish resources [31]. It also collects data on seabirds, with an ornithologist on board. This campaign occurred every summer from late June to early August.

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 were collected in 2022 and 2023, twice a year in spring and autumn, to cover pre-breeding and post-breeding bird migration periods.

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 distributions of twelve species were modeled (see species list in Table 3).

Seabird distribution can change between seasons, and some species are only present in the study area during certain periods of the year [30, 54]. For this reason, we split our analysis 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. In 2021 and 2022, the Migralion project deployed GPS-tracking devices on 51 Scopoli's shearwaters, 39 Yelkouan shearwaters, and 22 Sandwich terns were equipped with GPS at their nesting colonies in the South of France; 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 [55, 56] (<https://data.marine.copernicus.eu>) and MARSPEC database [57] (<http://www.marspec.org>). A total of eight geophysical and biotic covariates were included and are listed in Table 1 (see also Appendix A Figure 6 for maps). The choice of covariates was based on knowledge about seabirds ecology and by the most frequently used covariates for seabird distribution modeling [1, 30, 58, 59, 60].

The dynamic variables (dynamic and SST covariates in Table 1) were obtained with

**Table 1: Environmental covariates used in the study.** The column "Block" indicates the different groups of covariates used in model selection.

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 covariate
SSS	Sea Surface Salinity	Dynamic covariate
SSH	Sea Surface Height above geoid (in m)	Dynamic covariate
Chlorophyll A	Chlorophyll-A concentration (in mg/m <sup>3</sup> )	Dynamic covariate
Current velocity	Sea water velocity (in m/s)	Dynamic covariate

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 a measure of temporal variability of the covariate value. Sea Surface Temperature (SST) showed considerable seasonal variability. Thus, average values were calculated for each season. For each covariate, the values were calculated to align with 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 each species' seabird distribution in the Gulf of Lion. 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 presence probability  $\psi$  [32, 34].

#### Latent process

Let's consider  $J$  sites (i.e., hexagonal cells of the grid), each surveyed  $K$  times. We denote  $Z_j$  the true occupancy state at site  $j$ , where  $j \in 1, \dots, J$ .  $Z_j = 0$  if the species is absent and  $Z_j = 1$  if it is present. This latent occurrence variable is assumed to follow a Bernoulli process [32]:

$$\begin{aligned} Z_j &\sim \text{Bernoulli}(\psi_j) \\ \text{logit}(\psi_j) &= \beta_0 + X_j^T \beta \end{aligned} \quad (1)$$

Here,  $\psi_j$  represents the probability of species occurrence at site  $j$ . The occurrence probability is modeled by logistic regression of site-specific environmental covariates  $X_j$ , with  $\beta$  being a vector of regression coefficients [32, 34].

### 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 [61] (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, 62, 63, 64].

Let  $Y_{j,k,r}$  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, shared for all data sources. For a given data source  $r$ , the detection/non-detection data arises from a Bernoulli process conditional on the true latent occurrence process:

$$Y_{j,k,r} \sim \text{Bernoulli}(p_{j,k,r}Z_j)$$

$$\text{logit}(p_{j,k,r}) = \alpha_r^0 + \alpha_r^1 \text{transect length}_{j,k,r} \quad (2)$$

Where  $p_{j,k,r}$  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.

Multiple sampling occasions are needed to ensure that detection probability can be accurately identified. Counts were repeated for each dataset over different years. We considered these counts 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 colony locations that remain relatively fixed over time. Their population sizes are not expected to vary significantly between years, except for occasional migration variability.

### 2.2.2 Seabirds relative abundance modeling with count data

As a second step, we inferred seabird at-sea abundance using N-mixture models [41, 33]. Like 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  $N_j$  represent the true abundance at site  $j$ . This latent variable is assumed to follow a negative binomial distribution:

$$N_j \sim \text{NegBin}(\lambda_j, \kappa)$$

$$\log(\lambda_j) = \beta_0 + X_j^T \beta \quad (3)$$

Where  $\lambda_j$ , the mean of the distribution, is modeled as a function of site-specific covari-

ates, and  $\kappa$  is the overdispersion parameter. A negative binomial distribution was preferred to a Poisson distribution to account for overdispersion [65, 61], as the negative binomial allows for the standard deviation and the mean to be different (the variance is  $\lambda + \lambda^2/\kappa$ ). Overdispersion is frequent with seabird 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  $N_{j,k,r}^{obs}$  is assumed to follow a Binomial distribution conditional on the true number of individuals  $N_j$ :

$$\begin{aligned} N_{j,k,r}^{obs} &\sim \text{Binomial}(p_{j,k,r}, N_j) \\ \text{logit}(p_{j,k,r}) &= \alpha_r^0 + \alpha_r^1 \text{transect length}_{j,k,r} \end{aligned} \quad (4)$$

In other words, each individual at site  $j$  has a probability  $p_{j,k,r}$  of being detected. Detection is considered independent for each individual. The detection probability  $p_{j,k,r}$  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 the logistic regression of used and available locations is fitted simultaneously on the entire dataset.

We randomly selected 10 available points, or "pseudo-absences," for each recorded location on the GPS track," 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, 66].

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

$$\begin{aligned} K_j &\sim \text{Bernoulli}(\lambda_j) \\ \text{logit}(\lambda_j) &= \beta_0 + X_j^T \beta \end{aligned} \quad (5)$$

Where  $\lambda_j$  is the probability for a location being used knowing its environmental covariates values ( $\lambda_j = P(K_j = 1|X_j)$ ).  $\lambda_j$  is modeled as a logistic regression of location-specific covariates and regression coefficients  $\beta$ . Coefficients  $\beta$  represent the strength of the selection on the associated covariates.

To model potential population heterogeneity, we added a random effect on slope coefficients  $\beta$ . Let's  $I$  denote the total number of individuals, and take an individual  $i \in [1; I]$ , we have:

$$\beta_i \sim \text{Normal}(\beta, \tau) \quad (6)$$

With  $\beta$  being the slope coefficient and  $\tau$  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 [44, 45] showing that these models are similar to an Inhomogeneous Poisson Point process (IPP), Lauret et al. (under review) [47] proposed to integrate these two data sources into 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 with environmental covariates, as in occupancy, N-mixture, and RSF models. An IPP reflects the intensity  $\lambda$  of a spatial point process over a defined study area. N-mixture and RSF do not estimate the same IPP absolute intensity  $\lambda$ . In RSF, the intercept  $\beta_0$  has no biological meaning and only the relative habitat selection is estimated [67]. Since the relative IPP intensity  $\lambda$  is determined by the slope coefficients  $\beta$ , the 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  $\beta$  parameters.

The model is divided into the N-mixture submodel:

$$\begin{aligned} N_j &\sim \text{NegBin}(\lambda_j^{Nmix}, \kappa) \\ \log(\lambda_j^{Nmix}) &= \beta_0^{Nmix} + X_j^T \beta \end{aligned} \quad (7)$$

And the RSF submodel:

$$\begin{aligned} K_j &\sim \text{Bernoulli}(\lambda_j^{RSF}) \\ \text{logit}(\lambda_j^{RSF}) &= \beta_0^{RSF} + X_j^T \beta \end{aligned} \quad (8)$$

#### 2.2.5 Implementation and model fitting

All models were implemented in R (ref). Occupancy models were fitted with the package `spOccupancy` [68]. The package provides functionality for fast data integration of multiple single-species occupancy data sets and easy model selection, while using a joint likelihood framework and Markov chain Monte Carlo (MCMC). 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 using the `Nimble` package [69]. 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  $\hat{R} < 1.1$  [70] and by visual checks on trace plot.

#### 2.2.6 Model selection

The selection procedure was carried out exclusively using the occupancy model. For



Table 2: Priors specified for each model.

Model	Parameter	Prior
Occupancy	Occurrence regression coefficients $\beta$	$\beta \sim N(0, 2.72)$
	Detection regression coefficients $\alpha$	$\alpha \sim N(0, 2.72)$
N-mixture	Abundance regression coefficients $\beta$	$\beta \sim N(0, 1)$
	Detection regression coefficients $\alpha$	$\alpha \sim N(0, 1)$
	Overdispersion parameter $\kappa$	$\kappa \sim Unif(0, 100)$
RSF	Space-use regression coefficients $\beta$	$\beta \sim N(0, 1)$
	Inter-individual variability parameters $\tau$	$\tau \sim N(0, 100)$
RSF + N-mixture	Same priors than for RSF and N-mixture alone	

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 too long to perform model selection with the other models during the internship due to the computation time required.

To have the most accurate prediction of 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) [71] and cross-validation [72]. 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 [72]. 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. We 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 [73]. The `spOccupancy` package offers various spatial covariance models (spherical, exponential, and Gaussian) [68], which we all tested.

### 2.2.7 Interpretation of models outputs

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, the model output 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, implies that the latent ecological state of a site (i.e. a grid cell) remains unchanged between the repeated visits [74]. For seabirds, this as-

sumption is unrealistic, as they can easily travel from one site to another, and occupied sites are used only temporarily by individuals [75]. The size of a sampling grid cell is much smaller than their range of activity. In this situation, the occupancy estimator  $\psi_j$  represents the probability that grid cell  $j$  is used by the target species [76], and is interpreted as space-use by the seabird species. Similarly, N-mixture estimation does not capture the expected number of individuals at a site but reflects the relative number of individuals using the site when it is surveyed [34, 77]. 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 a relative intensity of space-use, which is consistent across different data sources and modeling approaches. Thus, we will use the term "relative space-use", despite subtle differences in what each method specifically measures.

## 2.3 Mapping seabirds' vulnerability to wind farms

### 2.3.1 Vulnerability indexes

Based on the results of the occupancy and N-mixture models, we constructed a synthetic index to evaluate species vulnerability to offshore wind farms integrating three main aspects of seabird vulnerability [48, 23]:

1. **Conservation Status (CS):** Defined in OFB and GISOM (2020) [78], this score

**Table 3: Vulnerability index for each species.** Blank space indicates no score for a species at a certain period due to its absence from the area. Breeding season last from April to August and Non-breeding season from September to March

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

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 space-use can vary greatly between these periods [78, 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]:

$$Vul = \frac{Alt + Man + Ptf + Noc}{4} \times \frac{Dis + Spe}{2} \times CS$$

Vulnerability indexes for every species are indicated in Table 3 and the values of the criteria used are indicating in Appendix B, Table 4.

