

# Comparing and integrating movement and counts data to estimate animal habitat selection

Valentin Lauret<sup>1,2\*</sup>, Nicolas Courbin<sup>1</sup>, Aurélien Besnard<sup>1</sup>

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

(2) France Energies Marines

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

---

**Abstract:** Both movement tracking and population monitoring contribute to understand how animals select the components of the landscape. The spatial and temporal scales at which animal habitat selection is studied affect the monitoring design. Counts and movement data capture complementary aspects of habitat selection. While animal tracking through movement provide detailed insights on habitat selection behaviour, population counts provide large-scale information about species or population distribution. As both movement and population methods estimate the relative selection for some ecological variables, we suggested that it is possible in some cases to compare and to combine both data sources.

Based on simulations and Sandwich terns (*Thalasseus sandvicensis*) data collected in the French Mediterranean Sea, we compared habitat selection estimates obtained by Resource Selection Function (RSF) using movement data, and estimates obtained by hierarchical Poisson generalized linear models (GLM) using counts data. We developed an integrated model that combine RSF and Poisson GLM, hence using movement and counts data. Our simulations illustrated that RSF, Poisson GLM, and the integrated model correctly estimated habitat selection while the integrated model achieved better accuracy and bias than models using the data sources separately. The case study on sandwich terns presented similar conclusions: the integrated model took information from both movement and counts data to estimate relative habitat selection. Our study represented a novel attempt towards the formal integration of movement and population data to estimate habitat selection. We hope to contribute to an promising research avenue as movement studies and count data are abundant in many ecological contexts.

**Keywords:** animal ecology, habitat selection, data integration, occupancy models, movement ecology, step selection function.

## 1 Introduction

Telemetry tracking and population counts provide insightful spatial information to understand how animals use the landscape, hence thriving from animal cognition (Robira et al., 2021) to conservation application (Courbin et al., 2023). The increasing accessibility of animal tracking devices and the ease of collecting population counts make movement data and counts widely available to estimate habitat selection (Matthiopoulos et al., 2022; Northrup et al., 2022; Thurfjell et al., 2014). Counts and movement data coexist in multiple ecological contexts (e.g. ungulates Gamelon et al. (2021), Fradin & Chamaillé-Jammes (2023), Chandler et al. (2022); seabirds Laran et al. (2017), Courbin et al. (2023)), hence capturing complementary aspects of habitat selection. However, comparing and combining both data sources to estimate habitat selection is not trivial whereas appealing (Matthiopoulos et al., 2022). Movement data and population counts focus on different ecological scales and the fundamental difference in the nature of the recorded data between population counts and individual animal movement constitute an important challenge to compare and combine the data (Lindenmayer et al., 2022; Michelot et al., 2019; Nathan et al., 2022).

## Movement

First, animal movement obtained through telemetry devices (e.g. GPS) provide detailed information about spatial selection at the individual level. Movement data help to tackle behavioural questions related to spatial habitat selection, (e.g. *do gorillas memorize the most efficient routes between food patches?* (Robira et al., 2021), or *How do wild boars select resting sites?* (Fradin & Chamaillé-Jammes, 2023)). At large temporal scales, movement data collected on multiple individuals allow to infer population distributions (Nathan et al., 2022), e.g. estimating year-round seabirds distribution in the Northeastern Atlantic (Fauchald et al., 2021). One strength of movement data to inform habitat selection is that telemetry technologies lead to the collection of thousands of locations used by one individual, which constitute a detailed knowledge of how an individual select the components of the available landscape.

Habitat Selection Functions (HSFs) constitutes the main modelling tool to quantify the selection behaviour of animals analysing movement data (Northrup et al., 2022). HSFs link used locations of individual obtained from tracking devices to available locations generated in the landscape to fit the models, i.e. use-availability design (Manly et al., 2002; Northrup et al., 2022). HSFs estimate the probability that an individual selects the used location over the available locations via a generalized linear regression of environmental predictors, often a logistic regression (see Methods section).  $\text{logit}(\omega) = \beta_0 + \beta X$ . The intercept  $\beta_0$  represents the ratio of used to available points when all other covariates are set to 0 and therefore is biologically meaningless (Fieberg et al., 2021; Muff et al., 2020). Dropping the intercept,  $\exp(\beta X)$  of the HSF represents the relative selection strength for habitat  $X$  (Northrup et al., 2022), and the  $\beta$ 's inform for selection or avoidance of environmental variable  $X$ .

However, up-scaling habitat selection coefficient estimated from data collected at individual level to the population level is not straightforward and raised methodological challenges to accommodate the scales of spatial selection Muff et al. (2020). Habitat selection coefficients at micro and macro scales may indeed differ as selection can be affected the environmental context and may be scale-dependent (Fieberg et al., 2021; Northrup et al., 2022). Under some biological conditions (see Methods and Discussion sections), Resource Selection Functions (RSFs) use telemetry data to estimate habitat selection at the population level (Fauchald et al., 2021; Northrup et al., 2022). Recent methodological developments, i.e. Step-Selection Functions (SSFs), enable to disentangle habitat selection from movement constraints when analyzing animal tracking (Avgar et al., 2017), allowing to account for variability between individuals (Muff et al., 2020), and to reconcile estimation of both habitat selection and movement descriptors simultaneously (Michelot et al., 2019).

## Counts

Second, counts data represent information collected at the population level informing the distribution of a given species or population. Easily collected over large spatial and temporal scales, counts (or presence-absence) data can tackle population dynamics and species distribution questions (e.g. mapping wolf recolonization in France over 20 years (Louvrier et al., 2017)). Counts data can inform about habitat selection of defined sampled populations based on standardized protocols (Royle & Nichols, 2003) or opportunistic data collection (Elith &

Leathwick, 2009). On the contrary of individual tracking that follow few individuals in details, counts protocols can collect information on a large number of individuals of the sampled population despite gathering less observations per individuals (MacKenzie, 2006).

While being easy to implement at a large spatial scale, sampling design collecting data at the species level hardly provide information on the behaviour of individuals, which we know affect habitat selection (e.g. do detected individuals stayed a long time or did they just cross the spatial unit (Rota et al., 2009)). From these data, we only know that the species used the spatial unit at a particular time. However, counts (or presence-absence) datasets are widely available and are used to estimate animal habitat selection in many ecological contexts (MacKenzie, 2006; Royle & Nichols, 2003). Species distribution models estimate habitat selection relying on counts, presence-absence, presence-only or detection/non-detection of species and possibly accounting for the specificities of the sampling design and imperfect detection (Mackenzie et al., 2002; Royle & Nichols, 2003). Species distribution models belong to the larger class of Inhomogeneous Point Process models (IPP) that correlate counts data (or presence-absence data) with environmental predictors through a generalized linear regression; often a logistic regression for presence-absence data  $\text{logit}(\psi) = \beta_0 + \beta X$ , or via a log-linear regression and a Poisson draw for counts data  $\log(\lambda) = \beta_0 + \beta X$ , i.e. hereafter named a Poisson Generalized Linear Model (Poisson GLM). Depending on the monitoring design and the ecology of targeted species,  $\psi$  is either interpreted as probability of species presence at the sampling unit or the probability that the species uses the sampling unit knowing it is present in the study area. For counts models  $\lambda$  is interpreted as the density of individuals per spatial sampling unit. In both modelling approaches, the  $\beta$ 's represent the relative selection of a particular environmental covariate by the individuals of the sampled population. Considering a discrete covariate indicating habitat A or B in linear regression  $\log(\lambda) = \beta_{\text{Habitat A or B}}$ ,  $\beta$  indicates whether animals prefer or avoid habitat A over B when both are available. When using dataset that include true absences, species distribution models estimate absolute habitat selection. The intercept is biologically meaningful as it translates the mean presence probability of the different sampled units when all covariates are set to 0. When true absence data are not available, one can generate pseudo-absence points hence following a use-availability modelling approach (described above, Fieberg et al. (2021)).

