|  |  |  |  |
| --- | --- | --- | --- |
| **Contents**  **1 Introduction** | | | **2** |
| **2** | **Methods** |  | **5** |
|  | 2.1 Study area and data collection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | 5 |
|  | 2.1.1 Study area . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | 5 |
|  | 2.1.2 Seabirds at-sea counts . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | 6 |
|  | 2.1.3 Telemetry data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | 6 |
|  | 2.1.4 Environmental data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | 7 |
|  | 2.2 Statistical analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | 8 |
|  | 2.2.1 Seabirds occupancy models with detection/non-detection data . . . . . . . . . | | 8 |
|  | 2.2.2 Seabirds abundance modeling with count data . . . . . . . . . . . . . . . . . . | | 9 |
|  | 2.2.3 Resource Selection Function space-use modeling based on Telemetry GPS data . . . . | | 10 |
|  | 2.2.4 Relative space use models integrating N-mixture and RSF models . . . . . . . | | 11 |
|  | 2.2.5 Implementation and model fitting . . . . . . . . . . . . . . . . . . . . . . . . . | | 11 |
|  | 2.2.6 Model selection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | 12 |
|  | 2.4 Mapping seabirds’ vulnerability to wind farms . . . . . . . . . . . . . . . . . . . . . . | | 13 |
|  | 2.4.1 Vulnerability indexes . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | 13 |
|  | 2.4.2 Vulnerability maps . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | | 14 |
| **3** | **Results** | | **15** |
|  | 3.1 Comparison of seabird occurrence and abundance modeling . . . . . . . . . . . . . . . | | 15 |
|  | 3.2 Comparison and integration of movement data with count . . . . . . . . . . . . . . . | | 18 |
|  | 3.3 Seabird spatial vulnerability to wind farm . . . . . . . . . . . . . . . . . . . . . . . . | | 19 |
| **I** | **Appendix A** | | **27** |
| **II Appendix B** | | | **28** |

# 1 Introduction

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

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

In addition to existing threats, new human activities are planned in the Gulf of Lion, including offshore wind farm projects [19]. While offshore wind energy is expected to significantly contribute to renewable energy development, its impact on seabird populations is a major environmental concern [18, 20, 21, 22]. Offshore wind farms affect seabirds through direct collision mortality with wind turbines [20, 22] and through population displacement due to disturbance created by the wind turbines and associated activities [21, 22]. Furthermore, wind farms alter the local ecosystem by changing wind and water circulation [23], transforming seabed communities [24], and restricting fishing activities [25]. These changes often make the area less attractive to seabirds, but in some cases, resources can become more abundant, making the area more appealing to certain species [26, 27]. However, while this increase in wind farm use could provide new feeding opportunities for seabirds and in turn improve their population dynamics, it also increases the risk of collisions [27].

To effectively mitigate the future impact of wind farms on seabird populations in the Gulf of Lion, a spatial assessment of seabirds’ vulnerability to these installations is crucial [28]. This requires understanding the spatial distribution of seabird species within the assessment area. However, a comprehensive assessment of seabird distribution at the scale of the Gulf of Lion has yet not been conducted. Previous studies have estimated seabird distribution in the French Mediterranean Exclusive Economic Zone (EEZ) [3, 29]. Still, these studies were either not conducted at the species level [3] or focused on only a few seabird species [29]. To be comprehensive, the assessment of species distribution should encompass a broader range of species. In the Gulf of Lion, several programs have collected data on seabirds from 2011 to 2023 [3, 30]. Most of the data consists of counts collected during boat or aerial surveys, each conducted with different protocols. These protocols vary, for example, in the number of observers and the counting platforms used. The probability of detecting all species and all individuals of a species at each site depends on the protocol used and differs for each survey. Besides, for a few species GPS-tracking data is now also available. Combining all these data will allow us to have a comprehensive and representative assessment of seabird distribution, while current assessments rely on single dataset approaches, hence discarding possibly important information.

The most reliable method to combine several data sources is to develop so-called “integrated species distribution models” [31, 32, 33]. Integrated species distribution models allow to take these differences into account [31, 32, 33]. Integrated models present several major advantages, with the primary one being enhanced accuracy in estimating ecological parameters of interest [34, 35]. 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 [36, 33].

Depending on the available data and the specific ecological question, different classes of species distribution models can be employed. Occupancy models [35, 37, 38] are suitable to use for detection/nondetection data, or if the ecological focus is on the presence or absence of a species at a site. These models provide information on the occurrence areas of a species. When count data are available and the ecological focus is on the abundance of a species at each site, N-mixture models can be utilized [35, 39, 40]. Abundance can also be modeled with distance-sampling or capture-recapture methods, but our data lack the necessary information for these approaches, observation distance, or individual identification, respectively. To guide public policy or understand species distribution, occupancy models are more widely used than N-mixture models [41], mainly because detection/non-detection data are more abundant. While N-mixture models can provide more detailed information, occupancy models might be sufficient in some cases [42] and could even serve as a surrogate for abundance estimation [43], particularly for rare species with low density [42]. However, this is less likely to be the case for seabirds. Additionally, occupancy models are often easier to fit [42]. Few studies have compared these two methods directly [42]. In the context of wind farm planning, it would be interesting to know if the use of detection/non-detection data could be sufficient.

Apart from count data, spatial distribution can also be determined from positions acquired by GPS

(Global Positioning System) devices [44]. As the quantity and variety of data increase, there’s a growing trend of using diverse data types to model species distribution [33], but employing a model-based approach to combine disparate data is not yet common practice. 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 also provide insights into species distribution [44].

Count and GPS data capture complementary aspects of animal space-use [1]. Count data provide broad-scale population-level information, while GPS data offer detailed, fine-scale information for a few individuals. Several studies have attempted to combine information from GPS and count data to determine species distribution, particularly in seabird research [1, 46, 47]. However, none of these studies utilized a model-based approach. Lauret et al. (under review) [48] developed a method to combine telemetry and count data into a single model. This model is hierarchical, incorporating a component for count data using an N-mixture model and a component for movement data through a

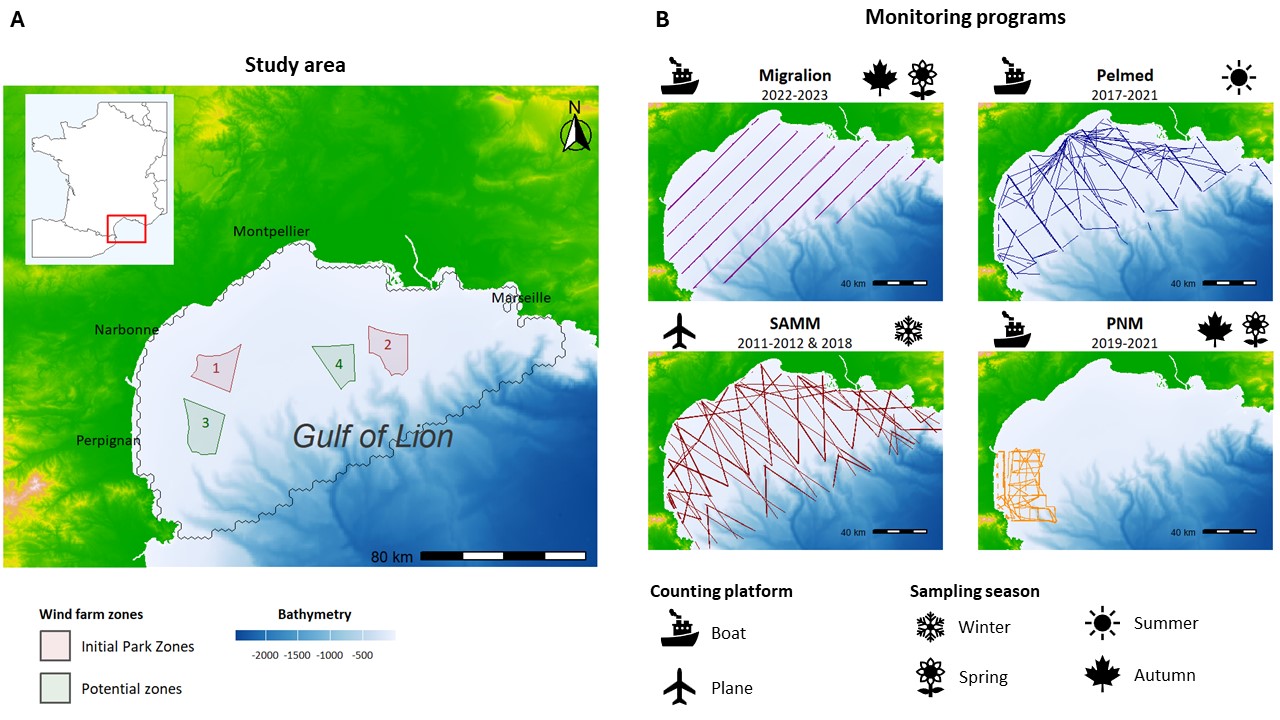
Resource Selection Function (RSF), the most commonly used method in habitat selection analysis [45].