### 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:

$$SV_{j,S} = \sum_{sp} Vul_{sp} \times \psi_{j,sp,S}$$

where  $\psi_{j,sp,S}$  represents the space-use probability of site  $j$  by species  $sp$  during season  $S$ .

With maps based on N-mixture, the seasonal seabird vulnerability  $SV$  at site  $j$  and season  $S$  was defined following Bradbury et al. (2014) [48]:

$$SV_{j,S} = \sum_{sp} Vul_{sp} \times \log(N_{j,sp,S} + 1)$$

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:

$$SV_{tot} = \text{mean}(SV_B, SV_{NB})$$

This approach ensured that both breeding and non-breeding seasons contributed to the overall assessment of seabird vulnerability.  $SV_{tot}$  was then centered-reduced to obtain a relative vulnerability score over the entire study area. The mean of relative vulnerability was 0, with positive values indicating areas with higher vulnerability and negative values indicating areas with lower vulnerability.

### 3 Results

All models effectively discriminated between less-used and most-used areas and showed good concordance with observable patterns in the data. Models that included spatial autocorrelation resulted in low predictive performance and overfitting; hence, none of the

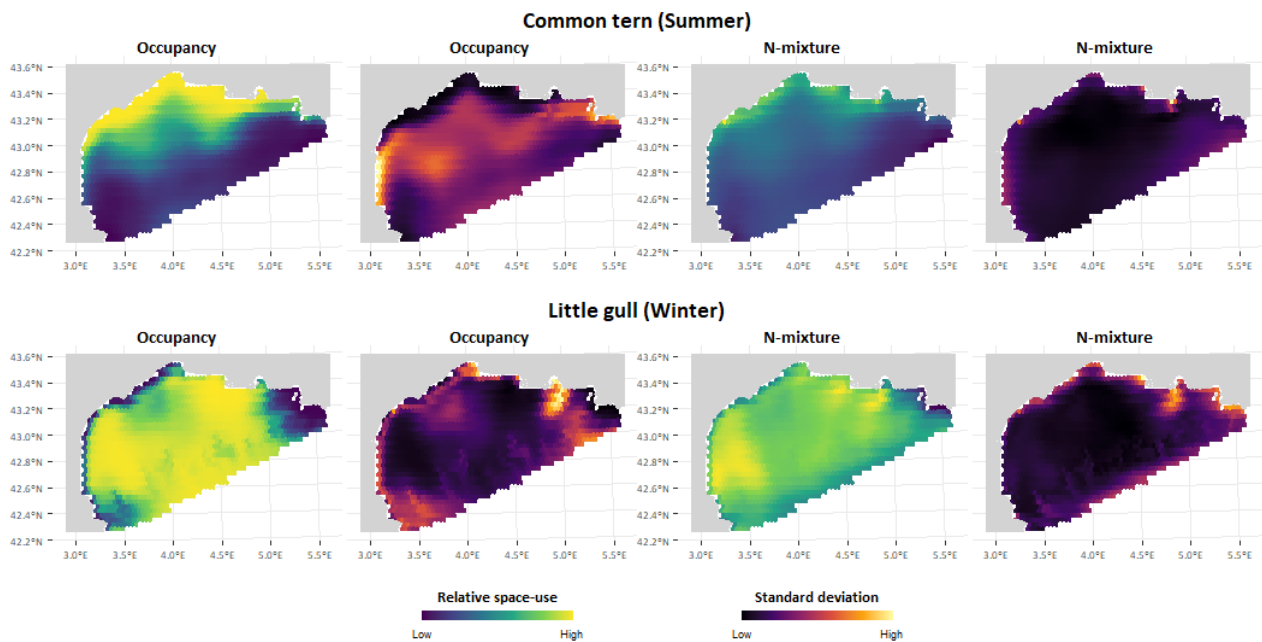
selected models included a spatial autocorrelation structure. The results of occupancy and N-mixture models are presented hereafter for four species representative of the pattern variations: Common Tern and Little Gull in Table 2, and Yellow-legged Gull and Small Shearwaters in Table 3. Results for other species are provided in Appendices C and D.

#### 3.1 Comparison of seabird space-use modeled with detection/non-detection and count data

**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



**Fig.2: Distribution of Common tern (*Sterna hirundo*) and Little gull (*Hydrocoloeus minutus*) 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 of the prediction.

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, Figure 4). Generally, the values remain close, but sometimes, the slope direction is different. For example, for the Yellow-legged gull (Figure 3), the  $\beta$  coefficient associated with autumn SST has a negative value in the occupancy model but a positive value in the N-mixture model.

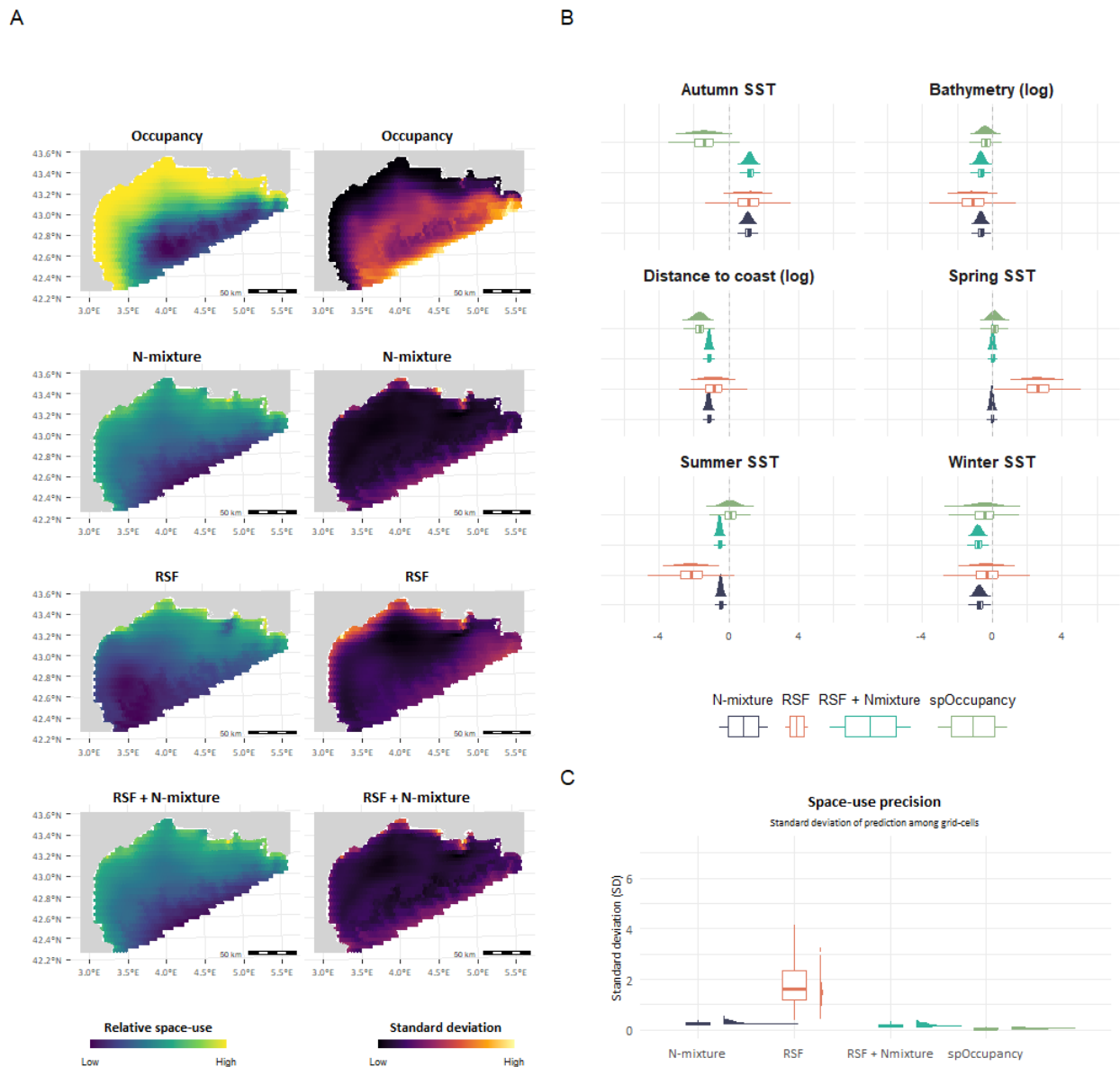
### 3.2 Comparison and integration of GPS-tracking data with counts

The results obtained with RSF using telemetry data are consistent with those derived from N-mixture using count data (Figure 3, Figure 4) (Appendix C 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