## IPP as a unifying framework

Species distribution and movement models do not infer the same absolute ecological process of habitat selection (Johnson, 1980). Yet, both methods provide the estimation of relative selection strength of habitat features via the  $\beta$ 's coefficients. Use-availability models using movement data estimate the relative probability that tracked individual chooses an habitat conditional on its availability (Avgar et al., 2015) while counts models estimate the probability the population uses a particular spatial unit (Mackenzie et al., 2002). However, both species distribution models and movement models estimate the relative selection of habitat components through the effect size of covariates in the linear models, the  $\beta$ 's. Aarts et al. (2012) showed the convergence of use-availability models (RSF or SSF) with inhomogeneous point process models (IPP) classically used for analysing counts data. As both RSF and Poisson GLM estimate habitat selection as a linear combination of environmental (biotic or abiotic) predictors, one can compare the relative habitat selection strength estimated by the coefficient  $\beta$ . The convergence of the relative habitat selection coefficient  $\beta$  opens perspectives to formally develop an integrated model that uses both data sources jointly.

## Integrated modelling

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

estimate a joint abundance or distribution process via a linear regression of environmental covariates, hence jointly informing the habitat selection coefficient  $\beta$ 's (D. A. W. Miller et al., 2019; Strebel et al., 2022). Under the conditions that RSF and Poisson GLM converge, we created an integrated model to estimate relative habitat selection integrating movement data via use-availability models and counts data via a Poisson GLM. We are not aware of any previous attempt to formally integrated movement and counts data to estimate habitat selection coefficient.

## In this paper

In this paper, we contributed to the long-lasting literature that aim at comparing habitat selection methods that use movement and counts data. We developed a novel framework to integrate both data sources into the same habitat selection model. Based on existing work, we explained when and how the inhomogeneous point process (IPP) framework can reconcile counts and movement data. Then, using simulations and on real data, we compared RSF and Poisson GLM and we showcased the possibilities of integrating movement and counts data to infer habitat selection process. First, we detailed our development of a Bayesian integrated model made of two main components : i) an Poisson GLM using counts data, ii) a RSF using movement data. Second, we performed simulations to test the validity of our integrated model and to assess its relevancy in a variety of situations. In Appendices, we extended our comparison to occupancy models that use presence-absence data and to step-selection functions (SSF). Then, we applied our integrated model to estimate relative habitat selection of Sandwich Terns (*Thalasseus sandvicensis*) in the Northwestern Mediterranean Sea. Finally, we discussed the opportunity of integrating movement and population data to understand habitat selection, and we suggested future extensions.

## 2 Methods

### 2.1 IPP as a unifying framework

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

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

where parameters to be estimated  $\beta_0$  is the density intercept and  $\beta$ 's are the regression coefficient of the environmental predictors  $X$ .

IPP constitute widely used tools to describe spatial ecological processes, including traditional Habitat-Selection Functions (HSFs) estimated from movement data (Aarts et al., 2012; Northrup et al., 2022) or density surface modelling for abundance estimation from counts data of mark-recapture data (e.g. within Spatial Distance Sampling, or Spatial Capture-Recapture D. L. Miller et al. (2013)). The work of Aarts et al. (2012) suggested that both the use-availability design of RSF and the Poisson align in the IPP framework to estimate the relative preference environmental covariates, resulting in similar  $\beta$ . When analysing movement data in an IPP, the points pattern represents the used locations of one individual. Whereas when analysing counts data in an IPP, the points pattern represent the density of individuals from the studied population. In ecological terms, the locations pattern of movement track, and ii) the expected number of individuals estimated when analysing population counts provide similar relative habitat selection coefficients. In the following sections, we detailed how to estimate relative habitat selection when analysing counts data and movement data.

## 2.2 Counts data - Poisson generalized linear model

To estimate habitat preferences from counts data, we fitted a Poisson generalized linear model (GLM) over a discrete geographical space where the expected count  $N$  in grid-cell follow the intensity of an IPP (Aarts et al., 2012). We used counts data collected over multiple grid-cells to constitute,  $N$  a counts dataset.  $N_s$  represents the expected counts in grid-cell  $s$  and is estimated as a Poisson draw in  $\lambda_s$  the IPP intensity at  $s$ .

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

### 2.2.0.1 Accounting for imperfect detection

One can extend the Poisson GLM to account for imperfect detection through a hierarchical formulation (Royle & Nichols, 2003). Let us name  $Y_s$  the observed counts at  $s$ . We estimated  $N_s$  the latent number of points in the grid-cell  $s$  through a binomial draw with detection probability  $p$ . Then,  $Y_s$  links with the IPP intensity through equations (1) and (2).



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

$p$  represent the probability of detecting one individual.  $p$  can be grid-cell specific and estimated as a logistic regression of covariates (see Sandwich tern application).

### 2.3 Movement data - Resource Selection Function

HSFs are classically modelled using a logistic regression where animal occurrences are coded as 1s and available distribution or ‘pseudo-absence’ points are 0s. Let us name  $K$  the dataset that stores the binary 1/0 (use/available) data. HSFs operate in continuous geographical space, and the  $K_s$  of each point  $s$  in  $S$  is modeled as a Bernoulli draw in the IPP intensity  $\lambda$  from equation (1).

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

However, when fitting logistic regression to use-availability points dataset, the intercept  $\beta_0$  is meaningless as it represents the ratio of used points when all covariates  $X$  are set to 0. Aarts et al. (2012) and Warton & Shepherd (2010) demonstrated the equivalence between an IPP and HSFs when dropping the intercept. Slope coefficients estimated from logistic regression are equivalent to slope coefficient estimated with IPP (i.e. with log-linear regression, Muff et al. (2020), Fithian & Hastie (2013), Warton & Shepherd (2010)). Thus, predictions of a HSF written as  $\omega = \exp(\beta X)$  is equivalent to the relative (not absolute) intensity of an IPP. Then,  $\beta$  estimates of the HSFs are indicating the relative habitat selection strength (Northrup et al., 2022).

Among HSFs used to analyse movement data, one can distinguish Resource Selection Functions (RSFs) and Step Selection Functions (SSFs). RSFs are a particular class of HSFs where the logistic regression of used ( $K=1$ ) vs available points ( $K=0$ ) is fitted simultaneously on the entire dataset (Manly et al., 2002). The process consists in 1) subsetting the tracking dataset to avoid autocorrelation between successive locations and to obtain a set of used locations, 2) generating a set of available points in the environment according to biological movement capacity of the animals (e.g. centered around a colony in the case of a central place forager), and 3) fitting a logistic regression on the use-available dataset. A SSF is relatively similar except it accounts for autocorrelation in the tracking data. To implement an SSF, we generate available points at each time step conditional of local movement pattern and fit a conditional logistic regression at each step (Avgar et al. (2015), see Appendix X for details about SSF). Accounting for local movement, SSFs estimate micro-scale habitat selection and coefficients  $\beta$  are not strictly equivalent to habitat



selection coefficients of RSF applied to the same tracks (Michelot et al., 2019; Signer et al., 2017). In this paper, we used RSF as being the simplest utilization of use-availability design that align with the IPP framework (Aarts et al., 2012). In the Appendices, we displayed the special case of SSFs, comparing *SSFs* estimates with RSFs, and we discussed its possible integration in an integrated model with counts data.