To complement distribution maps that describe fine patterns of risk for each seabird species, 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 [27, 20, 49]. To address species-specific differences, we calculated a vulnerability index for each species using the approach developed by Furness et al. (2013) [22] and Bradbury et al. (2014) [49]. This index considers the conservation status of the species and metrics determining collision and habitat loss risk [22]. 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 [47, 49].

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

# 2 Methods

## 2.1 Study area and data collection 2.1.1 Study area



**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 which need further investigations. The black line delimits the study area and corresponds to the continental shelf. **(B)** The transects followed by each monitoring program to visually count seabirds. Migralion program was performed every year during April and September from 2022 and 2023. PELMED (PELagic MEDiterranean) was conducted every year in July from 2017 to 2021. SAMM (Survol A´erien de la M´egafaune 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. It’s important to note that not every transect was sampled during each campaign.

The Gulf of Lion is located in the north-western Mediterranean Sea (Fig.1). It is one of the most productive areas in the Mediterranean Sea, due to a combination of freshwater input by the Rhone River [4, 5] and small-scale upwellings created by winds [50]. The study area was determined based on suitability to receive wind farms, which is mainly driven by the bathymetry. In consequence, the study area coincides with the continental shelf. The total area considered for the study had a surface of approximately 17,300 km². Within this area, assessment units were defined by 1380 hexagonal cells with a size of 12,6 km². All the ecological parameters are estimated at the scale of the grid cells, thus each grid cell defines a ’site’. Grid resolution was chosen to have a ratio of approximately 1 cell with detection for

10 cells without any detection on average for each species.

### 2.1.2 Seabirds at-sea counts

Count data were collected from four different programs, as summarized in **??**. Observations were conducted using strip-transect methodology (Fig.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].

SAMM (Survol A´erien de la M´egafaune Marine, Aerial Census of Marine Megafauna) [29, 54], was the only aerial survey. It was conducted from November to February 2011-2012 and from January to February 2019, covering the entirety of the French Exclusive Economic Zone. Although this aerial survey also occurred during the summer of 2012, it was not replicated and thus not used in our study, as our models require replicated data.

PELMED is an annual survey primarily conducted to assess small pelagic fish resources [30]. 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 was collected in 2022 and 2023, twice a year around April and September.

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

Seabirds distribution could change between seasons and some species are only present in the study area during a part of the year [29, 54]. For these reasons, our analysis was split into two periods: breeding season (from April to August) and non-breeding season (from September to March).

### 2.1.3 Telemetry data

GPS-tracking data within the Gulf of Lion was available for four species: Scopoli’s shearwater, Yelkouan shearwater, Yellow-legged gull, and Sandwich tern. For the Migralion project in 2021 and 2022, 51 Scopoli’s shearwaters, 39 Yelkouan shearwaters, and 22 Sandwich terns were equipped with GPS at their nesting colonies on Riou Island near Marseille, in Port-Cros, and at the Etang de Thau near Montpellier, respectively, all located in the South of France. Additionally, 106 Yellow-legged gulls were equipped with GPS in 2021 in the Natural Park of Camargue as part of ??? project.

We sampled recorded locations within the study area, and retained one location per hour to avoid

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

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

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.co and MARSPEC database [57] (http://www.marspec.org). A total of 8 geophysical and biotic covariates were included and are listed in Tab.1 (see also Appendix I Fig.6 for maps). Covariates choice was based on knowledge about seabird ecology and that are frequently used covariates for seabird distribution modeling [1, 29, 58, 59, 60].

The dynamic variables were obtained with monthly temporal resolution and averaged values were calculated from 2011 to 2023, covering the oldest to most recent seabird data available. We also computed the standard deviation for each covariate from the monthly temporal resolution and considered it as a measure of temporal variability of the covariate value. Sea Surface Temperature (SST) showed great seasonal variability, thus average values were calculated for each season. We extracted covariates at the spatial scale of each grid cell. When the correlation between two covariates was too high (*R*2 *>* 0*.*8), one of the covariates was excluded from the model selection procedure (see 2.2 statistical analysis). Were excluded in priority covariates highly correlated with more than one variable, and standard deviation before mean values.

## 2.2 Statistical analysis

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

We first fitted occupancy models [37] to estimate seabird distribution in the Gulf of Lion for each species. The fundamental concept of this modeling approach is to consider the occurrence of a species as a Bernoulli random variable where the success is the species presence, and the success probability is the presence probability *ψ* [35, 37].

### Latent process

Let’s consider J sites (i.e. hexagonal cells of the grid), each surveyed K times. We denote *Zj* the true occupancy state at site j, where *j* ∈ 1*,...,J*. *Zj* = 0 if the species is absent and *Zj* = 1 if it is present. This latent occurrence variable is assumed to follow a Bernoulli process [37]:

Une image contenant texte, Police, blanc, capture d’écran

Description générée automatiquement

Here, *ψj* represents the probability of species presence at site j. The presence probability is modeled by logistic regression of site-specific environmental covariates *Xj*, with *β* being a vector of regression coefficients [35, 37].

### 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 [35, 62, 63, 64].

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

*Yj,k,r* ∼ *Bernoulli*(*pj,k,rZj*)

(2) *logit*(*pj,k,r*) = *αr*0 + *αr*1*transect lengthj,k,r*

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

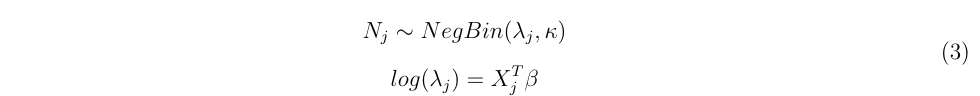
To ensure the detection probability could be accurately identified, multiple sampling occasions are needed. Counts were repeated for each dataset over different years. We considered these counts as identical replicates despite potential inter-annual variability. This decision was made to obtain a static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

### 2.2.2 Seabirds abundance modeling with count data

As a second step, we inferred seabird at-sea abundance using N-mixture models [40, 39]. Similarly to occupancy models, N-mixture models have a hierarchical structure where the observation is defined conditionally on the latent process describing the true abundance [40, 39].

### Latent process

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

Where *λj*, the mean of the distribution, was modeled as a function of site-specific covariates, and *κ* was the overdispersion parameter. A negative binomial distribution was preferred to a Poisson distribution to take overdispersion of the count data into account, the negative binomial allowing the standard deviation and the mean to be different (the variance is *λ* + *λ*2*/κ*). Overdispersion is frequent with seabirds counts as individuals aggregate in great numbers at the same place for many species.

### Observation process

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

*Nj,k,robs* ∼ *Binomial*(*pj,k,r,Nj*)

(4) *logit*(*pj,k,r*) = *αr*0 + *αr*1*transect lengthj,k,r*

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

**2.3 Resource Selection Function space-use modeling based on Telemetry**

## GPS data

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

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

travel.

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

*Kj* ∼ *Bernoulli*(*λj*)

(5) *logit*(*λj*) = *XjT β*

Where *λj* is the probability for a location being used knowing its environmental covariates values

(*λj* = *P*(*Kj* = 1|*Xj*)). *λj* is modeled as a logistic regression of location-specific covariates and regression coefficients *β*. Coefficients *β* represent the strength of the selection on the associated covariates.

To model potential population heterogeneity, we added a random effect on slope coefficients *β*. Let’s I denote the total number of individuals, and take an individual *i* ∈ [1;*I*], we have:

*βi* ∼ *Normal*(*β,τ*) (6)

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

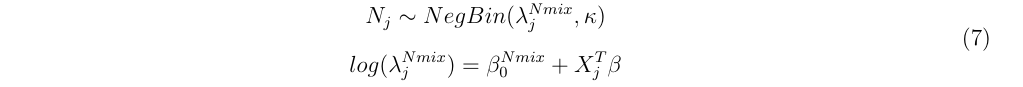
To complete the Bayesian specification of the model, normal priors are assigned to *β*, and uniform priors to *τ*.