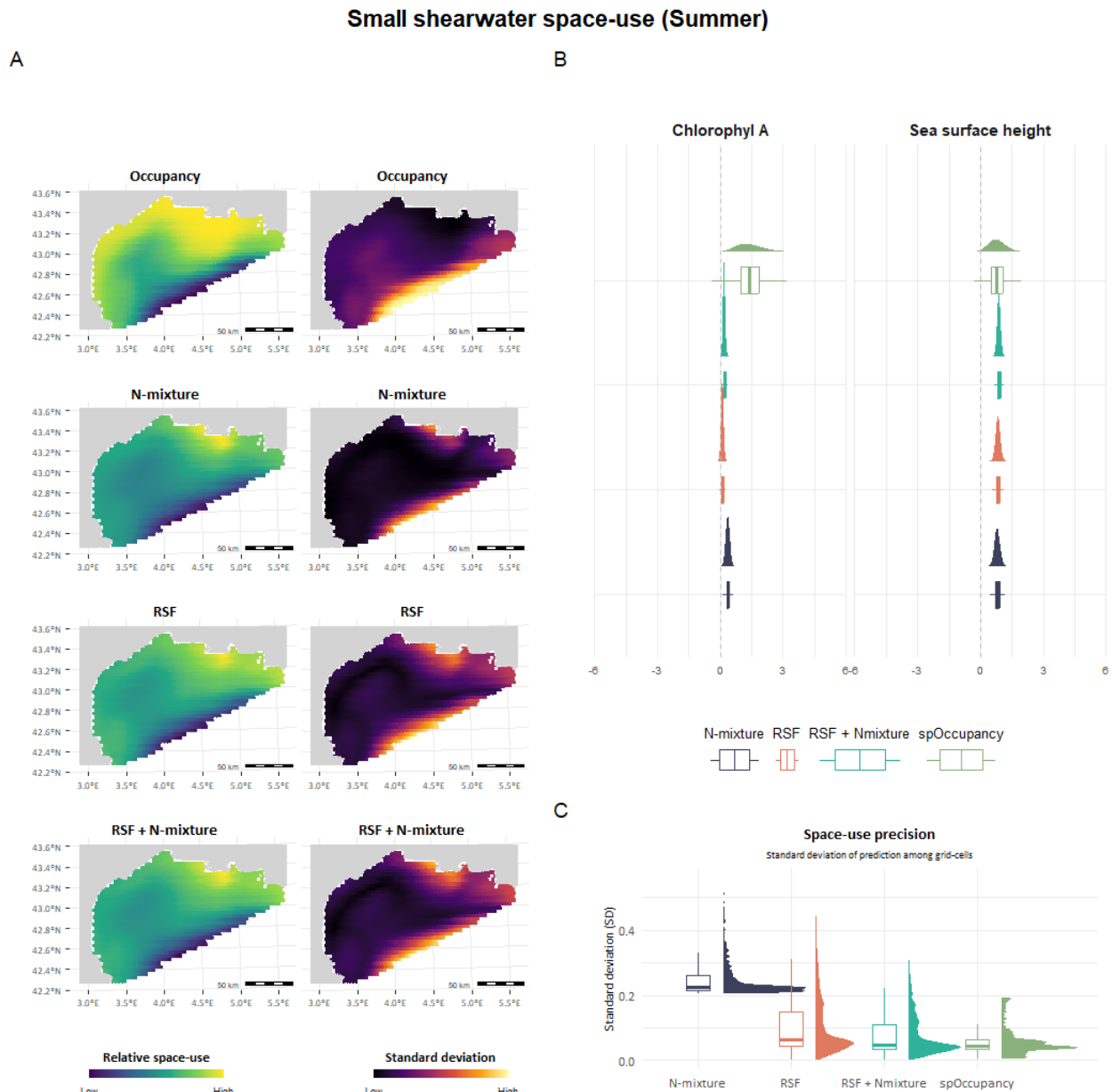
## Yellow-legged gull space-use (Winter)



**Fig.3: Yellow-legged gull (*Larus michaellis*) distribution in the Gulf of Lion.** At-sea space use is 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 in the results, computed as the standard deviation of the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Precision of the predicted space-use, computed as the standard deviation of the grid cells.

data yields contrasting results depending on the species. For small shearwaters (Figure 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 coeffi-

cients show considerable uncertainty. The final map and coefficient estimates 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



**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 GPS-tracking data, and an integrated model combining both GPS-tracking and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty in the results, computed as the standard deviation of the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Precision of the predicted space-use, computed as the standard deviation of the grid cells.

alone, resulting in identical maps.

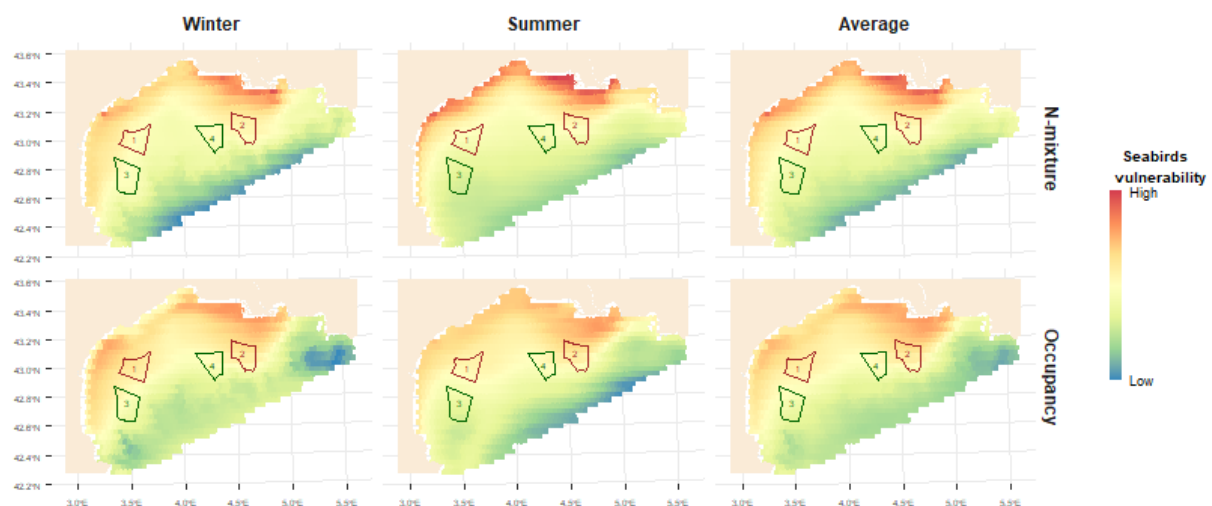
### 3.3 Seabird spatial vulnerability to wind farm

The species with the highest vulnerability scores are the Little Gull, the Mediterranean

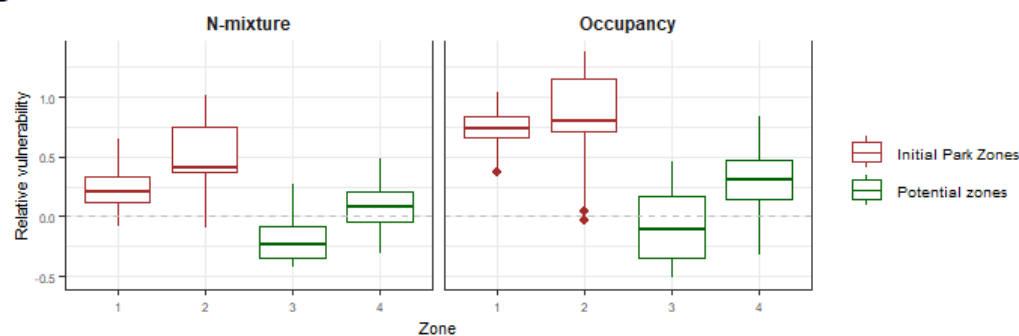
Gull, and the Yellow-legged Gull in winter (non-breeding season), and the Yellow-legged Gull, Black-headed Gull, and Mediterranean Gull in summer (breeding season) (Table 3).

Seasonal variation in the spatial distribution of relative vulnerability is minimal, regardless of the methods used for species dis-

A



B



**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 over 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.

tribution (Figure 5). The main difference appears in the western part of the Gulf, where there is an increased risk in winter due to the presence of the Little Gull during this period.

Throughout the area, higher vulnerability values are observed near the coast. In the maps based on N-mixture, vulnerability decreases rapidly with increasing distance from the coast, whereas it remains high over a larger area in the occupancy-based maps. 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 relative vulnerability of the different areas being considered for wind farms is directly linked to the vulnerability patterns shown on the maps (Figure 5). With N-mixture models, vulnerability decreases rapidly with the distance from the coast, so

that zones 3 and 4 show a relative vulnerability equivalent to that of the entire study area. In contrast, zones 1 and 2 have a much higher vulnerability than the rest of the study area. With occupancy models, the 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 Comparison of seabird space-use estimated with detection / non-detection and count data

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 between the two methods, with the main difference occurring in areas of high space-use. Maps of species space-use reflected differences that make sense considering the focus of each method. While detection / non-detection data used in occupancy models have the same weight no matter how many individuals have been detected, count data used in N-mixture allows for the differentiation between levels of observed abundance, resulting in more contrasted space-use intensity maps. Nevertheless, occupancy models based on detection/non-detection data still offered an accurate picture of seabird distribution.

When the distribution maps for all species are merged into a single map, both methods yield very similar results. For both meth-

ods, the relative vulnerability of future wind farm development areas appears higher in the same zones. However, the areas of highest vulnerability appear more dispersed on the maps obtained using occupancy than those derived from N-mixture models. This difference stems directly from the variations observed in individual species maps. In our assessment area, using detection/non-detection data rather than count data tends to provide more conservative recommendations for wind farm planning. However, this might not always be the case, as detection/non-detection data could potentially obscure areas of very high risk, corresponding to areas of really high space-use by multiple species.

Other studies assessing seabird risk related to wind farm construction have used abundance or density of seabirds [48, 79, 80], without using detection/non-detection data. However, detection/non-detection data are easier and cheaper to collect and are often more abundant [81, 82]. In some cases, detection/non-detection data allow more sites to be surveyed because it takes less time to detect the presence of a species than to count all individuals. However, this advantage might be less pronounced in open environments such as the sea. Compared to detection/non-detection data, count data are more sensitive to detection bias, notably when it is strongly heterogeneous. Furthermore, in N-mixture models, the assumption of independent detection among individuals is often violated for species that occur in groups, such as seabirds.

The primary focus of our study was to assess seabird vulnerability before wind farm construction. If the data also aim to evalu-

ate the wind farm's impact post-construction by comparing the state before and after, detection/non-detection data may not be as appropriate as count data. Indeed, while abundance might decrease in the area where the wind farm is built, species presence could remain unchanged (e.g. [83, 84]). Therefore, if detection/non-detection data might be relevant for initial wind farm planning, count data should be preferred when collecting new data.

## 4.2 Comparison of seabird space-use estimated with telemetry and count data

For four species, including Yellow-legged gull, Sandwich tern, Scopoli shearwater, and small shearwaters, we also modeled distribution using telemetry data with RSF. For telemetry data, the main bias arises from the fact that only a small number of individuals, generally breeding adults, all from the same colony, are equipped with GPS [85]. When the sample size is too small, it limits the ability to make accurate population-level inferences, resulting in low statistical power [86] and high bias due to individual-specific traits [87, 88]. Although we took into account individual heterogeneity in the RSF, limited sample size can affect the accuracy of inferences at the population level. For example, younger seabirds, not equipped with GPS devices, are likely to have different distributions and foraging behaviors compared to adults [89, 90].