### 2.3.0.1 Accounting for individual heterogeneity in habitat selection

To account for variation in habitat selection among  $I$  individuals, we included an individual random effect on slope parameters  $\beta$ 's with

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

for  $i \in [1, I]$

With  $\beta$  being the parameter of interest describing the relative habitat selection strength of covariate  $X$  for the population,  $\tau$  represents the between individual variance, i.e. population heterogeneity, regarding habitat selection of covariate  $X$ .

## 2.4 Data integration - Integrated RSF-counts model

Despite RSF and Poisson GLM quantify habitat selection through generalized linear regression on covariates  $X$  that is similar to an IPP (Aarts et al., 2012; Northrup et al., 2022), Poisson GLM and RSF do not estimate the same absolute IPP intensity  $\lambda$ . In HSFs, the intercept is meaningless and estimates can only be interpreted as a relative and not absolute selection strength Avgar et al. (2017). Poisson GLM and HSFs then align to estimate relative intensity of the IPP ( $\beta X$ ), i.e.  $\beta$ s estimated under a use-availability data result are consistent with  $\beta$  estimated from Poisson GLM fitted to counts data in discrete space (Aarts et al., 2012). Then, we built an integrated model that jointly estimates  $\beta$ 's whose reflect the relative selection intensity in the IPP. The integrated model keeps the hierarchical structure of the Poisson GLM and the logit-regression structure of the RSF, hence including two IPP-like models with specific intercepts and shared habitat selection coefficients.

Poisson GLM submodel:

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

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

RSF submodel:

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

$$K \sim \text{Bernoulli}(\lambda_{RSF})$$

For details, see Rcodes. Note that intercepts  $\beta_0^c$  of the Poisson GLM submodel and  $\beta_0^r$  of the RSF submodel are different. While backtransformed parameters  $\lambda_{counts}$  and  $\lambda_{RSF}$  are not equal,  $\beta$ s reflecting the relative selection for habitat covariates are shared among the two submodels. Thus, we obtained maps of relative habitat selection intensity by representing  $\beta X$  for RSF, Poisson GLM, and integrated model (Figure 2A).

## 3 Applications

### 3.1 Simulations

To test the validity and robustness of our models, we fitted it to several simulations. First, we generated a discrete landscape of two random habitats *A* and *B* equally available with low spatial autocorrelation on a study area of 1000 x 1000 cells. We used `{localGibbs}` Github package (Michelot et al., 2019) to simulate the landscape. We simulated a fictive population of 1000 individuals who moved in this landscape during 300 steps and based on  $\Delta$  a pre-defined habitat selection of habitat *A* over habitat *B*.  $\Delta = 4$  reflects that an individual is 4 times more likely to select habitat *A* over habitat *B* if both are available in equal quantities. We simulated tracks based on a biased correlated random walk to mimic a central place forager behaviour (see Rcodes in supplementary materials, Benhamou (2014)). Then, we obtained a fictive population of 1000 individuals with 300 steps each, with observed selection ratio  $\hat{\Delta} = \frac{\text{number of steps in habitat A}}{\text{number of steps in habitat B}}$ .

Due to movement constraints and stochasticity, observed habitat selection  $\hat{\Delta}$  is always lower than  $\Delta$ . For example, when we used  $\Delta = 4$ , we had a  $\hat{\Delta} = 3.38$ , hence the slope coefficient of the covariate to be estimated was  $\beta = \log(\hat{\Delta}) = 1.21$ . Besides, without individual heterogeneity in habitat selection, modelling stochasticity led to  $sd(\hat{\Delta}_n) = 0.17$  between the 1000 individuals (see Supplementary material for details about the simplest simulations).

In the main batch of simulations, we included individual heterogeneity of habitat selection. We assigned to each individual  $\Delta_n \sim \text{Normal}(\Delta, \tau)$ , with  $\tau = 0.5$  the individual heterogeneity for habitat selection.

To mimic tracking data, we randomly selected  $I$  individual among the 1000 and used their 300 simulated tracks to fit the RSF process described in the previous section.

Then, to mimic the monitoring of counts data, we simulated zigzag line transect length using `{dsims}` R package (Marshall, 2023). To build our grid for Poisson GLM analyses, we segmented the line transects in cells of the same size as habitat grid-cells. We filled the grid-cells with the number of locations of the 1000 individuals that fell into the grid-cell. At this point, we did not account for any imperfect detection fixing  $p = 1$ . Based on these count data, we fitted Poisson GLM to estimate habitat selection of A over B.

In the simulations, we did not investigate the effect of the monitoring conditions considering no measurement error in the tracking dataset neither imperfect detection when collecting counts data. We explored the effect of the quantity of each data source to be included in the model on  $\beta$ . Among the 1000 individuals, we kept subsets of  $I = 2$ ,  $I = 5$ , and  $I = 20$  individuals to fit RSF without accounting for individual heterogeneity in habitat selection. We analysed line transect over a range from 2000 to 15000 cell-units sampling effort to fit Poisson GLMs. We fitted the integrated RSF-Poisson GLM model using an increasing amount of data from counts and RSF data (Figure 1). In the main text, we exhibited the results with structural individual heterogeneity and  $\Delta = 4$  ( $\log(\hat{\Delta}) = 1.21$ ) (See supplementary materials for further simulations scenarios).

### 3.2 Sandwich tern case study

We applied our modelling development to estimate habitat selection of sandwich terns (*Thalasseus sandvicensis*) in the Gulf of Lion, French Mediterranean Sea. To estimate habitat selection of sandwich terns, we considered linear and quadratic effect of bathymetry (i.e. sea floor depth), and distance to coastline as environmental covariates. We checked that distance to coastline and bathymetry are not correlated in the Gulf of Lion study area. Covariates were scaled before the analysis.

In 2021, 22 sandwich terns were equipped in Etang de Thau nesting colony near Montpellier (Southern France, see supplementary materials, which constituted our tracking dataset ([link to](#)

*movebank?*). We focused on habitat selection during the 2022 nesting period covering June to August. To assess marine habitat selection, we removed all locations on the land and resampled data to keep one location every hour. Filtering our dataset, we obtained **9807** locations from 16 individuals, other individuals being dead, out of the study area, or the tracking device stopped functioning. Then, we simulated 10 available points for each used location in our dataset following a negative exponential distribution centered on the nesting colony (see Appendix for Rcodes). We fitted a Bayesian RSF accounting for individual heterogeneity in sandwich tern habitat selection.

Then for each individual  $i$  and each point  $s$ , RSF model is:

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

with

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

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

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

where  $\beta_{0,i}, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2, \tau_3$  are to be estimated.