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

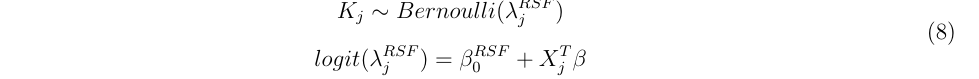
Lastly, we integrated GPS data with count data into a single model by combining an N-mixture model with a Resource Selection Function.

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 Inhomogeneous Poisson Point process (IPP), Lauret et al. (under review) [48] proposed to integrate these two data sources in a single model using IPP as a unifying framework. IPP describes the stochastic distribution of points in space when the expected density of points is correlated to environmental covariates, as in occupancy, N-mixture, and RSF models. IPP reflects the intensity *λ* of a spatial point process over a defined study area. N-mixture and RSF do not estimate the same IPP intensity *λ*. In RSF, the intercept has no biological meaning and only the relative habitat selection is estimated. IPP intensity *λ* being determined by slope coefficients *β*, coefficient estimates should be consistent between RSF and N-mixture. Then N-mixture and RSF estimate the same relative intensity of the IPP. The integrated model keeps the hierarchical structure of the N-mixture and the logistic regression of the RSF and jointly estimates *β* parameters.

The model is divided into the N-mixture submodel:



And the RSF submodel:



### 2.3.2 Implementation and model fitting

All models were implemented in R (ref). Occupancy models were fitted with the package spOccupancy

[65], allowing fast computation, and easy model selection. With this package, models are fit in a Bayesian framework. We ran 3 MCMC chains with 30,000 iterations each and a burn-in of 3,000

iterations. N-mixture, RSF, and the model integrating N-mixture and RSF were implemented with Nimble [66]. 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 Tab X in Appendix X). For all the models, convergence was checked by verifying that *R <* 1*.*1 [67].

Each method provides slightly different insights into seabird distribution. Occupancy models indicate the area of occurrence for a species, with the map values representing the probability of presence (*ψ*). N-mixture models aim to model abundance, with the map values showing the intensity of the inhomogeneous Poisson process (IPP) (*λ*). However, in our case, the assumption of closure between sites is not met. Consequently, occupancy and N-mixture models do not directly measure occurrence and abundance but rather provide a relative intensity of space use by seabirds. For the RSF, the intercept is not meaningful, and the output reflects relative habitat selection, indicating areas of higher or lower use. In the integrated RSF and N-mixture model, the output also represents relative space use, which is consistent across different data sources. Thus, we will use the term ”relative space-use” to describe the results obtained from each method, despite subtle differences in what each method specifically measures.

### 2.3.3 Model selection

The selection procedure was carried out exclusively using the occupancy model. For each species, the set of covariates used in the N-mixture, RSF, and the integrated N-mixture and RSF models was the same as that obtained through model selection on the occupancy models. This ensured that the results produced by each model were comparable.

To have the most accurate prediction on seabird distribution as possible to inform wind farm policies, the selection procedure aimed to maximize the predictive power of our models. For this purpose, we evaluated model performance using two criteria: the Widely Applicable Information Criterion (WAIC) [68] and cross-validation [69]. 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 [69]. When predictive scores were too close (*difference <* 3), we chose the model with fewer covariates.

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 and tested all possible combinations

Vulnerability index

Scientific name Common name

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

**Table 2: Species and their associated vulnerability index.** Blank space indicates no score for a species at a certain period due to its absence from the area.

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 [70]. The spOccupancy package offers various spatial covariance models (spherical, exponential, and Gaussian) [65], which we tested.

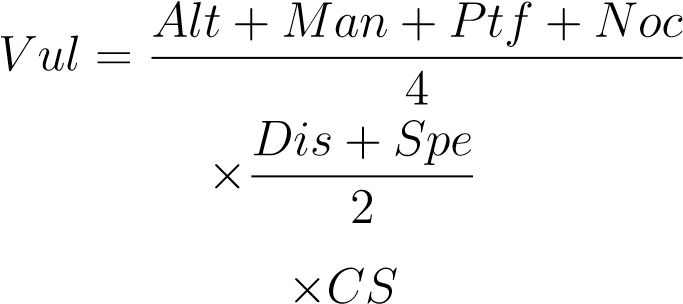
## 2.4 Mapping seabirds’ vulnerability to wind farms

### 2.4.1 Vulnerability indexes

Based on the results obtained with occupancy and N-mixture models, we constructed a synthetic index to evaluate species risk to offshore wind farms integrating three main aspects of seabird vulnerability [49, 22]:

1. **Conservation Status (CS):** Defined in [71], this score combines species status and representativeness of French populations. Species status is based on the worst IUCN Red List rankings or population trends. At the same time, representativeness is measured by the percentage of the global population in France and on the Mediterranean coast. Different representativeness values are defined for the breeding and non-breeding periods as seabird distribution can vary greatly between these periods [71, 3]. CS is the mean between representativeness and vulnerability.
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).

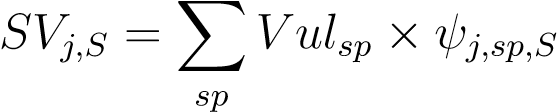
The vulnerability index was computed as follows [22]:



Vulnerability indexes for every species are indicated in Tab.2

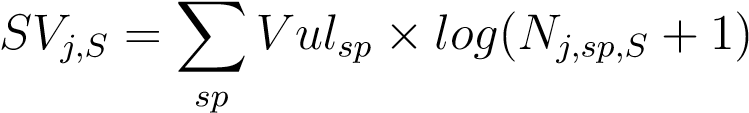
### 2.4.2 Vulnerability maps

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



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

With abundance maps, the seasonal seabird vulnerability *SV* at site *j* and season *S* was defined following Bradbury et al. (2014) [49]:



These calculations were performed for the breeding (B) and non-breeding (NB) seasons. The final vulnerability map was obtained by averaging the two seasonal maps at each site:

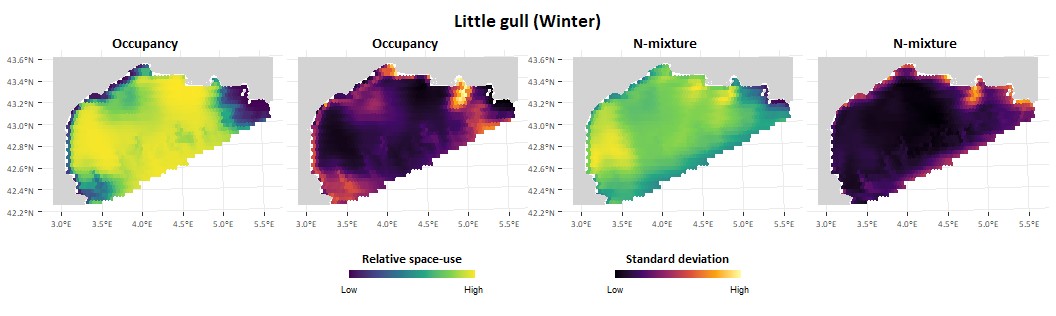
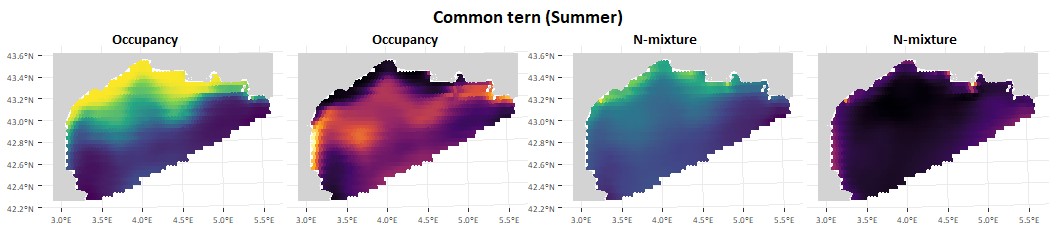
*SVtot* = mean(*SVB,SVNB*)

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

# 3 Results

All models were able to effectively differentiate between less used and most used areas, showing good concordance with observable patterns in the data. The models selected after the model selection process are summarized in Appendix X, Table X. Models that included spatial autocorrelation resulted in low predictive performance and overfitting, so the final models did not incorporate a spatial autocorrelation structure.