Count data from at-sea surveys have the advantage of recording all individuals regardless of their behavioral traits, age, reproduc-

tive status, or colony of origin. However, these surveys are limited in time and are only conducted during good weather conditions and daylight hours [3, 85]. In contrast, telemetry data provide information continuously, regardless of time or weather. Since space-use can depend on weather and time of day [91, 92], these factors may explain differences in distribution predicted by telemetry and count data. Carroll et al. (2019) [85] explored the differences between distributions obtained from GPS devices and boat-based surveys for two species. They found a stronger match between the two data types when considering only daytime data for one species.

Other drawbacks of count data include limited taxonomic resolution for closely related species and observation errors. Despite these drawbacks, count data have the advantage of indicating both the presence and absence of a species in an area (although non-detection does not always equate to absence), unlike telemetry data, which only show locations visited by sampled individuals and do not indicate their absence [44].

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, and Appendix C Figure 7).

For Sandwich terns, the distribution maps and relationships between distribution and environmental covariates obtained with N-mixture and RSF were less consistent than for shearwaters (Appendix C, Figure 8, Figure 9). One reason could be the small sample size of only 22 individuals equipped with



GPS devices. Additionally, collar failures, individual deaths, and emigration outside the study area further reduced the available telemetry data for this species. The distribution of Yellow-legged gulls (Figure 3) based on telemetry data showed a bias compared to the map obtained with the N-mixture, with greater predicted space use on the eastern side of the gulf. This could be due to a poor estimate of the effect of specific covariates, as all birds were equipped on the east side of the gulf.

In contrast, for Scopoli's shearwaters (Appendix C, Figure 7), 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 obtained with RSF and N-mixture. This is also true for Yelkouan shearwaters (Figure 4). Although count data did not differentiate between Balearic and Yelkouan shearwaters, Yelkouan shearwaters are much more abundant than Balearic shearwaters.

Telemetry data offers 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 disturbance [93, 86, 94]. This can be particularly useful in the context of wind farm development where such data have been used to assess the behavioral response of seabirds to new wind farms [95, 96, 97], or to estimate flight height [98], a key determinant of collision risk [23].

Financial resources often determine the choice between data types. While telemetry

data require fewer human resources than at-sea surveys, telemetry device costs remain high, although they are decreasing [86]. 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 [99]. Conducting a comprehensive assessment using only telemetry data is challenging because typically, only a few well-studied species are equipped [100], and small species, such as 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 [86]. Additionally, it would be crucial to equip birds from multiple colonies [85], further increasing both human and financial costs.

When we modeled species distribution by combining telemetry and count data in an integrated model (hereafter referred to as the integrated model), we expected the results from this integrated model to be less biased and more accurate [101, 47]. 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, possibly indicating less bias. For Yellow-legged gulls, the maps from the integrated model closely matched those from the count data alone, likely because the count data had many non-zero counts, suggesting it contained more information than the telemetry data. Conversely, the integrated model sometimes produced results identical to those of the RSF, suggesting that telemetry data had more weight

for this species (e.g., summer distribution of Sandwich tern, Appendix C, Figure 8).

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

### 4.3 Seabird vulnerability to wind farms

Despite substantial variation in seabird distribution among species, averaging the spatial vulnerability across species reveals two distinct and consistent patterns (Figure 5). The first pattern is a strong inshore-offshore gradient, with very high vulnerability near the coast and lower vulnerability offshore. The second pattern is 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 [48], the east [80] and west [79] coasts of the United States, and Norway [49]. This gradient is likely partly explained by seabird behavior during the breeding season. During this period, seabirds are 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). 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 the distribution map).

Other studies assessing seabird vulnerability to wind farms did not directly highlight the productivity gradient. Although highly productive areas are known to be seabird hotspots [102, 103], and Goodale et al. (2014) [80] emphasized the importance of excluding seabird hotspots from wind farm infrastructure projects. In these areas, habitat loss would significantly impact seabird populations as estuaries correspond to critical areas for them. Collision mortality is also expected to be higher in these areas 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 explains why zone 2, one of the zones that might receive the first wind farm, exhibited a really high relative vulnerability (Figure 5).

Avoiding coastal and productive areas is good general advice, supported by studies on seabird vulnerability to wind farms [80, 49], but this simplification may mask variation in spatial vulnerability between species. Moving wind farms away from the coast reduces impacts on seabirds as a group. Yet, when looking at the different seabird guilds, the positive effects of moving wind farms away from the coast may be more nuanced [80]. For instance, wind farms constructed far from shore would pose a greater risk to

pelagic species [80].

In our study, pelagic species includes Yelkouan shearwater (*Puffinus yelkouan*), Balearic shearwater (*Puffinus mauretanicus*), Scopoli's shearwater (*Calonectris diomedea*), Little gull (*Hydrocoloeus minutus*), European Storm petrel (*Hydrobates pelagicus melitensis*) and Atlantic puffin (*Fratercula arctica*). Shearwaters, storm petrels, and Atlantic puffins fly close to the water surface and spend much of their time below the rotor blade height [104, 98], thus having a low collision risk. In addition, shearwaters and storm petrels exhibit a low displacement risk [23, 48]. In comparison, Little gulls face a low risk of displacement but a high risk of collision [48] because they spend a significant amount of time flying, with about 15% of that time at rotor blade height (Appendix B, Table 4). Little gull is one of the most abundant seabird species in winter, and its distribution overlaps with the future wind farm zones (Figure 2). The potential additional mortality caused by wind farms raises significant concern for this species.

The risk posed by future wind farms from the currently projected zones 1 and 2 should be lower for coastal than pelagic species. However, the risk persists, as many coastal species still use the projected wind farm areas, albeit much less than coastal areas. Furthermore, post-construction studies in the North Sea show that some species do not avoid wind farms and even find them attractive [27, 28]. This attraction stems from opportunities to roost on above-water structures [28] and the higher abundance of prey. For instance, 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-bottomed areas [25], altering the seabed community and attracting more fish. This phenomenon is known as the 'reef effect' [25]. Furthermore, the benefits of excluding trawling, allowing the soft-bottom community to recover [105], should also act positively on fish abundance. Whether these changes will result in long-term beneficial effects for seabird populations remains unknown. Additionally, attraction to wind farms may increase 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 gull species [28, 27], one could hypothesize that the widespread Yellow-legged gulls would find wind farms attractive and may thus suffer increased collision risk.

In contrast to birds attracted by wind farms, species showing a strong avoidance behavior [27] often exhibit a low collision risk and vice versa [79]. 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 [49]. In our case, using two scores and taking the highest one did not significantly alter the relative vulnerability maps.

Synthesizing multiple seabird distribution maps with vulnerability scores involves substantial uncertainty and biases [106]. Uncer-

tainty arises from both the distribution maps and the vulnerability scores. Our integrated approach of combining multiple data sources should minimize uncertainty in the distribution maps [38, 34]. However, predicted maps have associated uncertainty not transferred to the spatial vulnerability. One method to account for this uncertainty could have been using the highest estimated space-use intensity within the 95% confidence interval. The second source of variability lies in the vulnerability scores [106]. The reliability of these scores depends on the methods used to determine the values of each criterion and the amount of supporting literature [106]. For instance, uncertainty in 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 [106]. There is a lack of research on flight altitude for Mediterranean species using such data, and available GPS-tracking data will be used to fill this gap [104, 107]. Therefore, it is important to remember that vulnerability scores and maps could evolve as more knowledge is gained.

In addition to wind farms, seabirds face numerous other threats at sea, such as episodic pollution incidents, fisheries bycatch, fish stock depletion, and disturbances from recreational activities. These threats can exert additive or compensatory effects when occurring in the same area [108]. Incorporating all these diverse threats into a cumulative impact assessment [109, 110] would improve recommendations for wind farm planning. The impact of wind farms may differ between two areas with similar seabird vulner-

ability if other threats already exist. In addition to local threats, climate change alters marine ecosystems, affecting seabird distribution and abundance. Therefore, current assessments remain only valid for a limited time. While current assessments consider seabirds, offshore wind farms can also endanger migratory birds (e.g., Geese [111]), as well as a diverse array of species, including bats, marine mammals, and fish [112].

## 5 Conclusion

Seabird vulnerability in the Gulf of Lion increases near the coast and in the Camargue region, where the Rhone River and other factors drive a high primary productivity. As a result, the proposed zones for the first wind farms fall in highly vulnerable areas for seabirds. A 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 including a zero-inflated structure. Integrating telemetry with count data is promising, especially if other species, such as the Mediterranean gull, are equipped. Telemetry data should also be used to determine the flight height of the species already equipped. Adding other anthropogenic threats to the spatial assessment would be the logical continuation 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).
- [2] 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).
- [3] 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).
- [4] 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).
- [5] Marion Frayssé 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).
- [6] 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).
- [7] 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).
- [8] D. Bănară 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).
- [9] Carles Carboneras. “Seabirds in the Gulf of Lions shelf and slope area” (2013).
- [10] 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).
- [11] 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-031R2.1](https://doi.org/10.1644/06-MAMM-A-031R2.1).
- [12] 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).
- [13] 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).
- [14] 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).
- [15] 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).
- [16] Marina Codina-García 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).
- [17] 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).
- [18] Stefan Garthe and Ommo Hüppop. “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).
- [19] Ministère de la transition écologique. *Projets de parcs éoliens flottants et de leurs raccordements en Mer Méditerranée. Synthèse du dossier de saisine de la Commission nationale du débat public*. 2020.
- [20] *Paris agreement*. Dec. 12, 2015.
- [21] 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.



- [22] 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: 978-1-78427-131-2.
- [23] 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).
- [24] Göran Broström 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.
- [25] 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.
- [26] 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).
- [27] 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).
- [28] 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/s10750-014-2088-x](https://doi.org/10.1007/s10750-014-2088-x).
- [29] Martin R. Perrow. "Wildlife and Wind Farms, Conflicts and Solutions. Vol. 3 Offshore: Potential effects." Pelagic Publishing, Exeter, UK, Sept. 5, 2019. ISBN: 978-1-78427-131-2.
- [30] 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).
- [31] BOURDEIX Jean-Hervé. "PELMED - PELAGIQUES MEDITERRANÉE" (1985). DOI: [10.18142/19](https://doi.org/10.18142/19).
- [32] 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[2248:ESORWD]2.0.CO;2).
- [33] J. Andrew Royle. "N-Mixture Models for Estimating Population Size from Spatially Replicated Counts". *Biometrics* (2004). DOI: [10.1111/j.0006-341X.2004.00142.x](https://doi.org/10.1111/j.0006-341X.2004.00142.x).
- [34] "Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS". Ed. by Marc Kéry and J. Andrew Royle. Academic Press, Jan. 1, 2020. ISBN: 978-0-12-823768-7. DOI: [10.1016/B978-0-12-809585-0.12001-0](https://doi.org/10.1016/B978-0-12-809585-0.12001-0).
- [35] 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).
- [36] 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).
- [37] 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).
- [38] 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).
- [39] 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).
- [40] 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).
- [41] 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).
- [42] 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).
- [43] Laurentiu Rozyłowicz 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).

