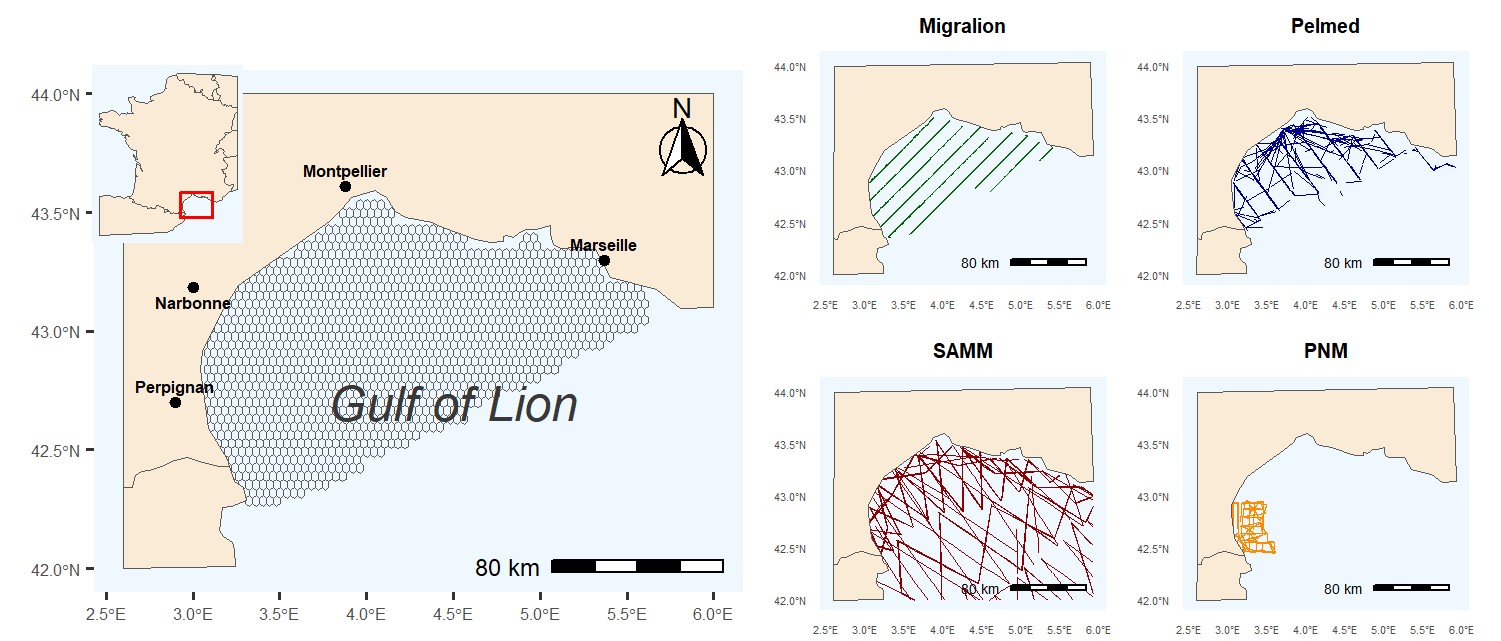
**1 Methods**

# 1.1 Study area



**Fig.1: Study area and transect lines -** (Left) Location of the Gulf of Lion. The gridded area represents the extent of our study area. (Right) The transects followed for each program. It’s important to note that not every transect was sampled during each campaign. Transects that extend beyond the study area were excluded from our analysis.

The Gulf of Lions 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 [1, 2] and small-scale upwellings created by winds [3]. The total area considered for the study had a surface of approximately 17,300 km². Within this area, assessment units were defined by polygonal 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 good ratio between occupied vs unoccupied sites for every species included in the analysis.

# 1.2 Seabirds data

## 1.2.1 Count Data

Count data were collected from four different programs, as summarized in 1. Observations were conducted using strip-transect methodology (Fig.1) [4, 5, 6], employing ships or aircraft as counting platforms.

SAMM (Survol A´erien de la M´egafaune Marine, Aerial Census of Marine Megafauna) [7, 8], was the only aerial survey. It was conducted from November to February in 2011-2012 and from January to February in 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 focused on assessing small pelagic fish resources [9]. 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 are conducted twice a year since 2019, in autumn and late spring.