## 3.1 Comparison of seabird occurrence and abundance modeling

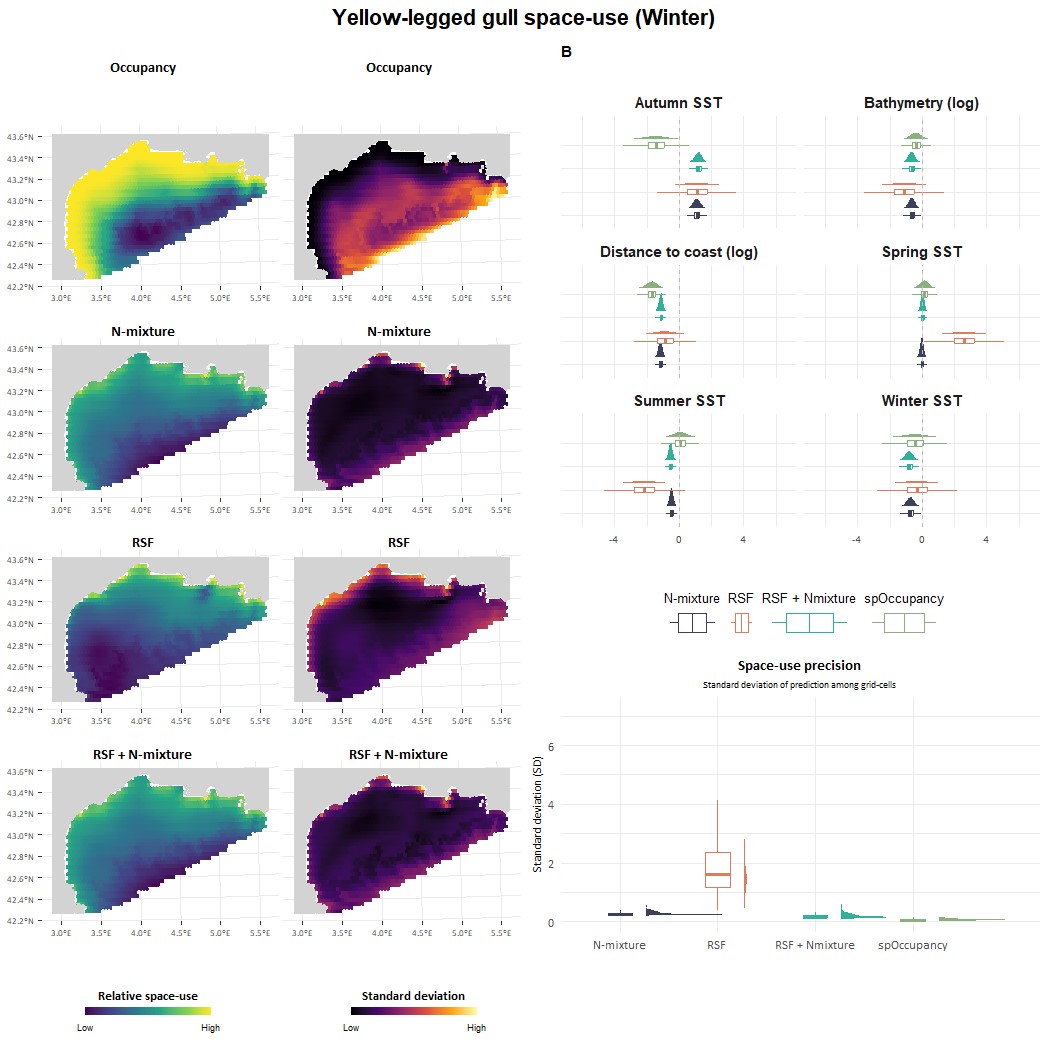


**Fig.2: Common tern (Sterna hirundo) and Little gull (Hydrocoloeutus minutus) distribution in the Gulf of Lion.** At-sea distribution estimated using an occupancy model with detection/non-detection data and an N-mixture model with count data. The figure shows the distribution maps and the associated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

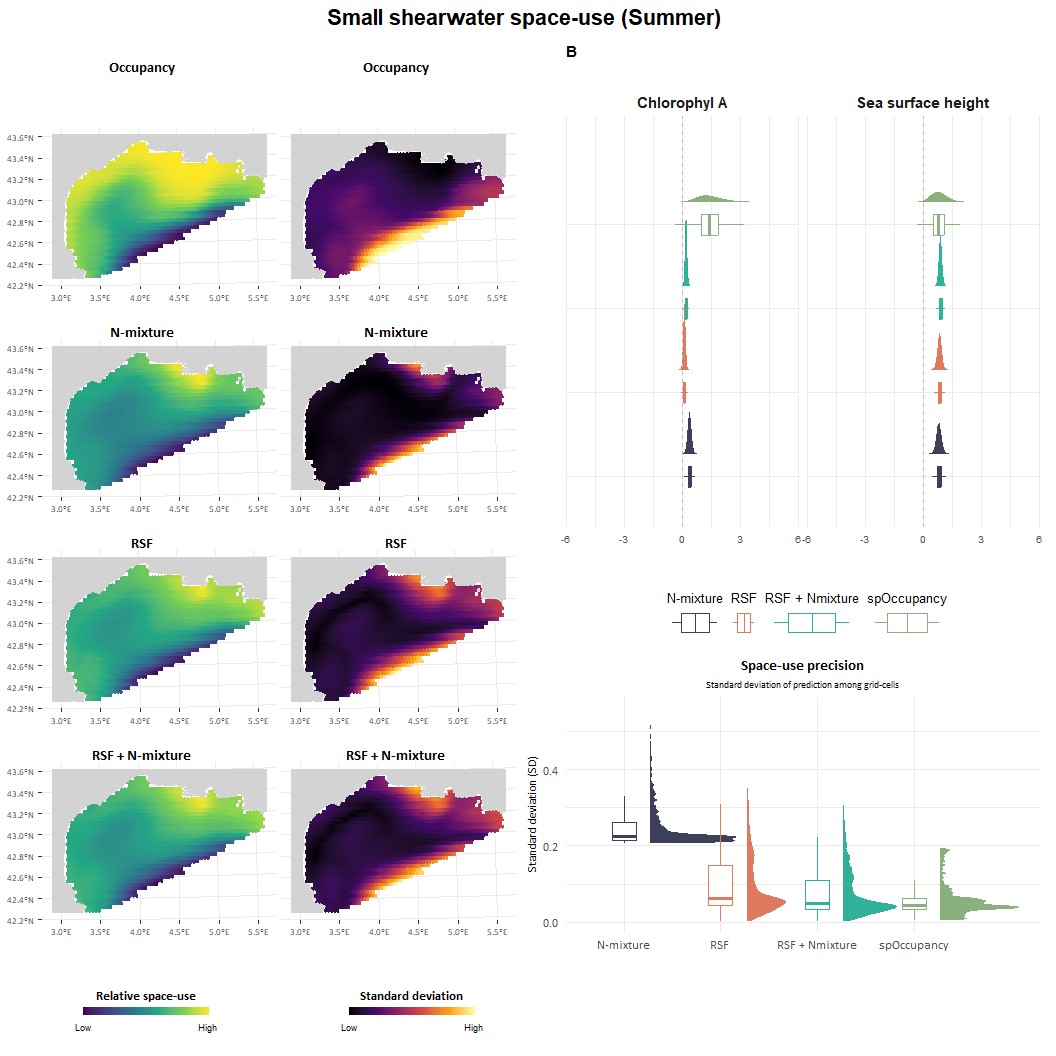
The results for occupancy and N-mixture models are presented for four representative species:

Common Tern and Little Gull in Table 2, and Yellow-legged Gull and Small Shearwaters in Table 3.

Results for other species are provided in Appendix C, Figure XX.



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



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

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

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

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

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

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

## 3.2 Comparison and integration of movement data with counts

The results obtained from movement data using RSF are consistent with those derived from count data using N-mixture (Figure **??**) (refer to the appendix for other species). For small shearwaters, the maps produced are similar. However, a notable difference is observed for the Yellow-legged Gull: the

RSF map displays a West-East pattern with minimal predicted space use in the western part of the Gulf, whereas the N-mixture model predicts similar space use for both the western and eastern parts of the Gulf. In other words, higher space-use is predicted on the eastern side with RSF, corresponding to the area where Yellow-legged gulls were equipped.