We used counts data from PELMED line transects monitoring program ([Laran et al., 2017](#)). PELMED is a large-scale monitoring program targeting small pelagic fisheries in the Gulf of Lion. In addition to fishes, they collect marine megafauna data, including seabirds data during one month every summer (see Supplementary materials for details about monitoring programs and datasets). We divided our study area in 471 contiguous hexagonal grid-cells (Figure 2, and supplementary materials). We fitted a Poisson GLM using counts data collected during 2017-2021 summers, hence considering five sampling occasions assuming no temporal variation of  $N$  the number of individual per grid-cell. We included sampling effort calculated as the transect length (in km) prospected in each grid-cell during each sampling occasion as an observation covariate affecting detection probability  $p$ .

$$\text{logit}(p_s) = \rho_0 + \rho_1 \text{Effort}$$

We estimated the IPP intensity  $\lambda$  using the same environmental covariates as the RSF model, i.e. a linear and quadratic effect of bathymetry and a linear effect of distance to coastline. As we used counts data collected during 2017-2021 summer, we must assume that Sandwich tern habitat selection remains constant during the 2017-2022 nesting periods regarding the bathymetry and the distance to coastline. Then, the Poisson GLM submodel is:

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

Finally, we implemented the integrated model joining the RSF and the Poisson GLM as described in section above.

### 3.3 Bayesian implementation

We fitted all models with NIMBLE R-package (de Valpine et al., 2017). We ran two MCMC chains with 100 000 iterations each, and we assessed convergence ensuring that Gelman test provide  $\hat{R} < 1.1$  (Gelman et al., 2013). R codes are available in supplementary materials (*mettre le lien*).

Regarding the simulations, we displayed the posterior distribution of  $\beta$  for all fitted model in Figure 1. In the main text, we displayed the results of one random scenario of simulations to have the RSF, the Poisson GLM, and the integrated model fitted to the same datasets, hence enabling to observe the effect of data integration without the “blur” of multiple data simulations. Concerning the sandwich tern case study, we displayed the posterior distribution of the slope parameters of covariates in Figure 2B. We predicted the relative habitat selection intensity plotting the linear predictor

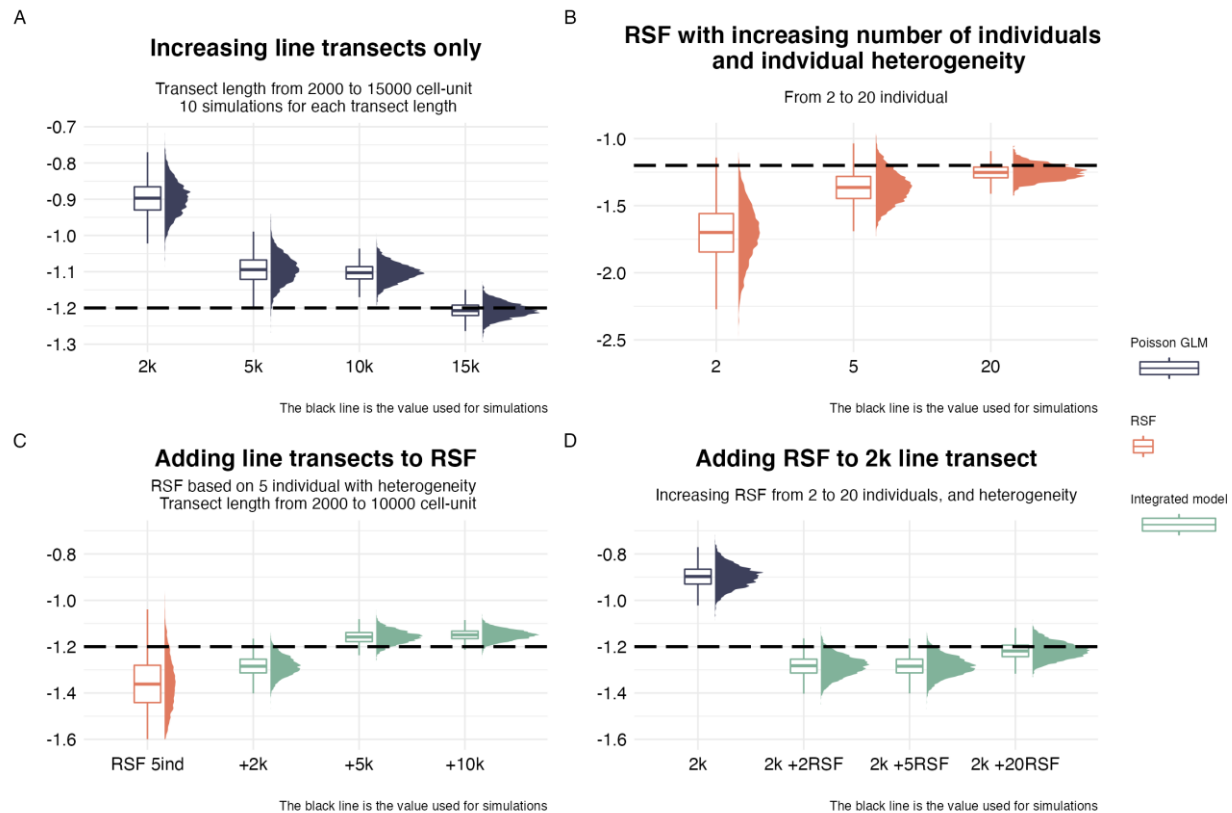
$$\theta = \beta_1 \text{Bathymetry} + \beta_2 \text{Bathymetry}^2 + \beta_3 \text{Distance to coastline}$$

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

## 4 Results

### 4.1 Simulations

RSF, Poisson GLM, and integrated models correctly detected the effect of habitat selection  $\beta$  used for simulations (Figures 1 and S1). Poisson GLM produced better precision and lower bias of  $\beta$  with increasing amount of data (Figure 1A), while increasing number of individuals included in the RSF also increased the precision and reduced the bias associated with habitat selection coefficient (Figure 1B). Overall, integrated models combining RSF and Poisson GLM increased precision and reduced bias associated with relative habitat selection estimated compared to models using each dataset alone (Figures 1C and 1D). Integrating an increasing amount of counts data to RSF models fitted with 5 individuals significantly reduced bias and increased precision except when adding the lowest 2k amount of counts data (Figure 1A). Similarly, integrating increasing number of individuals via an RSF to a Poisson GLM with a sampling effort of 2000 cell-unit line transect increased precision and reduced bias compared to the 2K Poisson GLM in isolation (Figure 1B).