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

**Table 1:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Program | Years | Period of the year | Area covered | Counting platform |
| Migralion | 2022-2023 | Spring and autumn | Gulf of Lion | Boat |
| SAMM | 2011-2012 & 2018-2019 | Summer and winter | French ZEE | Aerial |
| Pelmed | 2017-2021 | Summer | Gulf of Lion | Boat |
| Megaobs | 2019-2021 | Early summer and late autumn | National Marine Parc of The Gulf of Lion | Boat |

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. This affected Balearic shearwater (*Puffinus mauretanicus*) and Yelkouan shearwater (*Puffinus yelkouan*), both labeled as small shearwaters, as well as Pomarine skua (*Stercorarius pomarinus*) and Arctic skua (*Stercorarius parasiticus*), labeled as small skuas. The datasets contained sufficient information to model the distribution of twelve different species 3.

**1.2.2 Telemetry data**

# 1.3 Environmental data

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

Equ2: **Environmental covariates used in the study.**

We extracted environmental variables from the E.U. Copernicus Marine [10, 11] (https://data.marine.copernic and MARSPEC [12] (http://www.marspec.org) databases. A total of 8 geophysical, biotic, and climate covariates were included and are listed in Tab.2. These 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. From the monthly temporal resolution, we also computed the standard deviation for each covariates and considered it as a measure of temporal variability of the covariate value. Sea Surface Temperature (SST) showed a great seasonal variability, thus average value were calculated for each season. The spatial resolution for all covariates was set to be consistent with the spatial resolution used for our grid. When the correlation between two covariates was too high (R²¿0.8), one of the covariates was excluded from the selection procedure.

# 1.4 Seabirds occupancy modeling with detection/non-detection data

The most reliable method to combine several data sources is to develop integrated species distribution models [13], because each data source was collected using different protocols (number of observers, counting platforms, speed...) leading to different detection biases. Integrated models are often hierarchical models, i.e. sequence of related models ordered by their conditional probability structure [14]. In hierarchical models, there are one or more intermediate levels involving latent variables. In our case, the models have a component for the observations and a component for the latent state describing the ecological state of our system [14].

We initially employed occupancy modeling [15] to estimate 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 *ψ* [14, 15].

## 1.4.1 Latent process

Let’s consider J sites, 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 [15]:

*Zj* ∼ Bernoulli(*ψj*)

(1) logit(

Here, *ψj* represents the probability of species presence at site j. The presence probability is modeled by a logistic regression of site-specific environmental covariates *Xj*, with *β* being a vector of regression coefficients [14, 15].

## 1.4.2 Observation process

The non-detection of a species at a site does not necessarily mean it is not using that site. The observation process can yield false absence (*Y* = 0) while the species is occupying a site (*Z* = 1) (false negative errors). Not accounting for imperfect detection could lead to wrong inferences [14, 16, 17, 18].

Let *Yj,k,r* be the observed detection (1) or non-detection (0) at site j during survey k for dataset r. Each dataset has its own bias and detection probability, which is why the observation process is modeled separately for each data source. However, the observation process is conditional on the latent process described in 1, which is common and 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*) = *αr1* + *αr2 transect lengthj,k,r*

Where *pj,k,r* is the detection probability at site j for replicate 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.

## 1.4.3 Model fitting

This model was implemented in R (ref) using the package spOccupancy [19], allowing fast computation, and easy model selection. 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. Convergence was checked by verifying that *R <*ˆ 1*.*1

[20].

# 1.5 Seabirds abundance modeling with count data

In a second time, we tried to infer seabird abundance using N-mixture models [21, 22]. Similarly to occupancy models, N-mixture models have a hierarchical structure where the observation is defined conditional on the latent process describing the true abundance [21, 22].

## 1.5.1 Latent process

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

*Nj* ∼ *Poisson*(*λj*)

(3)

*log*(*λj*) = *Xj β*

The Poisson mean *λ* is a function of site specific covariates.

## 1.5.2 Observation process

Again, we employ an integrated model with a separate model for each dataset to accommodate differences in the observation process. The real number of individuals at a site is consistent across all datasets, but the number of observed individuals varies for each dataset. The observed number of individuals *Nj,k,robs* is assumed to follow a Bernoulli distribution conditional on the true number of individual *Nj*:

*Nj,k,robs* ∼ *Bernoulli*(*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.

## 1.5.3 Model fitting

This model was implemented using Nimble package [23]. We ran 3 MCMC chains with 150,000 iterations each and a burn-in of 15,000 iterations. Convergence was checked by verifying that *R <*ˆ 1*.*1

[20].