Integrating GPS-tracking data with count data yields contrasting results depending on the species. For small shearwaters (Fig. 4), coefficient estimates are similar between RSF and N-mixture models, and integration results in more precise estimates. Conversely, for Yellow-legged gulls, the coefficient estimates differ significantly, and RSF coefficients show considerable uncertainty. The final map and coefficient 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 alone, resulting in identical maps.

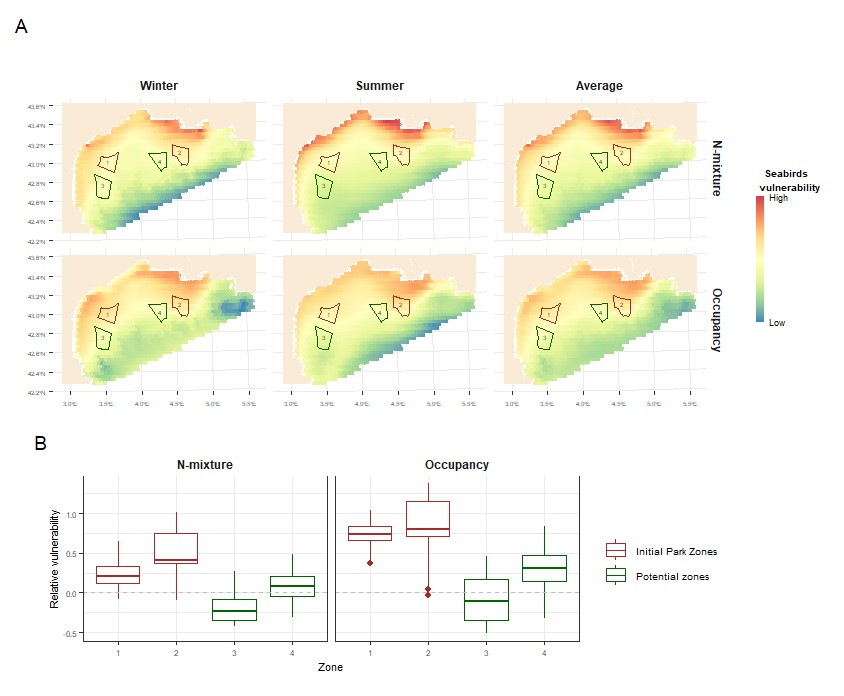
## 3.3 Seabird spatial vulnerability to wind farm

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 2).

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

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

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



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

# 4 Discussion

**4.1**

# 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.](https://doi.org/10.1111/j.1466-8238.2011.00697.x)

[1111/j.1466-8238.2011.00697.x.](https://doi.org/10.1111/j.1466-8238.2011.00697.x)

1. 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)
2. 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.](https://doi.org/10.1016/j.csr.2006.05.012)

[2006.05.012.](https://doi.org/10.1016/j.csr.2006.05.012)

1. Marion Fraysse et al. “Intrusion of Rhone River diluted water into the Bay of Marseille: Generation processes and impacts on ecosystem functioning”. *Journal of Geophysical Research: Oceans* (2014). doi: [10.1002/2014JC010022.](https://doi.org/10.1002/2014JC010022)
2. 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)
3. 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)
4. D. B˘anaru et al. “Trophic structure in the Gulf of Lions marine ecosystem (north-western Mediterranean Sea) and fishing impacts”. *Journal of Marine Systems* (Feb. 1, 2013). doi:

[10.1016/j.jmarsys.2012.09.010.](https://doi.org/10.1016/j.jmarsys.2012.09.010)

1. Carles Carboneras. “Seabirds in the Gulf of Lions shelf and slope area” (2013).
2. Richard Zotier, Vincent Bretagnolle, and Jean-Claude Thibault. “Biogeography of the marine birds of a confined sea, the Mediterranean”. *Journal of Biogeography* (1999). doi: [10.1046/j.](https://doi.org/10.1046/j.1365-2699.1999.00260.x)

[1365-2699.1999.00260.x.](https://doi.org/10.1046/j.1365-2699.1999.00260.x)

1. 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)

1. L. Ruffino et al. “Invasive rats and seabirds after 2,000 years of an unwanted coexistence on Mediterranean islands”. *Biological Invasions* (Aug. 1, 2009). doi: [10.1007/s10530-0089394-z.](https://doi.org/10.1007/s10530-008-9394-z)
2. K. Laneri et al. “Trawling regime influences longline seabird bycatch in the Mediterranean: new insights from a small-scale fishery”. *Marine Ecology Progress Series* (2010).
3. Philippe M. Cury et al. “Global Seabird Response to Forage Fish Depletion—One-Third for the

Birds”. *Science* (Dec. 23, 2011). doi: [10.1126/science.1212928.](https://doi.org/10.1126/science.1212928)

1. 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)
2. Marina Codina-Garc´ıa et al. “Plastic debris in Mediterranean seabirds”. *Marine Pollution*

*Bulletin* (Dec. 15, 2013). doi: [10.1016/j.marpolbul.2013.10.002.](https://doi.org/10.1016/j.marpolbul.2013.10.002)

1. 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)
2. Stefan Garthe and Ommo Hu¨ppop. “Scaling possible adverse effects of marine wind farms on seabirds: developing and applying a vulnerability index”. *Journal of Applied Ecology* (2004). doi:

[10.1111/j.0021-8901.2004.00918.x.](https://doi.org/10.1111/j.0021-8901.2004.00918.x)

1. Minist`ere de la transition ´ecologique. *Projets de parcs ´eoliens flottants et de leurs raccordements en Mer M´editerran´ee. Synth`ese du dossier de saisine de la Commission nationale du d´ebat public*.

2020.

1. 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.
2. 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.
3. 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)
4. Go¨ran Brostro¨m et al. “Atmosphere and ocean dynamics (2019). In: Perrow, M.R. (ed.) Wildlife and Wind Farms, Conflicts and Solutions. Vol. 3 Offshore: Potential effects, Chapter: 3.” Pelagic Publishing, Exeter, UK, Sept. 5, 2019. isbn: 978-1-78427-131-2.
5. 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.
6. 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)
7. 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)

1. 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)
2. 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.
3. 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)
4. BOURDEIX Jean-Herv´e. “PELMED - PELAGIQUES MEDITERRANEE” (1985).´ doi: [10.](https://doi.org/10.18142/19)

[18142/19.](https://doi.org/10.18142/19)

1. David A. W. Miller et al. “The recent past and promising future for data integration methods to estimate species’ distributions”. *Methods in Ecology and Evolution* (2019). doi: [10.1111/2041210X.13110.](https://doi.org/10.1111/2041-210X.13110)
2. Robert J. Fletcher Jr. et al. “A practical guide for combining data to model species distributions”. *Ecology* (2019). doi: [10.1002/ecy.2710.](https://doi.org/10.1002/ecy.2710)
3. Nick J. B. Isaac et al. “Data Integration for Large-Scale Models of Species Distributions”. *Trends in Ecology & Evolution* (Jan. 1, 2020). doi: [10.1016/j.tree.2019.08.006.](https://doi.org/10.1016/j.tree.2019.08.006)
4. Elise F. Zipkin, Brian D. Inouye, and Steven R. Beissinger. “Innovations in data integration for modeling populations”. *Ecology* (2019). doi: [10.1002/ecy.2713.](https://doi.org/10.1002/ecy.2713)
5. “Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS”. Ed. by Marc K´ery and J. Andrew Royle. Academic Press, Jan. 1, 2020. isbn: 978-0-12-823768-7. doi: [10.1016/B978-0-12-809585-0.12001-0.](https://doi.org/10.1016/B978-0-12-809585-0.12001-0)
6. 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)
7. 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.](https://doi.org/10.1890/0012-9658(2002)083%5B2248:ESORWD%5D2.0.CO;2)

[0.CO;2.](https://doi.org/10.1890/0012-9658(2002)083%5B2248:ESORWD%5D2.0.CO;2)

1. 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)

1. 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)

1. 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)

1. Laurentiu Rozylowicz et al. “Occupancy and N-mixture modeling applications in ecology: A bibliometric analysis”. *Global Ecology and Conservation* (Apr. 1, 2024). doi: [10.1016/j. gecco.2024.e02838.](https://doi.org/10.1016/j.gecco.2024.e02838)
2. Robert J. Ward et al. “Optimising monitoring efforts for secretive snakes: a comparison of occupancy and N-mixture models for assessment of population status”. *Scientific Reports* (Dec. 22, 2017). doi: [10.1038/s41598-017-18343-5.](https://doi.org/10.1038/s41598-017-18343-5)
3. D.I. MacKenzie and J.D. Nichols. “Occupancy as a surrogate for abundance estimation — Animal

Biodiversity and Conservation” (Mar. 29, 2007).