- [44] 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).
- [45] 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).
- [46] Clara Péron 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).
- [47] Valentin Lauret, Nicolas Courbin, and Aurélien Besnard. “Comparing and integrating movement and counts data to estimate animal habitat selection” ().
- [48] 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).
- [49] 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.2024.1335224](https://doi.org/10.3389/fmars.2024.1335224).
- [50] 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).
- [51] Grégoire 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).
- [52] L. L. Eberhardt. “Transect Methods for Population Studies”. *The Journal of Wildlife Management* (1978). DOI: [10.2307/3800685](https://doi.org/10.2307/3800685).
- [53] 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).
- [54] 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).
- [55] 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).
- [56] 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).
- [57] 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).
- [58] Juan Serratos 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.2019.00838](https://doi.org/10.3389/fmars.2019.00838).
- [59] 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.2022.105860](https://doi.org/10.1016/j.marenvres.2022.105860).
- [60] 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).
- [61] Francisco V. Dénes, Luís Fábio 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: [10.1111/2041-210X.12333](https://doi.org/10.1111/2041-210X.12333).
- [62] 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).
- [63] José 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).

- [64] Marc Kéry. “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).
- [65] 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).
- [66] Maria P. Dias, José 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).
- [67] 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).
- [68] 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).
- [69] 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).
- [70] 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).
- [71] Sumio Watanabe. “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory” (2010).
- [72] 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).
- [73] Jérôme Guélat and Marc Kéry. “Effects of spatial autocorrelation and imperfect detection on species distribution models”. *Methods in Ecology and Evolution* (2018). DOI: [10.1111/2041-210X.12983](https://doi.org/10.1111/2041-210X.12983).
- [74] 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/B978-0-12-407197-1.00023-5](https://doi.org/10.1016/B978-0-12-407197-1.00023-5).
- [75] 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).
- [76] 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).
- [77] Frank A. Fogarty and Erica Fleishman. “Bias in estimated breeding-bird abundance from closure-assumption violations”. *Ecological Indicators* (Nov. 1, 2021). DOI: [10.1016/j.ecolind.2021.108170](https://doi.org/10.1016/j.ecolind.2021.108170).
- [78] OFB and GISOM. *Identification et priorisation de la responsabilité de chaque sous région marine pour les enjeux ornithologiques*. 2020.
- [79] 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.2018.08.051](https://doi.org/10.1016/j.jenvman.2018.08.051).
- [80] 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).
- [81] 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).
- [82] Jacob F. Pollock. “Detecting Population Declines over Large Areas with Presence–Absence, 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).
- [83] 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).
- [84] 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).

- [85] 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).
- [86] 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.2010.0087](https://doi.org/10.1098/rstb.2010.0087).
- [87] 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).
- [88] 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).
- [89] 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).
- [90] 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).
- [91] 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).
- [92] 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).
- [93] 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).
- [94] 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).
- [95] 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).
- [96] 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.2016.11.005](https://doi.org/10.1016/j.seares.2016.11.005).
- [97] 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).
- [98] Nicolas Courbin et al. "Flight heights of Scopoli's shearwaters *Calonectris diomedea* in the context of offshore wind farm developments" (May 2023). DOI: [10.1101/2023.05.14.540698](https://doi.org/10.1101/2023.05.14.540698).
- [99] 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).
- [100] 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).
- [101] Elise F. Zipkin et al. "Integrating count and detection-nondetection data to model population dynamics". *Ecology* (2017). DOI: [10.1002/ecy.1831](https://doi.org/10.1002/ecy.1831).
- [102] 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).
- [103] 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).
- [104] 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).

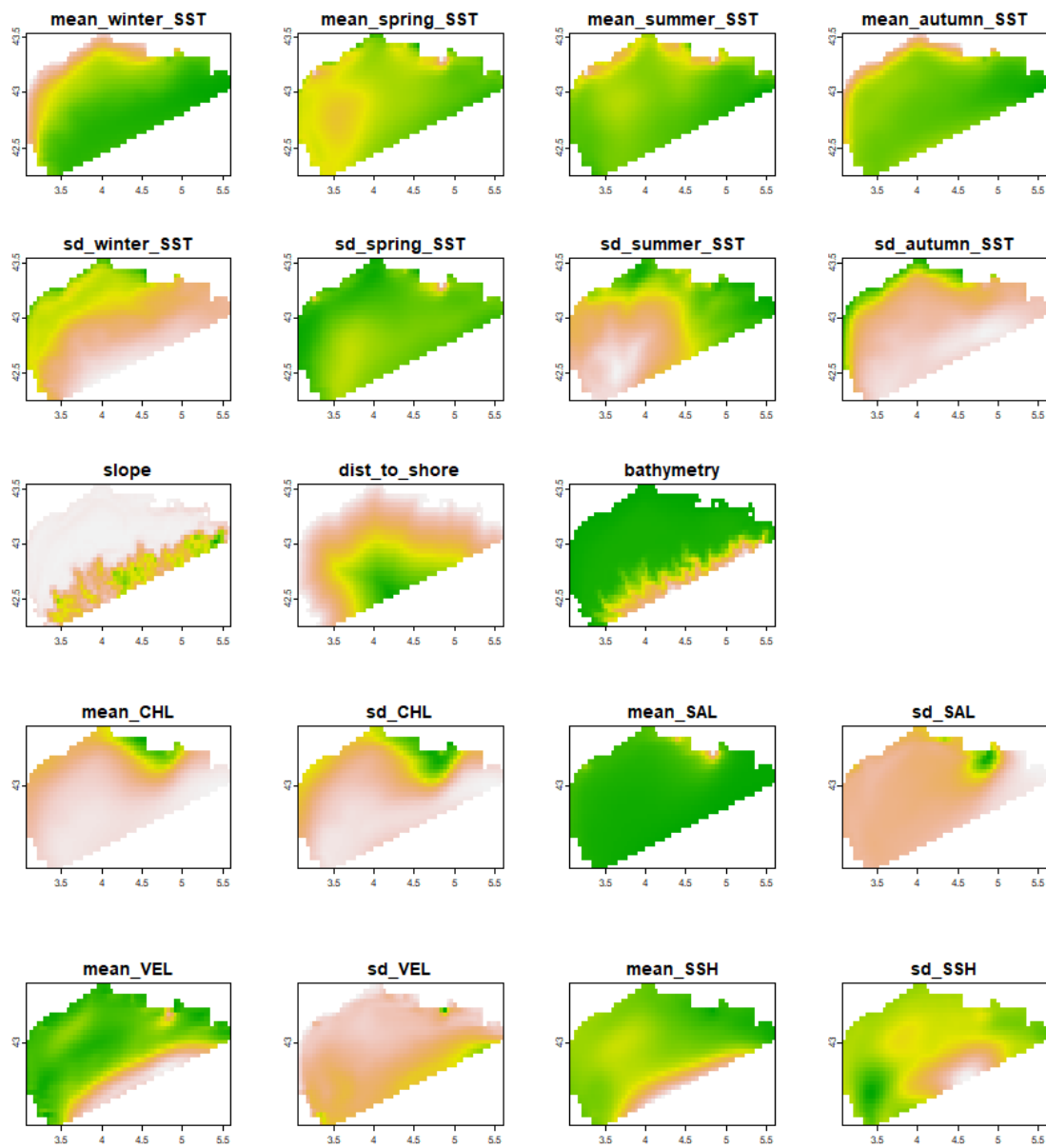


- [105] Gerard C. A. Duineveld, Magda J. N. Bergman, and Marc S. S. Lavaleye. “Effects of an area closed to fisheries on the composition of the benthic fauna in the southern North Sea”. *ICES Journal of Marine Science* (July 1, 2007). DOI: [10.1093/icesjms/fsm029](https://doi.org/10.1093/icesjms/fsm029).
- [106] 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).
- [107] Viola H. Ross-Smith et al. “Modelling flight heights of lesser black-backed gulls and great skuas from GPS: a Bayesian approach”. *Journal of Applied Ecology* (2016). DOI: [10.1111/1365-2664.12760](https://doi.org/10.1111/1365-2664.12760).
- [108] Michael Elliott, Angel Borja, and Roland Cormier. “Activity-footprints, pressures-footprints and effects-footprints – Walking the pathway to determining and managing human impacts in the sea”. *Marine Pollution Bulletin* (June 1, 2020). DOI: [10.1016/j.marpolbul.2020.111201](https://doi.org/10.1016/j.marpolbul.2020.111201).
- [109] 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).
- [110] 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).
- [111] Emily L. Weiser et al. “Geese migrating over the Pacific Ocean select altitudes coinciding with offshore wind turbine blades”. *Journal of Applied Ecology* (2024). DOI: [10.1111/1365-2664.14612](https://doi.org/10.1111/1365-2664.14612).
- [112] Martin R. Perrow. “Wildlife and Wind Farms - Conflicts and Solutions, Volume 4. Offshore: Monitoring and Mitigation.” Pelagic Publishing, Exeter, UK, Sept. 5, 2019.



## A Appendix A

Fig.6: Environmental covariates.