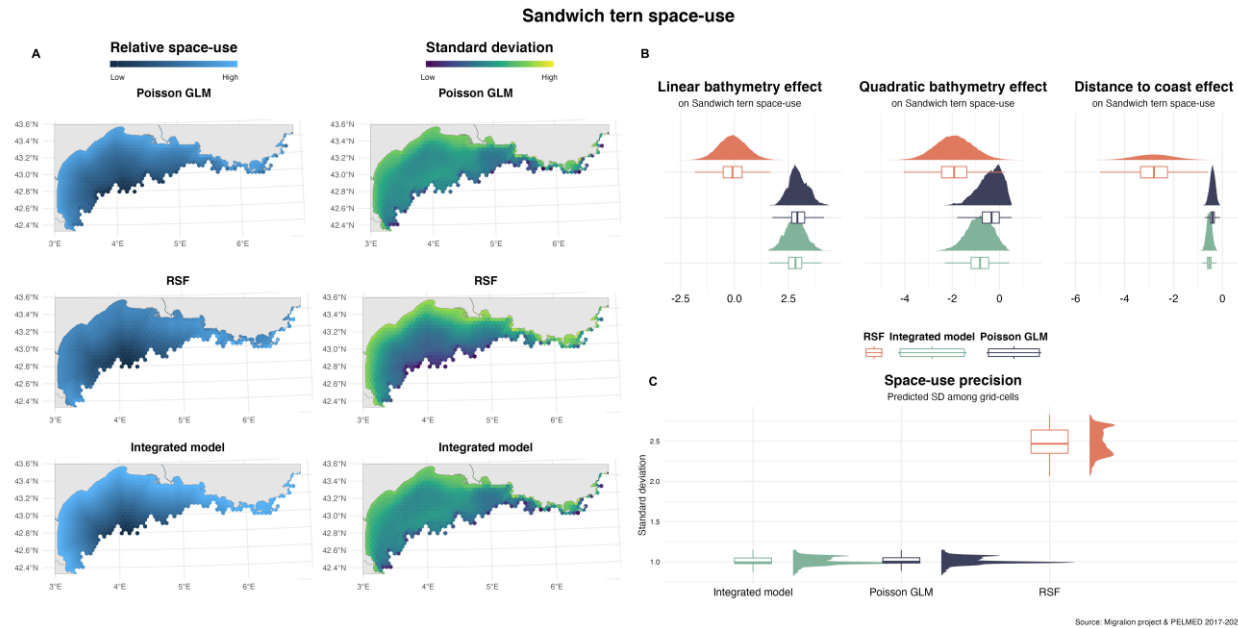
*Figure 1: Estimation of habitat selection coefficient from simulated data. To constitute our fictive population, we simulated 300 steps tracks for 1000 individuals over two discrete habitats A and B and following defined preference for habitat A. We fitted RSF, Poisson GLM, and an integrated RSF-Poisson GLM model to data subsets. A. Coefficients of habitat selection estimated via Poisson GLM models with increasing amount of data. B. Coefficients of selection estimated via RSF with increasing number of tracks. C. Based on a 5-individuals RSF, we added increasing line transect data to assess habitat selection. D. Based on a 2,000 cell-units line transects dataset, we fitted a Poisson GLM model to assess habitat selection and we fitted the integrated RSF-Poisson GLM model adding 2, 5, and 20 individuals to the RSF submodel.*

## 4.2 Sandwich terns in the French Mediterranean Sea

RSF, Poisson GLM, and integrated models estimated a significant negative effect of distance to coast on sandwich terns habitat selection (Figure 2B). RSF estimated a quadratic effect of bathymetry, which led as an optimum of space-use probability for mid-range sea depth (Figure 2A). However, Poisson GLM and integrated model estimated a linear effect of bathymetry and no significant quadratic effect, which reflected the coastal sandwich terns space-use prediction (Figure 2A-B). Overall, estimates produced by the integrated model were closer to estimates of the Poisson GLM model than to the estimates of the RSF models (Figure 2B). In terms of relative intensity of spatial use (Figure 2A), Poisson GLM predicted a higher use of shallower seas compared to mid-



range depth and pelagic seas. The integrated model predicted a high use of coastal and mid-range depths seas, possibly reflecting the influence of movement data compared to the Poisson GLM model coastal use prediction. All models predicted a low spatial use for pelagic and deep seas. When comparing standard deviation associated with relative space-use selection (Figure 2C), the integrated model displayed a similar precision to Poisson GLM model, RSF resulting in the lower precision (i.e. higher standard deviation).



**Figure 2: Estimation of Sandwich tern (*Thalasseus sandvicencis*) habitat selection in the Gulf of Lion, French Mediterranean Sea.** We fitted Resource Selection Function (RSF), Poisson GLM model, and an integrated model combining movement and counts data. We estimated habitat selection as a function of a linear and quadratic effect of bathymetry, and a linear effect of distance to coastline as environmental covariates. A. Relative sandwich terns space-use and associated standard deviation predicted by RSF, Poisson GLM, and the integrated model. B. Covariate effect on habitat selection estimated by RSF, Poisson GLM, and the integrated model. C. Distribution of standard deviation of sandwich terns space-use predictions among the grid-cells of the study area for RSF, Poisson GLM, and the integrated model.

## 5 Discussion

We showcased the integration of movement and counts data to estimate relative habitat selection. Our simulations and the sandwich tern case study supported the benefit of the integrated model under some biological conditions.

## 5.1 Simulations

The integrated model correctly inferred the simulated habitat selection when the data are abundant and unbiased (Figure 1 and Appendix X), but also when the data quantity is limited or when individual heterogeneity in habitat selection occurred (Figure 1). Simulations outlined that adding counts data to RSF or the opposite (i.e. adding movement data to Poisson GLM) improved the estimation of relative habitat selection compared to models using datasets in isolation (Figure 1C-D). All integrated models displayed an increased or equivalent precision and a reduced bias compared to single data models. While fitting Poisson GLM and RSF models to abundant data produced unbiased and precise estimates, we emphasized the benefit of the integrated models when data are limited. Due to the lack of data, the Poisson GLM fitted on 2000 cells and the two-individuals RSF model produced biased estimation (Figure 1A and 1B), whereas the integrated model with two-individuals RSF submodel and the 2000 cells Poisson GLM submodel significantly reduced the bias and increased the precision of the relative habitat selection estimation (Figure 1D). Our simulation study supported the trend that more data we have, the better the estimates (Figure 1A-B). However, RSF with 20 individuals produced unbiased and precise relative habitat selection estimates (Figure 1B) and adding 2000 cells count data fitted via a Poisson GLM produced did not increase precision of the integrated model (“2k + 20RSF” model, Figure 1D). We supported that caution has to be taken when performing data integration as adding biased data without explicitly accounting for the bias would lead to biased or unprecise ecological estimations compared to using an unbiased dataset alone (Simmonds et al., 2020). In our simulations with few data, Poisson GLM models produced an underestimation of  $\beta$  and RSF an overestimation of  $\beta$ . Then, the integrated model with few data produced a weighted-average estimation of  $\beta$  that performed better and produced lower bias compared to  $\beta$ s estimated from single data models (Figure 1D). When fitting models with limited data, the biased estimation of  $\beta$  reflected the stochasticity of low sampling effort, this sampling bias being reduced with increasing sampling effort. However, if few counts data and few movement data both lead to the under- or overestimation of  $\beta$ , it is likely that adding integrated model would not improve the estimation of  $\beta$ .

## 5.2 Sandwich terns

Applying a comparative approach between RSF, Poisson GLM, and integrated models to Sandwich terns in the Northwestern Mediterranean Sea, we exhibited that integrating movement and counts data benefit to the estimation of relative habitat selection (Figure 2C).

Poisson GLM and the integrated model produced closer estimation of habitat selection than RSF model. RSF detected a significant quadratic effect with an optimum depth, while the integrated model and counts models predicted a preference for coastal seas (Figure 2A). Quadratic effect of bathymetry are reported for similar seabirds species (Pratte et al., 2021), while a significant preference for shallow and close to colony waters is also documented on sandwich terns (van Bemmelen et al., 2023). Bathymetry is known to affect sandwich terns behavior and spatial use, however its role remains unclear (Fijn et al., 2022) and seems to depend on the local context and the spatial scale studied. Besides, most studies found a negative effect of increasing distance to coastline or distance from the colony for nesting sandwich tern (Fauchald et al., 2021; Fijn et al., 2022; van Bemmelen et al., 2023), which support the results of the three models. The effect of distance to the coast in the integrated model were closest to the estimation of Poisson GLM than to the RSF model. Overall, relative habitat selection estimates of the integrated model were mainly driven by counts data (Figure 2) and the precision associated with relative selection strength is equivalent between Poisson GLM and the integrated model. However, compared to predictions of the Poisson GLM, maps of the integrated model predicted a more intense use of mid-range pelagic seas, possibly reflecting the effect of RSF submodel. However, we emphasized that data integration have potential conservation implications for the sandwich terns study as the integrated model predicted a more pelagic habitat selection compared to the Poisson GLM (and to RSF model, Figure 2A).