1. 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)
2. 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)
3. Clara P´eron et al. “Importance of coastal Marine Protected Areas for the conservation of pelagic seabirds: The case of *Vulnerable* yelkouan shearwaters in the Mediterranean Sea”. *Biological Conservation* (Dec. 1, 2013). doi: [10.1016/j.biocon.2013.09.006.](https://doi.org/10.1016/j.biocon.2013.09.006)
4. 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)

1. Valentin Lauret, Nicolas Courbin, and Aur´elien Besnard. “Comparing and integrating movement and counts data to estimate animal habitat selection” ().
2. Gareth Bradbury et al. “Mapping Seabird Sensitivity to Offshore Wind Farms”. *PLOS ONE*

(Sept. 11, 2014). doi: [10.1371/journal.pone.0106366.](https://doi.org/10.1371/journal.pone.0106366)

1. 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)

1. Gr´egoire Certain and Vincent Bretagnolle. “Monitoring seabirds population in marine ecosystem: The use of strip-transect aerial surveys”. *Remote Sensing of Environment*. Earth Observations for Marine and Coastal Biodiversity and Ecosystems Special Issue (Aug. 15, 2008). doi: [10.](https://doi.org/10.1016/j.rse.2008.01.019)

[1016/j.rse.2008.01.019.](https://doi.org/10.1016/j.rse.2008.01.019)

1. L. L. Eberhardt. “Transect Methods for Population Studies”. *The Journal of Wildlife Management*

(1978). doi: [10.2307/3800685.](https://doi.org/10.2307/3800685)

1. 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)
2. Sophie Laran et al. “Seasonal distribution and abundance of cetaceans within French waters-

Part I: The North-Western Mediterranean, including the Pelagos sanctuary”. *Deep Sea Research*

*Part II: Topical Studies in Oceanography*. Abundance, distribution and habitats of Atlantic and Mediterranean marine megafauna (July 1, 2017). doi: [10.1016/j.dsr2.2016.12.011.](https://doi.org/10.1016/j.dsr2.2016.12.011)

1. Gianpiero Cossarini et al. “High-Resolution Reanalysis of the Mediterranean Sea Biogeochemistry (1999–2019)”. *Frontiers in Marine Science* (Nov. 29, 2021). doi: [10.3389/fmars.2021.](https://doi.org/10.3389/fmars.2021.741486)

[741486.](https://doi.org/10.3389/fmars.2021.741486)

1. Romain Escudier et al. “A High Resolution Reanalysis for the Mediterranean Sea”. *Frontiers in*

*Earth Science* (Nov. 24, 2021). doi: [10.3389/feart.2021.702285.](https://doi.org/10.3389/feart.2021.702285)

1. 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)
2. Juan Serratosa et al. “Environmental Drivers of Seabird At-Sea Distribution in the Eastern South Pacific Ocean: Assemblage Composition Across a Longitudinal Productivity Gradient”. *Frontiers in Marine Science* (Feb. 5, 2020). doi: [10.3389/fmars.2019.00838.](https://doi.org/10.3389/fmars.2019.00838)
3. 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)

1. Steffen Oppel et al. “Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds”. *Biological Conservation*. Seabirds and Marine Protected Areas planning (Nov. 1, 2012). doi: [10.1016/j.biocon.2011.11.013.](https://doi.org/10.1016/j.biocon.2011.11.013)
2. Francisco V. D´enes, Lu´ıs F´abio Silveira, and Steven R. Beissinger. “Estimating abundance of unmarked animal populations: accounting for imperfect detection and other sources of zero inflation”. *Methods in Ecology and Evolution* (2015). doi: [10.1111/2041-210X.12333.](https://doi.org/10.1111/2041-210X.12333)
3. 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)
4. Jos´e J. Lahoz-Monfort, Gurutzeta Guillera-Arroita, and Brendan A. Wintle. “Imperfect detection impacts the performance of species distribution models”. *Global Ecology and Biogeography* (2014). doi: [10.1111/geb.12138.](https://doi.org/10.1111/geb.12138)
5. Marc K´ery. “Towards the modelling of true species distributions”. *Journal of Biogeography*

(2011). doi: [10.1111/j.1365-2699.2011.02487.x.](https://doi.org/10.1111/j.1365-2699.2011.02487.x)

1. 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)
2. Perry de Valpine et al. “Programming with models: writing statistical algorithms for general model structures with NIMBLE”. *Journal of Computational and Graphical Statistics* (Apr. 3, 2017). doi: [10.1080/10618600.2016.1172487.](https://doi.org/10.1080/10618600.2016.1172487) arXiv: [1505.05093[stat].](https://arxiv.org/abs/1505.05093%20%5Bstat%5D)
3. 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)

1. Sumio Watanabe. “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable

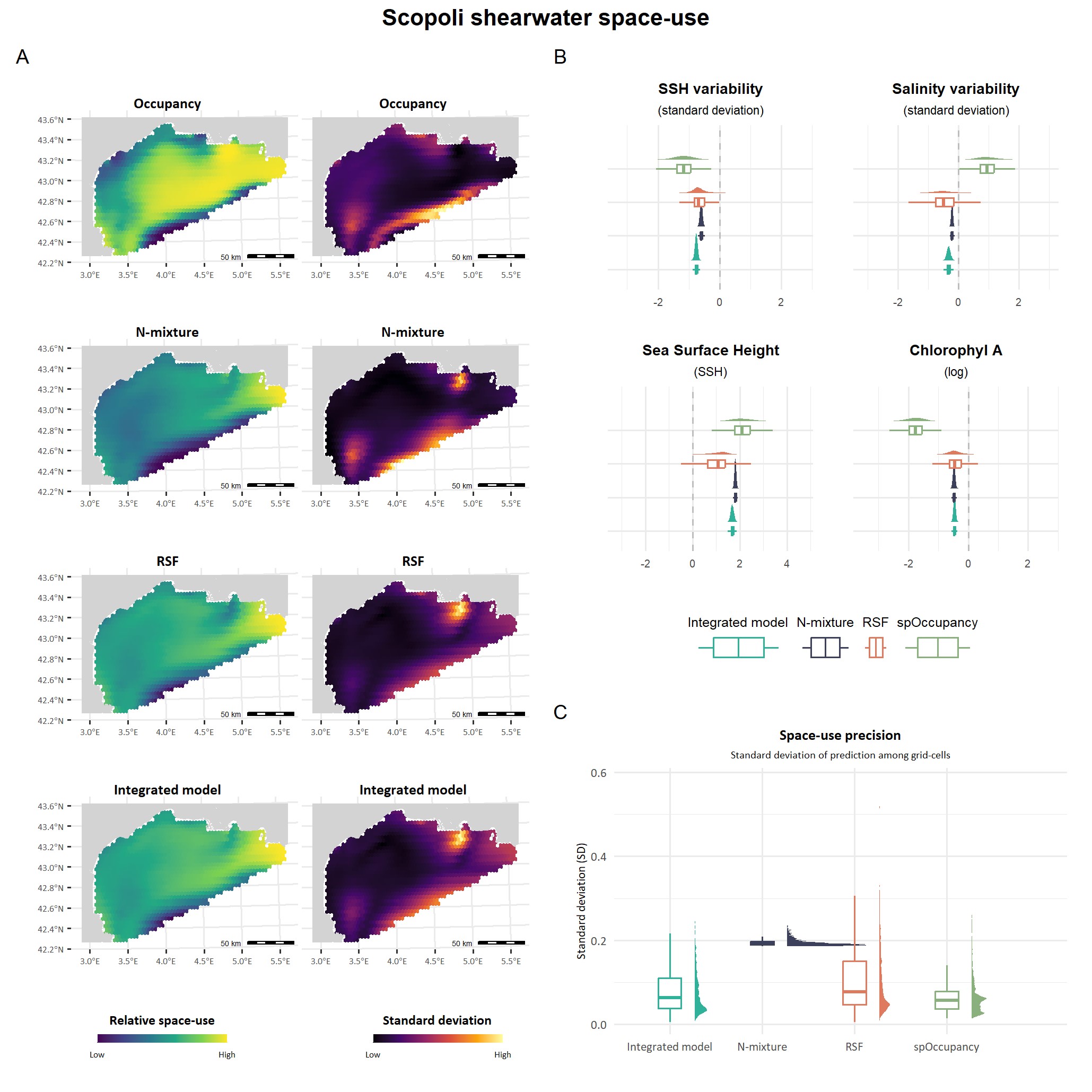
Information Criterion in Singular Learning Theory” (2010).