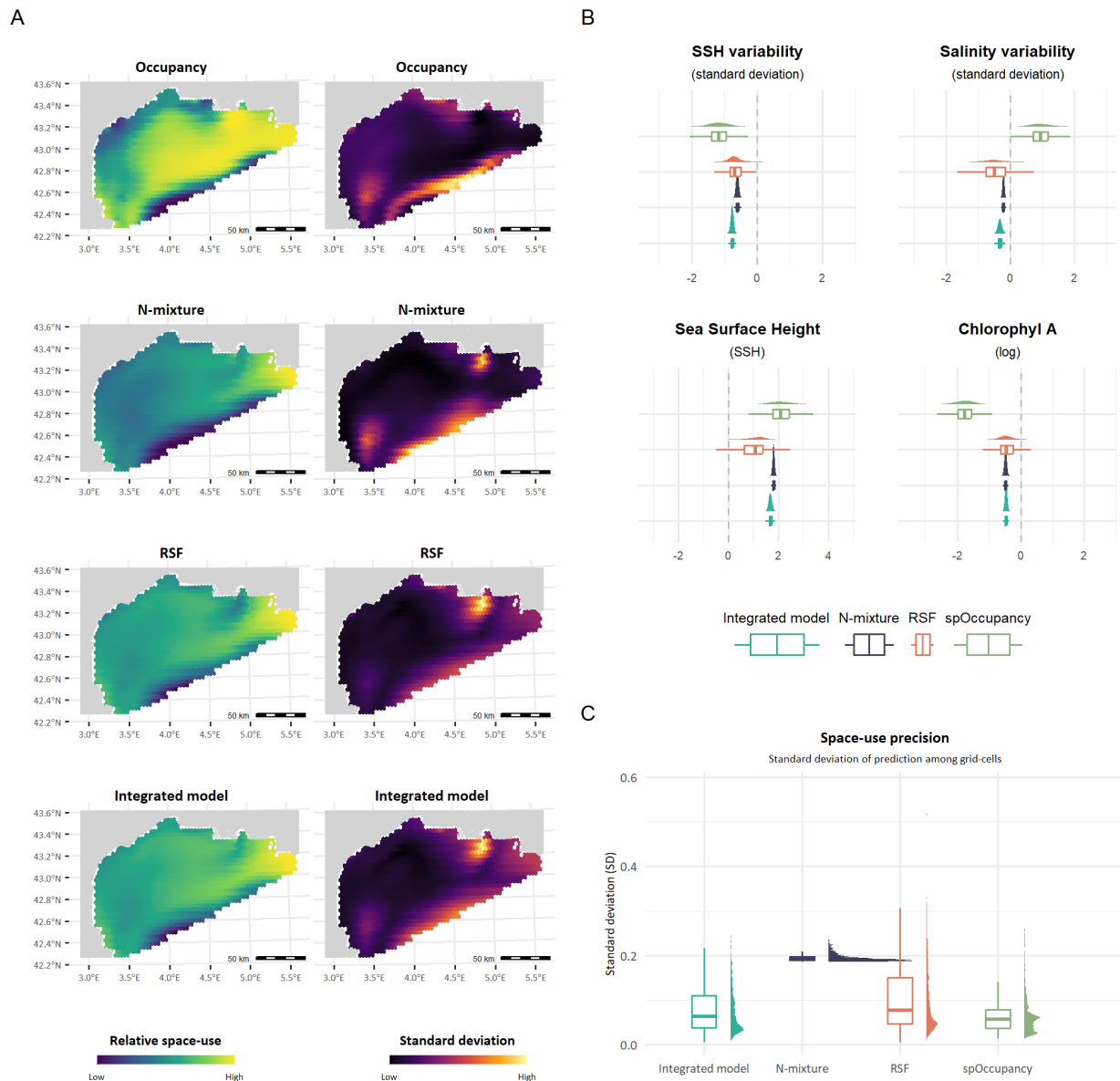
## B Appendix B

**Table 4: Values of the criteria used to compute vulnerability index.** **Alt:** percentage of time flying at blade height. **Man:** flight manoeuvrability score (/5). **Ptf:** score of the proportion of time flying. **Noc:** nocturnal activity score. **Dis:** score of disturbance sensibility to wind farms and surrounding activities. **Spe:** score of specialization in the choice of feeding areas.

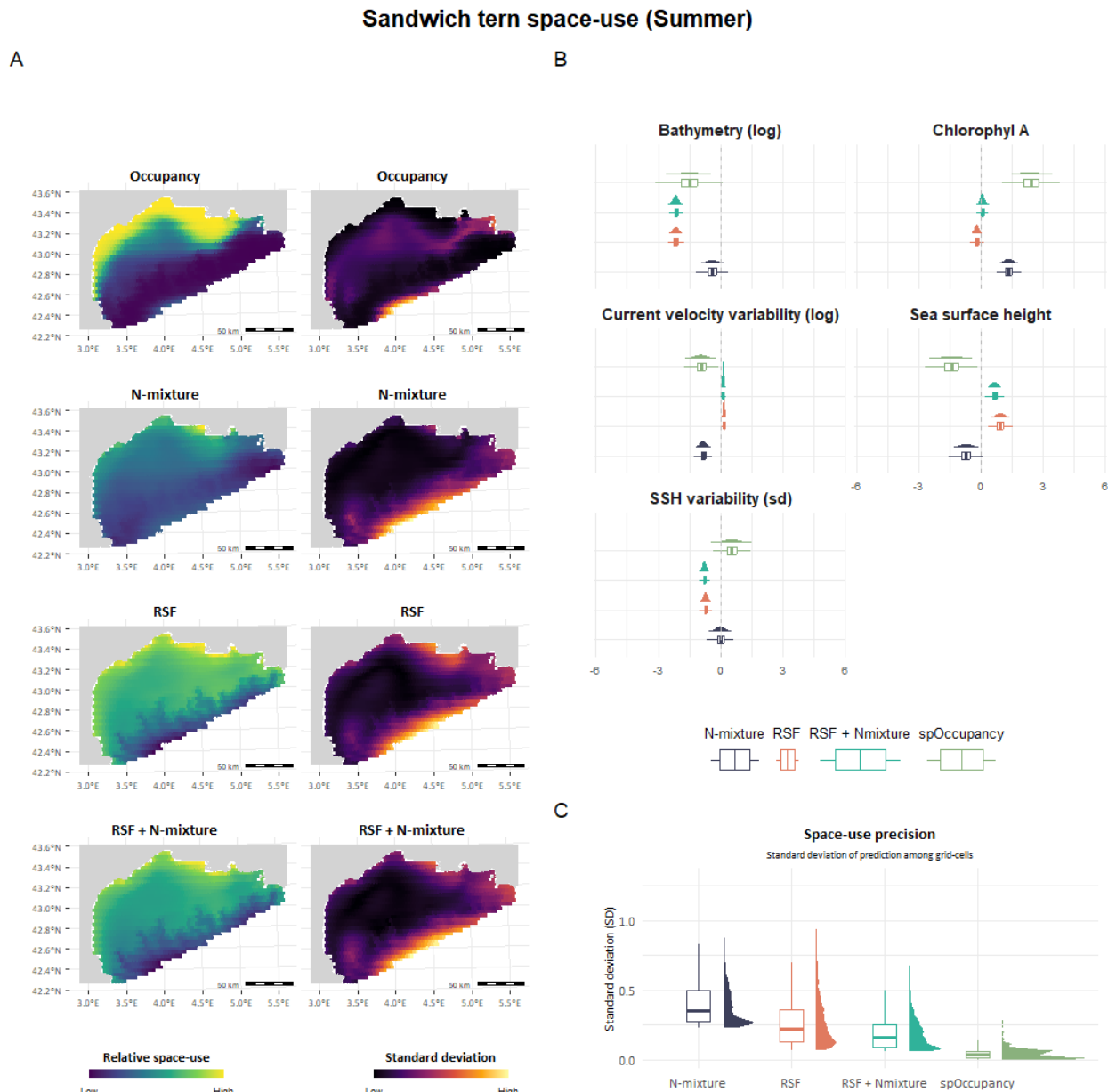
Scientific names	Common names	Alt	Man	Ptf	Noc	Dis	Spe	Conservation score	
								Summer	Winter
<i>Calonectris diomedea</i>	Scopoli shearwater	0	3	3	3	1	1	5.3	
<i>Chroicocephalus ridibundus</i>	Black-headed gull	20	1	1	2	2	2	3.3	1.3
<i>Fratercula arctica</i>	Atlantic puffin	0.5	3	1	1	2	3		4.3
<i>Hydrocoloeus minutus</i>	Little gull	15	1	3	2	1	3		5.5
<i>Hydrobates pelagicus melitensis</i>	European storm petrel	2	1	3	4	1	1	7.8	
<i>Larus melanocephalus</i>	Mediterranean gull	25	1	2	2	2	2	2.3	2.8
<i>Larus michahellis</i>	Yellow-legged gull	35	2	2	3	2	1	3.5	2.8
<i>Morus bassanus</i>	Northern gannet	12	3	3	2	2	1		1.8
<i>Puffinus mauretanicus</i> & <i>Puffinus yelkouan</i>	Small shearwaters	0	3	3	3	1	1	6.5	7
<i>Stercorarius pomarinus</i> & <i>Stercorarius parasiticus</i>	Small skuas	10	1	5	1	1	2		3
<i>Sterna hirundo</i>	Common tern	10	1	5	1	2	3	2.5	2
<i>Sterna sandvicensis</i>	Sandwich tern	10	1	5	1	2	3	1.5	