## 5.3 Condition of use and possible extensions

Our modelling development contributed to the formal integration of movement and counts models, which continue the ongoing path towards increasing links between species distribution and movement modelling Converse et al. (2022).

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

tracked individual to the population level remain challenging (Hooten et al., 2016; Potts & Börger, 2023; Signer et al., 2017). For our sandwich terns case study, we must assume that relative habitat selection regarding bathymetry and distance to coastline span to all individuals nesting in the different colonies of the study area and remain consistent throughout the time period when the data were collected. The application of our integrated RSF-counts model is conditional to the respect of biological hypothesis regarding movement data. Although we supported that such hypothesis was respected for sandwich tern in the Gulf of Lion regarding static covariates, it is the duty of the researcher to verify the applicability of our modelling approach.

The convergence of Poisson GLM and RSF under the IPP framework allows estimating relative habitat selection,  $\exp(\beta X)$ , and not absolute presence probability (Northrup et al., 2022). The predicted intensity  $\exp(\beta X)$  in a grid-cell should be interpreted as the relative use of the grid-cell conditional on being available to the animal/population. Estimating relative habitat selection strength is common and relevant in many ecological contexts, including for conservation purpose where the relative pattern of space-use intensity is often more informative than the absolute value within a grid-cell (e.g.  $\psi = 0.3$  or  $\psi = 0.8$ ). However, when looking for temporal trends in habitat selection, absolute value may be needed and then estimations of  $\beta X$  remain insufficient. We suggested that absolute habitat selection can be derived from an integrated model. Through the IPP framework, a Poisson GLM produces an absolute estimation that reflects the expected number of individual per area unit, i.e.  $\lambda$ . Integrating RSF and Poisson GLM helped to inform the slopes of covariates that are shared but not the intercept that is specific to each submodel. Then, one should obtain absolute estimation of density reconstructing the  $\lambda_{\text{counts}}$  using shared slopes but specific Poisson GLM intercept.

Overall, the ability to combine movement and counts data should not return in a systematic application of integrated modelling as soon as both datasets are available. Integrated models provide a useful statistical option to assess the relative habitat selection, but not in every situation and one should wonder whether it is worth integrating multiple datasets. When one dataset is abundant and produced precise estimates, engaging in data integration might not significantly improve the outputs (Figure 1). However, if interested in a large-scale habitat selection assessment where by both movement data and population counts covered distinct spatial areas, integrated models can benefit from both data sources and capture complementary aspects of the targeted

population (Lauret et al., 2022). Then, prior to engage in integrated models development, a reflexive thought about the pros and cons of data integration is needed and constitutes an important work question that must be done when considering the modelling tools available (Simmonds et al., 2020). We acknowledged that further investigation of our modelling development would help to better understand to what extent and in which situations combining RSF and Poisson GLM is beneficial, hence providing guidance for ecological applications.

While we built our integrated model based on RSF and Poisson GLM, many straightforward extensions are possible for models using regression methods to estimate habitat selection.

For movement data, SSFs and integrated SSFs (iSSFs) constitute widely used methods to estimate habitat selection, hence representing promising candidates for integration with population data (Avgar et al., 2015; McLoughlin et al., 2010). However, prediction from coefficient  $\beta$  do not converge between SSFs and RSFs fitted to the same data, except when the scale of habitat available at each time step increases (Michelot et al., 2019; Moorcroft & Barnett, 2008), and when the time scale of observations is identical (Hooten et al., 2016). Recent methodological developments unified individual and population-level inferences from SSF via simulations (Signer et al., 2017), and via a Markov Chain Monte Carlo (MCMC) movement model implemented in a SSF (Michelot et al., 2019). However, under the hypothesis that i) sampled individuals and time periods are representative of the long-term population state; ii) landscape availability is constant for animals, i.e. if a tracked animal is able to reach any point of its home range from its current position or if spatial autocorrelation in the landscape is low (Signer et al., 2017), and iii) changes in behavioural states are not affecting habitat selection (McClintock & Lander, 2023), inferences from SSF and RSF should converge (Signer et al., 2017). If these conditions are fulfilled, it is possible to integrate SSF with Poisson GLM to obtain joint inference of relative habitat selection coefficient  $\beta$ . We acknowledge that conditions to extrapolate SSF estimates to population-level predictions would be rarely respected in real ecological data and to ensure that hypothesis are verified is the reasearcher's duty. Following Signer et al. (2017) work, we questioned how much the biased induced by a naïve SSF prediction ( $\exp(\beta X)$ ) compared to a RSF prediction would be acceptable when being interested in large-scale ecological inference, e.g. for conservation purpose. In the appendices, we built a Bayesian SSF to assess habitat selection from multiple individuals fitted on our simulation and on our sandwich tern data (Muff et al., 2020). The estimations of  $\beta$  were similar between RSF

and naïve SSF for simulated data, although being different when applied to sandwich terns data. Naïve SSF estimate produced more precise estimation of  $\beta$  whereas resulting in close space-use prediction (i.e. quadratic effect of bathymetry, negative effect of distance to coastline). Besides, we briefly outlined that SSF and MCMC SSF (Michelot et al., 2019) produced similar estimation of *beta* on our simulated data, but diverged when applied to sandwich terns data (Appendix X).

When considering population-level data, species distribution models widely used presence-absence (or presence-only) data to estimate habitat selection (e.g. occupancy models, Mackenzie et al. (2002), Tyre et al. (2003)). A high resolution grid and the use of complementary loglog link function facilitate the convergence of logistic regression fitted to presence-absence data with the slope parameters estimated with an IPP fitted to count data (Aarts et al., 2012). Thus, occupancy models represent possible candidates to integrate with movement model to estimate relative habitat selection. In the appendices, we developed and fitted an integrated SSF-Occupancy model to our simulations and sandwich tern data (Appendix X). Among species distribution models, one can consider to adapt the RSF-counts model to any niche model, presence-only model, occupancy models, or integrated species distribution models (Farr et al., 2020; Mackenzie et al., 2002; D. A. W. Miller et al., 2019; Royle, 2004; Tyre et al., 2003).

## 5.4 Conclusion

Overall, our integrated model reflected the complementary of movement and counts data for habitat selection studies. Individual tracking provided a large amount of location data but on a restricted subset of the population while population monitoring programs can cover large spatial extent despite collecting fewer data. In the recent years, animal tracking and counts monitoring programs are increasingly available although often targeting different ecological questions and being supported by different people and motivation (Lindenmayer et al., 2022; Nathan et al., 2022). However, under some conditions, data integration of movement and counts constitutes a relevant opportunity to investigate habitat selection. We supported that our integrated model combining existing movement and counts datasets can be a promising tool that would echo many ecological contexts.