1. M. B. Hooten and N. T. Hobbs. “A guide to Bayesian model selection for ecologists”. *Ecological*

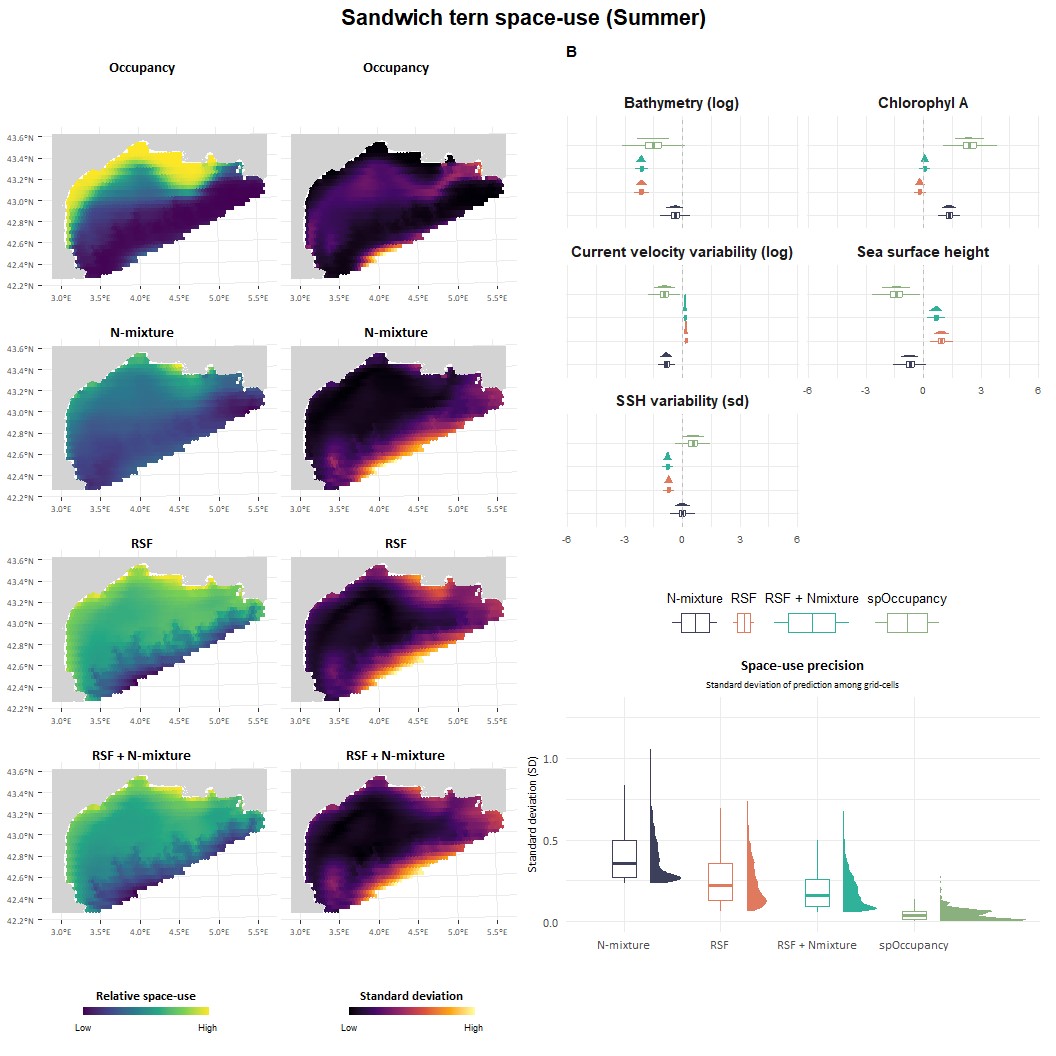
*Monographs* (2015). doi: [10.1890/14-0661.1.](https://doi.org/10.1890/14-0661.1)

1. J´erˆome Gu´elat and Marc K´ery. “Effects of spatial autocorrelation and imperfect detection on species distribution models”. *Methods in Ecology and Evolution* (2018). doi: [10.1111/2041210X.12983.](https://doi.org/10.1111/2041-210X.12983)
2. OFB and GISOM. *Identification et priorisation de la responsabilit´e de chaque sous r´egion marine pour les enjeux ornithologiques.* 2020.