## C Appendix C

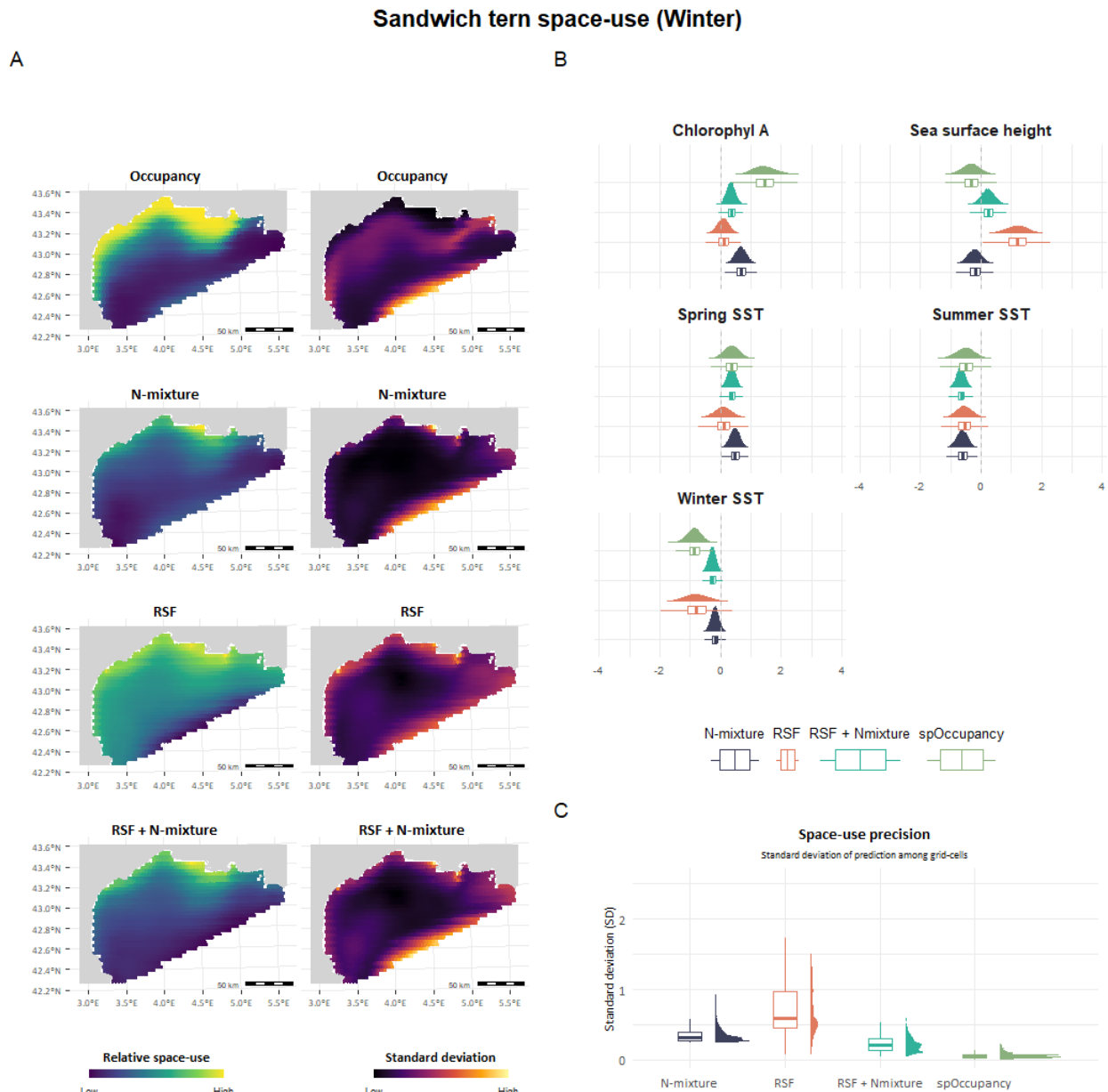
### Scopoli shearwater space-use



**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.

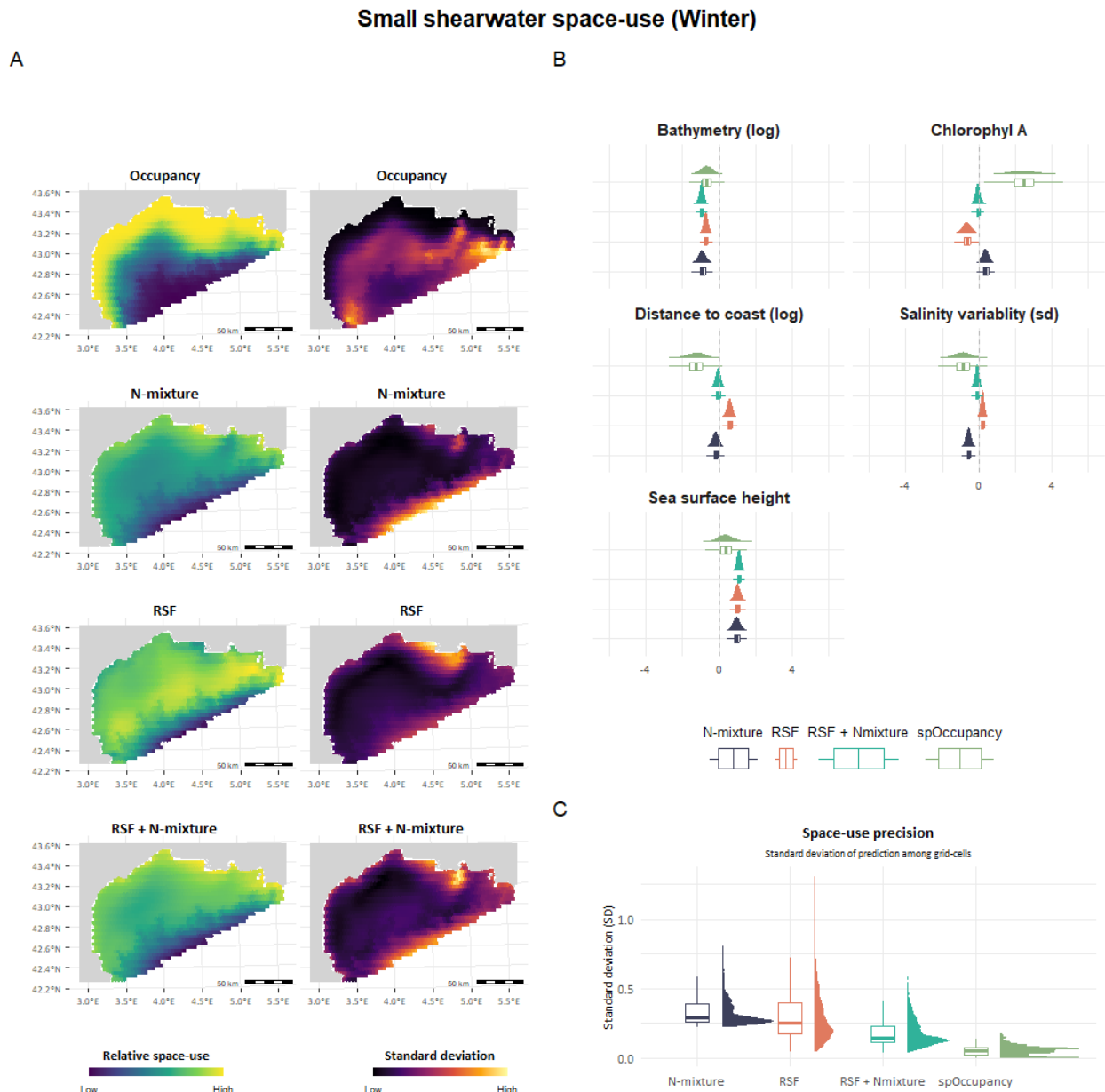


**Fig.8: Sandwich tern (*Sterna sandvicensis*) distribution in the Gulf of Lion.** Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining 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.

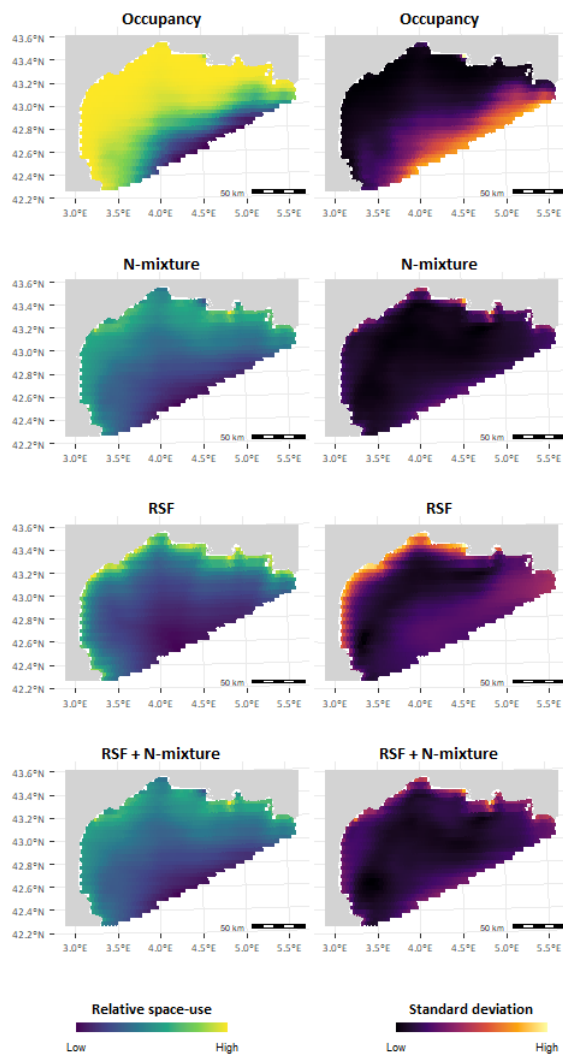




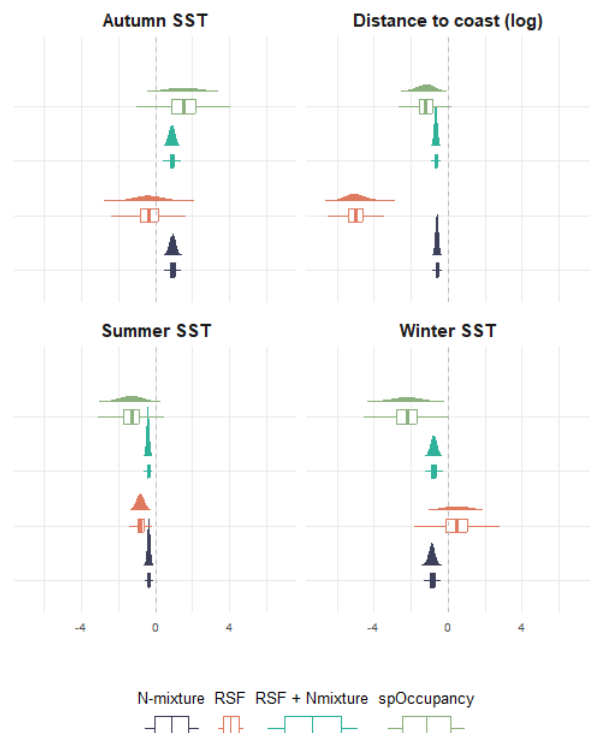
**Fig.10: Small shearwaters (*Puffinus* sp) winter distribution in the Gulf of Lion.** Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining 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.