## References

- van Bemmelen, R. S. A., Leemans, J. J., Collier, M. P., Green, R. M. W., Middelveld, R. P., Thaxter, C. B., & Fijn, R. C. (2023). Avoidance of offshore wind farms by Sandwich Terns in the North Sea increases with turbine density. *Ornithological Applications*, duad055. <https://doi.org/10.1093/ornithapp/duad055>
- Courbin, N., Besnard, A., Boncourt, E., & Grémillet, D. (2023). *Flight heights of Scopoli's shearwaters *Calonectris diomedea* in the context of offshore wind farm developments* (p. 2023.05.14.540698). bioRxiv. <https://doi.org/10.1101/2023.05.14.540698>
- Fradin, G., & Chamaillé-Jammes, S. (2023). Hogs sleep like logs: Wild boars reduce the risk of anthropic disturbance by adjusting where they rest. *Ecology and Evolution*, 13(7), e10336. <https://doi.org/10.1002/ece3.10336>
- Marshall, L. (2023). *Dsims: Distance Sampling Simulations*. Distance development.
- McClintock, B. T., & Lander, M. E. (2023). A multistate Langevin diffusion for inferring behavior-specific habitat selection and utilization distributions. *Ecology*, n/a(n/a), e4186. <https://doi.org/10.1002/ecy.4186>
- Potts, J. R., & Börger, L. (2023). How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection. *Journal of Animal Ecology*, 92(1), 16–29. <https://doi.org/10.1111/1365-2656.13832>
- Lindenmayer, D. B., Lavery, T., & Scheele, B. C. (2022). Why We Need to Invest in Large-Scale, Long-Term Monitoring Programs in Landscape Ecology and Conservation Biology. *Current Landscape Ecology Reports*. <https://doi.org/10.1007/s40823-022-00079-2>
- Fijn, R., Thaxter, C., Aarts, G., Adema, J., Middelveld, R., & Van Bemmelen, R. (2022). Relative effects of static and dynamic abiotic conditions on foraging behaviour in breeding Sandwich terns. *Marine Ecology Progress Series*, 692, 137–150. <https://doi.org/10.3354/meps14076>
- Nathan, R., Monk, C. T., Arlinghaus, R., Adam, T., Alós, J., Assaf, M., Baktoft, H., Beardsworth, C. E., Bertram, M. G., Bijleveld, A. I., Brodin, T., Brooks, J. L., Campos-Candela, A., Cooke, S. J., Gjelland, K. Ø., Gupte, P. R., Harel, R., Hellström, G., Jeltsch, F., ... Jarić, I. (2022). Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science*, 375(6582), eabg1780. <https://doi.org/10.1126/science.abg1780>
- Strebel, N., Kéry, M., Guélat, J., & Sattler, T. (2022). Spatiotemporal modelling of abundance from multiple data sources in an integrated spatial distribution model. *Journal of Biogeography*, jbi.14335. <https://doi.org/10.1111/jbi.14335>
- Chandler, R. B., Crawford, D. A., Garrison, E. P., Miller, K. V., & Cherry, M. J. (2022). Modeling abundance, distribution, movement and space use with camera and telemetry data. *Ecology*, 103(10), e3583. <https://doi.org/10.1002/ecy.3583>
- Converse, S. J., McClintock, B. T., & Conn, P. B. (2022). Special Feature: Linking capture-recapture and movement. *Ecology*, n/a(n/a), e3770. <https://doi.org/10.1002/ecy.3770>



- Lauret, V., Labach, H., Turek, D., Laran, S., & Gimenez, O. (2022). Integrated spatial models foster complementarity between monitoring programmes in producing large scale bottlenose dolphin indicators. *Animal Conservation*, 11. <https://doi.org/10.1111/acv.12815>
- Matthiopoulos, J., Wakefield, E., Jeglinski, J. W. E., Furness, R. W., Trinder, M., Tyler, G., Mccluskie, A., Allen, S., Braithwaite, J., & Evans, T. (2022). Integrated modelling of seabird-habitat associations from multi-platform data: A review. *Journal of Applied Ecology*, 59(4), 909–920. <https://doi.org/10.1111/1365-2664.14114>
- McClintock, B. T., Abrahms, B., Chandler, R. B., Conn, P. B., Converse, S. J., Emmet, R. L., Gardner, B., Hostetter, N. J., & Johnson, D. S. (2022). An integrated path for spatial capturerecapture and animal movement modeling. *Ecology*, 103(10), e3473. <https://doi.org/10.1002/ecy.3473>
- Northrup, J. M., Vander Wal, E., Bonar, M., Fieberg, J., Laforge, M. P., Leclerc, M., Prokopenko, C. M., & Gerber, B. D. (2022). Conceptual and methodological advances in habitat-selection modeling: Guidelines for ecology and evolution. *Ecological Applications*, 32(1), e02470. <https://doi.org/10.1002/eap.2470>
- Fauchald, P., Tarroux, A., Amélineau, F., Bråthen, V., Descamps, S., Ekker, M., Helgason, H., Johansen, M., Merkel, B., Moe, B., Åström, J., Anker-Nilssen, T., Bjørnstad, O., Chastel, O., Christensen-Dalsgaard, S., Danielsen, J., Daunt, F., Dehnhard, N., Erikstad, K., ... Strøm, H. (2021). Year-round distribution of Northeast Atlantic seabird populations: Applications for population management and marine spatial planning. *Marine Ecology Progress Series*, 676, 255–276. <https://doi.org/10.3354/meps13854>
- Robira, B., Benhamou, S., Masi, S., Llaurens, V., & Riotte-Lambert, L. (2021). Foraging efficiency in temporally predictable environments: Is a long-term temporal memory really advantageous? *Royal Society Open Science*, 8(9), 210809. <https://doi.org/10.1098/rsos.210809>
- Pratte, I., Ronconi, R., Craik, S., & McKnight, J. (2021). Spatial ecology of endangered roseate terns and foraging habitat suitability around a colony in the western North Atlantic. *Endangered Species Research*, 44, 339–350. <https://doi.org/10.3354/esr01108>
- Farr, M. T., Green, D. S., Holekamp, K. E., & Zipkin, E. F. (2021). Integrating distance sampling and presence-only data to estimate species abundance. *Ecology*, 102(1), e03204. <https://doi.org/10.1002/ecy.3204>
- Fieberg, J., Signer, J., Smith, B., & Avgar, T. (2021). A “How to” guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, 90(5), 1027–1043. <https://doi.org/10.1111/1365-2656.13441>
- Gamelon, M., Nater, C. R., Baubet, É., Besnard, A., Touzot, L., Gaillard, J.-M., Lebreton, J.-D., & Gimenez, O. (2021). Efficient use of harvest data: A size-class-structured integrated population model for exploited populations. *Ecography*, 44(9), 1296–1310. <https://doi.org/10.1111/ecog.05738>