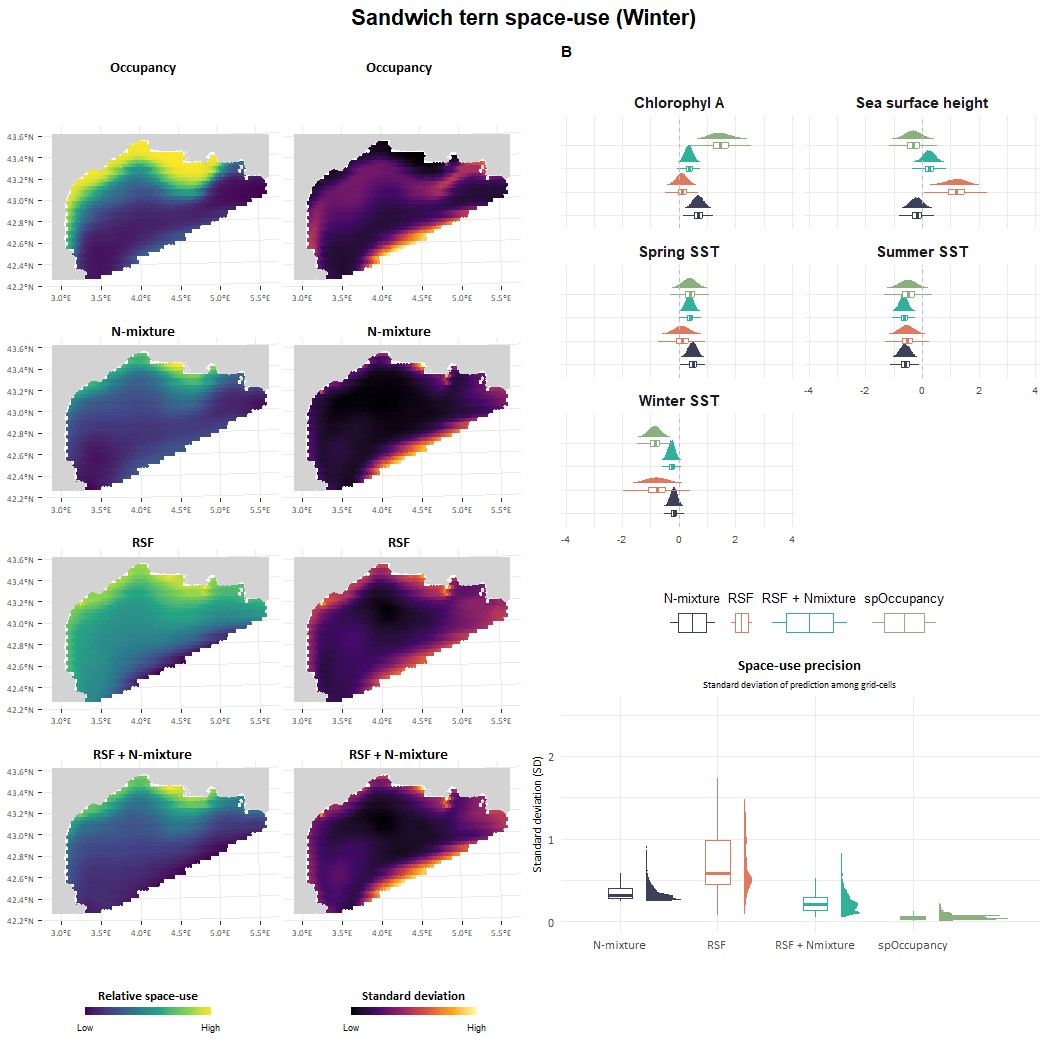
# II Appendix B



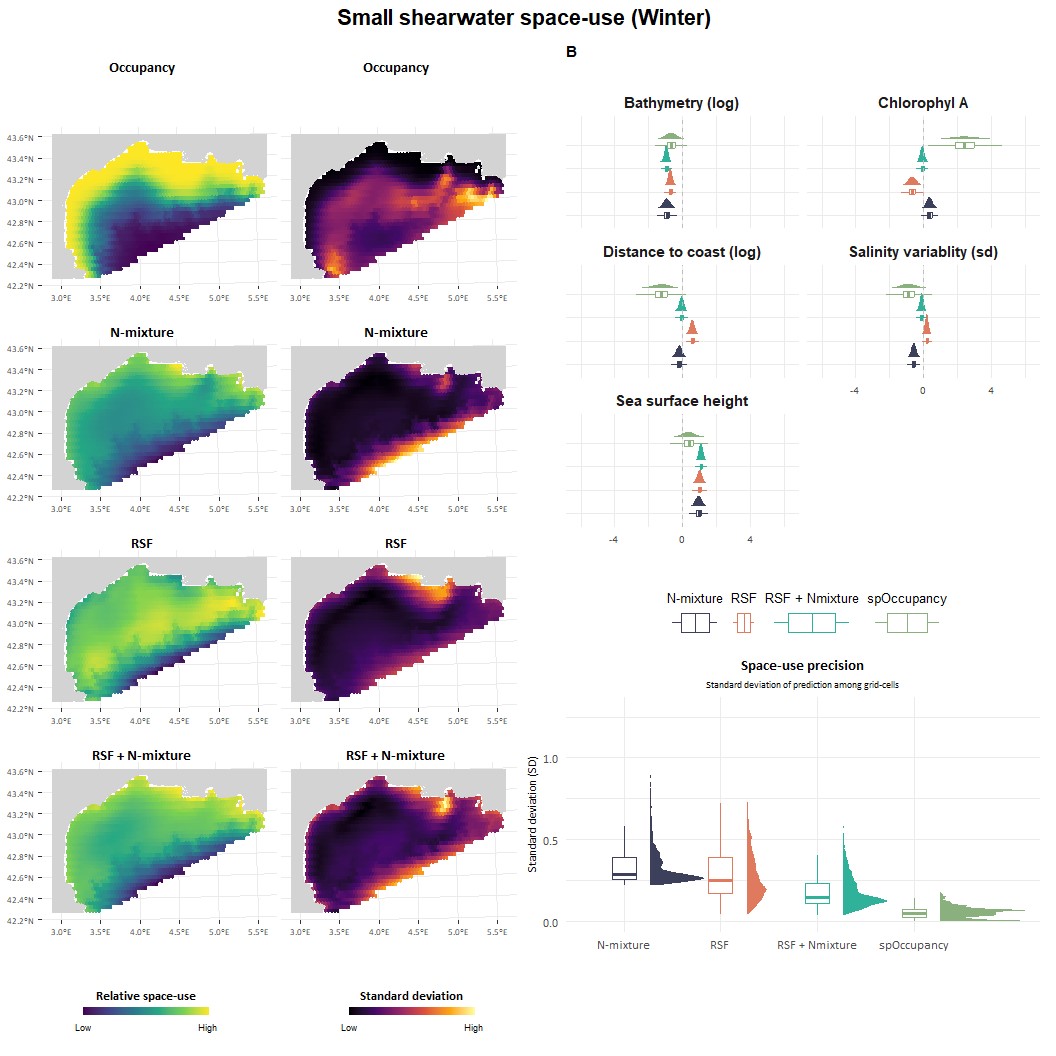
**Fig.7: Scopoli’s shearwater (***Calonectris diomedea***) distribution in the Gulf of Lion.** Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining both movement and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty on the results, computed as the standard deviation on the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Prediction on the predicted space use, computed as the standard deviation of the grid cells.



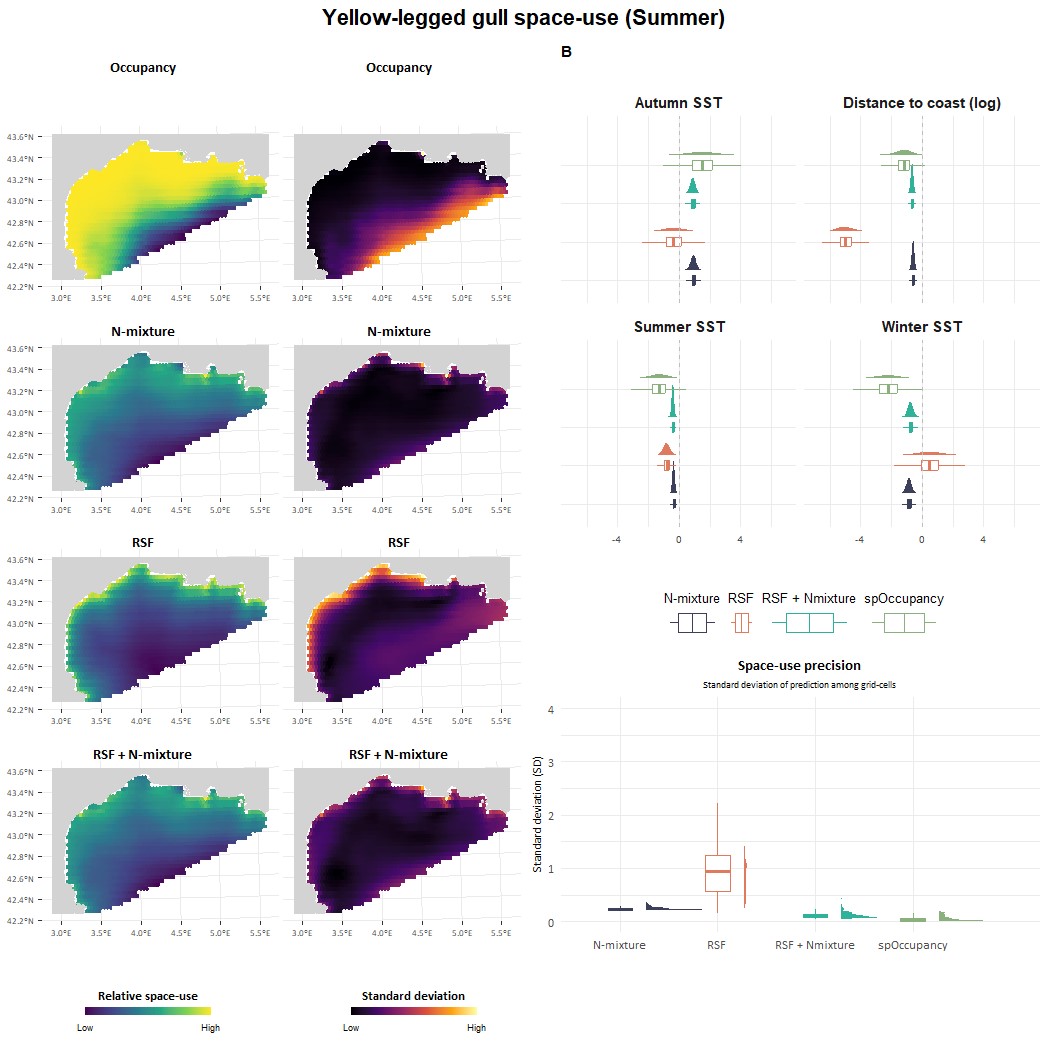
**Fig.8: Sandwich tern (***Sterna sandvicensis***) distribution in the Gulf of Lion.** Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining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



**Fig.9: Sandwich tern (***Sterna sandvicensis***) winter distribution in the Gulf of Lion.** Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining both movement and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty on the results, computed as the standard deviation on the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Prediction on the predicted space use, computed as the standard deviation of the grid cells.



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



**Fig.11: Yellow-legged gull (***Larus michaellis***) summer distribution in the Gulf of Lion.** Estimated in four different ways: using an occupancy model with detection/non-detection data, an N-mixture model with count data, a Resource Selection Function (RSF) with movement data, and an integrated model combining both movement and count data (RSF + N-mixture). **(A)** Maps of the relative space-use intensity and the associated uncertainty on the results, computed as the standard deviation on the prediction. **(B)** Regression coefficient estimates for each method and each covariate. **(C)** Prediction on the predicted space use, computed as the standard deviation of the grid cells.

# III Appendix C