## Yellow-legged gull space-use (Summer)

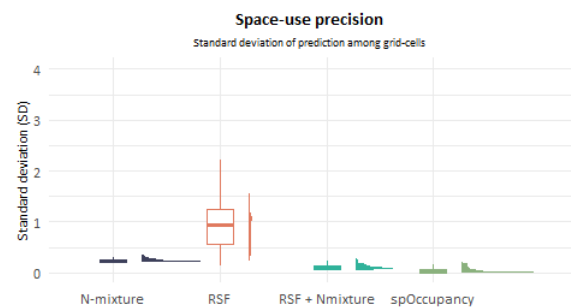
A



B



C



**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.

## D Appendix D

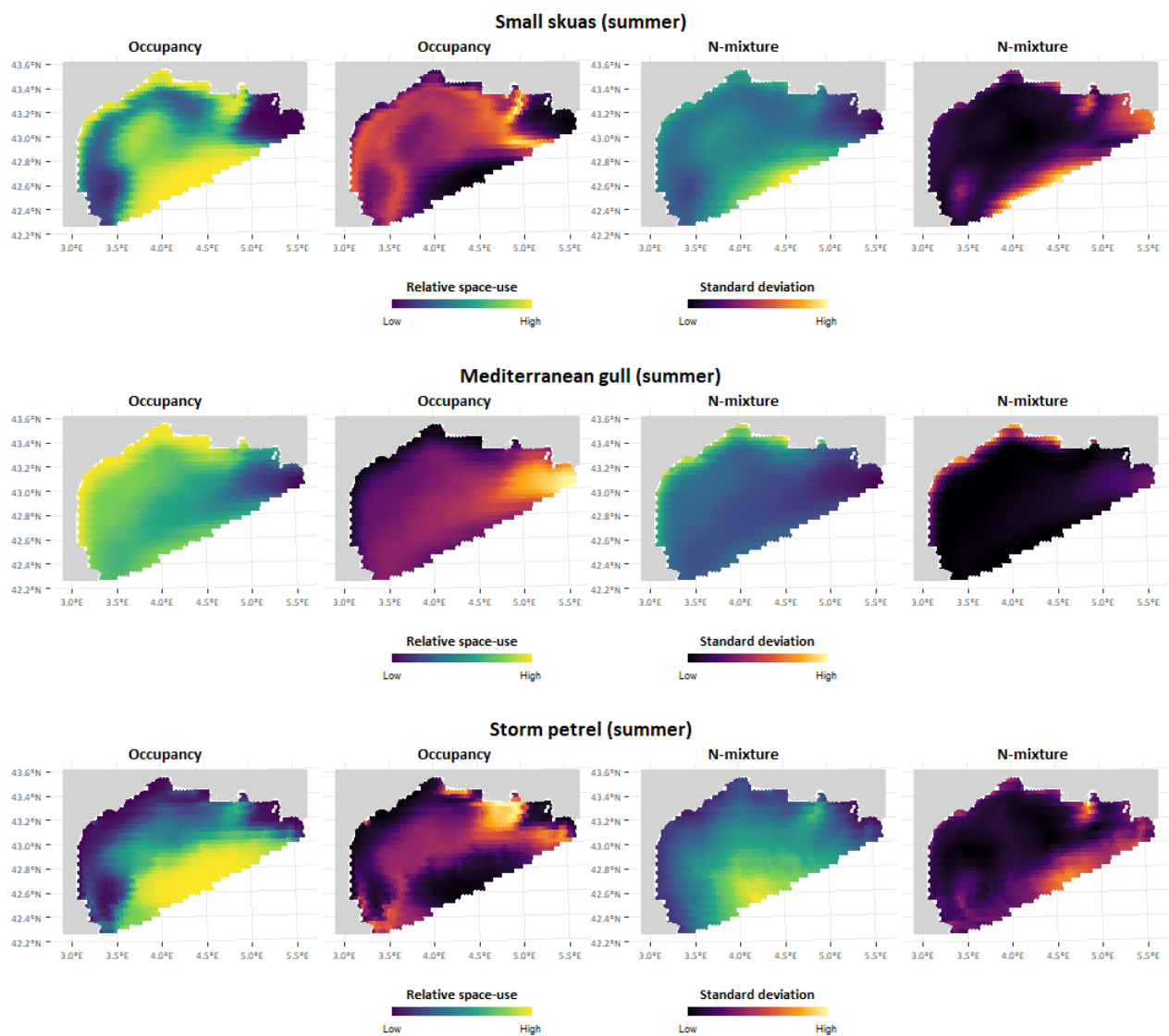


Fig.12: Summer distribution of small skuas, Mediterranean gull, and European Storm petrel.

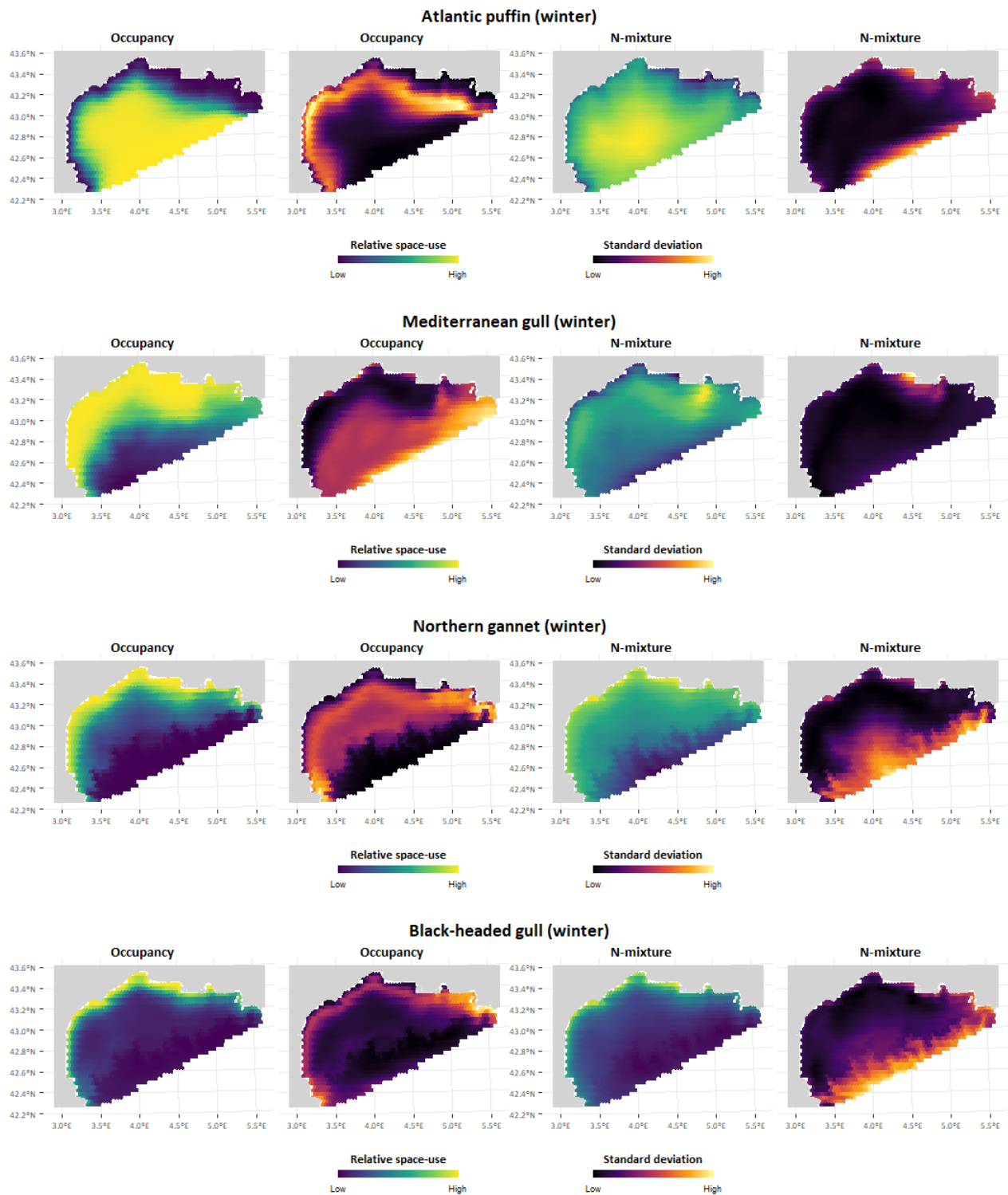


Fig.13: Winter distribution of Mediterranean gull, Atlantic puffin, Northern gannet, and Black-headed gull.

## Résumé

Pour faire face à la demande croissante en énergie décarbonée, la France développe des parcs éoliens en mer dans le golfe du Lion. Cette région maritime est essentielle pour les oiseaux marins, déjà confrontés à de multiples menaces. Il est donc nécessaire de procéder à une évaluation précise 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 oiseaux marins dans le golfe du Lion en combinant différentes sources de données. Ces sources comprennent quatre campagnes d'échantillonnage en mer et des données de télémétrie. En combinant ces données à l'aide de modèles intégrés, nous avons pu cartographier la distribution d'un large éventail d'espèce d'oiseaux marins. Nous avons comparé plusieurs approches de modélisation, en exploitant des données de détection/non-détection, de comptage et de télémétrie, 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 informations utiles dans le cadre de la planification éoliennes, la modélisation des données de comptage est l'approche à privilégier. La télémétrie est particulièrement intéressante et deviendra de plus en plus utile grâce à sa disponibilité croissante. L'utilisation conjointe de la télémétrie et des données de comptage représente une nouvelle approche prometteuse. 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 particulièrement vulnérables pour l'avifaune marine.

## Summary

To meet the growing demand for low-carbon energy, France is installing wind farms in the Gulf of Lion, located in the South of France, in the northwestern Mediterranean. This area is of great importance for seabirds, which are already impacted by numerous threats. Protecting Mediterranean seabirds is crucial due to their conservation status and endemism. Consequently, conducting a comprehensive assessment of seabird vulnerability to wind farms in this area becomes 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 unanalyzed 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 could estimate seabird distribution with great accuracy and target a broad range of species. We compared several modeling approaches, utilizing detection/non-detection, count, and telemetry data. Our findings suggest that while detection/non-detection data can provide accurate recommendations for wind farm planning, count data remains 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 in isolation. Our analysis reveals high seabird vulnerability in coastal areas and near the Camargue estuary. Therefore, the zones planned for the first wind farms fall within highly vulnerable locations.