**1.6 Integration of telemetry data for relative space-use modeling**

# 1.7 Model selection

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) and k-fold cross-validation. In k-fold cross-validation, the model is fitted k times, each time excluding a fraction J/k of the data, with J being the total number of sites, and then making predictions. Predictive performance is assessed by computing the deviance [24]. When predictive scores were too close, 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.
* Secondly, we divided the covariates into three blocks and tested all possible combinations within each block, selecting the one with the best predictive performance.

Vulnerability index

Scientific name Common name

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

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

* Then, with the three covariate 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 tested whether adding a spatial autocorrelation structure improved predictive performance. Different autocorrelation structures proposed in spOccupancy were tested, but they all resulted in overfitting and poor performance in cross-validation. Therefore, we did not pursue further analysis of spatial autocorrelation.

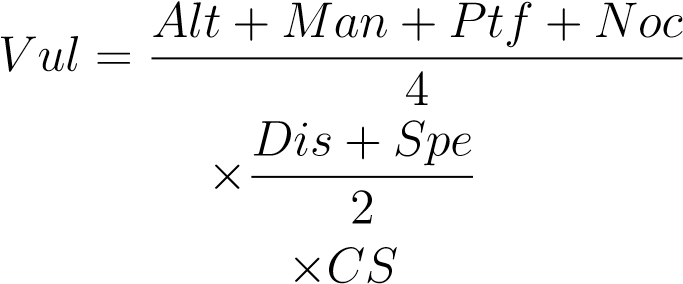
# 1.8 Mapping seabirds’ vulnerability to wind farms

## 1.8.1 Vulnerability indexes

The synthetic index constructed to evaluate species risk to offshore wind farms integrates three main aspects of seabird vulnerability [25, 26]:

1. **Conservation Status (CS):** Defined in [27], 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 period as seabirds distribution can vary greatly between these periods. 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** is calculated using two criteria: the sensibility to disturbance by wind turbines, boats, and helicopters (Dis) and habitat specialization score (Spe).

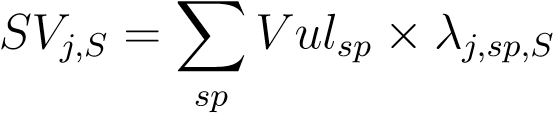
The vulnerability index is computed as follows [26]:



Vulnerability indexes for every species are indicated in Tab.3

## 1.8.2 Vulnerability maps

To create a single vulnerability map, the values of individual maps are first rescaled to fall within the range of 0 to 1. Then, all the maps are combined, using the vulnerability index as a weighting factor. The seasonal seabird vulnerability *SV* at site *j* and season *S* is defined as:



where *λj,sp,S* represents the space-use intensity of site *j* by species *sp* during season *S*. This

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

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

This approach ensures that both breeding and non-breeding seasons contribute to the overall assessment of seabird sensitivity.

**References**

1. Julie Gatti et al. “The Rhone river dilution zone present in the northeastern shelf of the Gulf of Lion in December 2003”. *Continental Shelf Research* (Oct. 1, 2006). doi: [10.1016/j.csr.](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. 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)
3. 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)
2. Mark L. Tasker et al. “Counting Seabirds at Sea from Ships: A Review of Methods Employed and a Suggestion for a Standardized Approach”. *The Auk* (July 1, 1984). doi: [10.1093/auk/ 101.3.567.](https://doi.org/10.1093/auk/101.3.567)
3. 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. Sophie Laran et al. “Seasonal distribution and abundance of cetaceans within French watersPart 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. 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. 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)
2. Elizabeth J. Sbrocco and Paul H. Barber. “MARSPEC: ocean climate layers for marine spatial ecology”. *Ecology* (2013). doi: [10.1890/12-1358.1.](https://doi.org/10.1890/12-1358.1)
3. 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)
4. “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)
5. 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. 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)
2. 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)
3. 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)
4. 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)
5. Andrew Gelman and Donald B. Rubin. “Inference from Iterative Simulation Using Multiple Sequences”. *Statistical Science* (Nov. 1, 1992). doi: [10.1214/ss/1177011136.](https://doi.org/10.1214/ss/1177011136)
6. 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)
7. 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)
8. 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)
9. 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)
10. 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)
11. 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)
12. GISOM OFB. *Identification et priorisation de la responsabilit´e de chaque sous r´egion marine pour les enjeux ornithologiques.* 2020.