- Zipkin, E. F., Zylstra, E. R., Wright, A. D., Saunders, S. P., Finley, A. O., Dietze, M. C., Itter, S., & Tingley, M. W. (2021). Addressing data integration challenges to link ecological processes across scales. *Frontiers in Ecology and the Environment*, 9.
- Farr, M. T., Green, D. S., Holekamp, K. E., & Zipkin, E. F. (2020). Integrating distance sampling and presence-only data to estimate species abundance. *Ecology*.  
<https://doi.org/10.1002/ecy.3204>
- Simmonds, E. G., Jarvis, S. G., Henrys, P. A., Isaac, N. J. B., & O'Hara, R. B. (2020). Is more data always better? A simulation study of benefits and limitations of integrated distribution models. *Ecography*, ecog.05146. <https://doi.org/10.1111/ecog.05146>
- Muff, S., Signer, J., & Fieberg, J. (2020). Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation. *Journal of Animal Ecology*, 89(1), 80–92.  
<https://doi.org/10.1111/1365-2656.13087>
- Kéry, M., & Royle, J. (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in r and bugs: Volume 2: Dynamic and advanced models: Vols. Ch 10* (1st ed.). Elsevier. <https://doi.org/10.1016/C2015-0-04070-9>
- Zipkin, E. F., Inouye, B. D., & Beissinger, S. R. (2019). Innovations in data integration for modeling populations. *Ecology*, e02713. <https://doi.org/10.1002/ecy.2713>
- Fletcher, R. J., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology*, e02710. <https://doi.org/10.1002/ecy.2710>
- Miller, D. A. W., Pacifici, K., Sanderlin, J. S., & Reich, B. J. (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution*, 10(1), 22–37. <https://doi.org/10.1111/2041-210X.13110>
- Michelot, T., Blackwell, P. G., & Matthiopoulos, J. (2019). Linking resource selection and step selection models for habitat preferences in animals. *Ecology*, 100(1), e02452.  
<https://doi.org/10.1002/ecy.2452>
- Laran, S., Pettex, E., Authier, M., Blanck, A., David, L., Dorémus, G., Falchetto, H., Monestiez, P., Van Canneyt, O., & Ridoux, V. (2017). Seasonal distribution and abundance of cetaceans within French waters- Part I: The North-Western Mediterranean, including the Pelagos sanctuary. *Deep Sea Research Part II: Topical Studies in Oceanography*, 141, 20–30.  
<https://doi.org/10.1016/j.dsr2.2016.12.011>
- Avgar, T., Lele, S. R., Keim, J. L., & Boyce, M. S. (2017). Relative Selection Strength: Quantifying effect size in habitat- and step-selection inference. *Ecology and Evolution*, 7(14), 5322–5330. <https://doi.org/10.1002/ece3.3122>
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2), 403–413.

673 Louvrier, J., Duchamp, C., Lauret, V., Marboutin, E., Cubaynes, S., Choquet, R., Miquel, C., &  
 674 Gimenez, O. (2017). Mapping and explaining wolf recolonization in France using dynamic  
 675 occupancy models and opportunistic data. *Ecography*.  
 676 Signer, J., Fieberg, J., & Avgar, T. (2017). Estimating utilization distributions from fitted step-  
 677 selection functions. *Ecosphere*, 8(4), e01771. <https://doi.org/10.1002/ecs2.1771>  
 678 Hooten, M. B., Buderman, F. E., Brost, B. M., Hanks, E. M., & Ivan, J. S. (2016). Hierarchical  
 679 animal movement models for population-level inference. *Environmetrics*, 27(6), 322–333.  
 680 <https://doi.org/10.1002/env.2402>  
 681 Avgar, T., Potts, J. R., Lewis, M. A., & Boyce, M. S. (2015). Integrated step selection analysis:  
 682 Bridging the gap between resource selection and animal movement. *Methods in Ecology and*  
 683 *Evolution*, 12.  
 684 Thurfjell, H., Ciuti, S., & Boyce, M. S. (2014). Applications of step-selection functions in  
 685 ecology and conservation. *Movement Ecology*, 2(1), 4. [https://doi.org/10.1186/2051-3933-](https://doi.org/10.1186/2051-3933-2-4)  
 686 [2-4](https://doi.org/10.1186/2051-3933-2-4)  
 687 Benhamou, S. (2014). Of scales and stationarity in animal movements. *Ecology Letters*,  
 688 17(3), 261–272. <https://doi.org/10.1111/ele.12225>  
 689 Fithian, W., & Hastie, T. (2013). Finite-sample equivalence in statistical models for  
 690 presence-only data. *The Annals of Applied Statistics*, 7(4), 1917–1939.  
 691 <https://doi.org/10.1214/13-AOAS667>  
 692 Miller, D. L., Burt, M. L., Rexstad, E. A., & Thomas, L. (2013). Spatial models for distance  
 693 sampling data: Recent developments and future directions. *Methods in Ecology and*  
 694 *Evolution*, 4(11), 1001–1010. <https://doi.org/10.1111/2041-210X.12105>  
 695 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian*  
 696 *data analysis*. Chapman and Hall/CRC.  
 697 Sollmann, R., Gardner, B., Parsons, A. W., Stocking, J. J., McClintock, B. T., Simons, T. R.,  
 698 Pollock, K. H., & O'Connell, A. F. (2013). A spatial markresight model augmented with  
 699 telemetry data. *Ecology*, 94(3), 553–559. <https://doi.org/10.1890/12-1256.1>  
 700 Aarts, G., Fieberg, J., & Matthiopoulos, J. (2012). Comparative interpretation of count,  
 701 presence-absence and point methods for species distribution models: *Species Distribution*  
 702 *as Spatial Point Process*. *Methods in Ecology and Evolution*, 3(1), 177–187.  
 703 <https://doi.org/10.1111/j.2041-210X.2011.00141.x>  
 704 McLoughlin, P. D., Morris, D. W., Fortin, D., Vander Wal, E., & Contasti, A. L. (2010).  
 705 Considering ecological dynamics in resource selection functions. *Journal of Animal Ecology*,  
 706 79(1), 4–12. <https://doi.org/10.1111/j.1365-2656.2009.01613.x>  
 707 Warton, D. I., & Shepherd, L. C. (2010). Poisson Point Process Models Solve the "Pseudo-  
 708 Absence Problem" for Presence-Only Data in Ecology. *The Annals of Applied Statistics*, 4(3),  
 709 1383–1402. <https://www.jstor.org/stable/29765559>

710 Rota, C. T., Fletcher Jr, R. J., Dorazio, R. M., & Betts, M. G. (2009). Occupancy estimation and  
 711 the closure assumption. *Journal of Applied Ecology*. [https://doi.org/10.1111/j.1365-](https://doi.org/10.1111/j.1365-2664.2009.01734.x)  
 712 [2664.2009.01734.x](https://doi.org/10.1111/j.1365-2664.2009.01734.x)

713 Elith, J., & Leathwick, J. R. (2009). *Species Distribution Models: Ecological Explanation and*  
 714 *Prediction Across Space and Time*. 24.

715 Moorcroft, P. R., & Barnett, A. (2008). Mechanistic Home Range Models and Resource  
 716 Selection Analysis: A Reconciliation and Unification. *Ecology*, 89(4), 1112–1119.  
 717 <https://doi.org/10.1890/06-1985.1>

718 MacKenzie, D. I. (Ed.). (2006). *Occupancy estimation and modeling: Inferring patterns and*  
 719 *dynamics of species*. Elsevier.

720 Royle, J. A. (2004). N-Mixture Models for Estimating Population Size from Spatially  
 721 Replicated Counts. *Biometrics*, 60(1), 108–115. [https://doi.org/10.1111/j.0006-](https://doi.org/10.1111/j.0006-341X.2004.00142.x)  
 722 [341X.2004.00142.x](https://doi.org/10.1111/j.0006-341X.2004.00142.x)

723 Royle, J. A., & Nichols, J. D. (2003). ESTIMATING ABUNDANCE FROM REPEATED  
 724 PRESENCEABSENCE DATA OR POINT COUNTS. *Ecology*, 84(3), 777–790.  
 725 [https://doi.org/10.1890/0012-9658\(2003\)084\[0777:EAFRPA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2)

726 Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003).  
 727 Improving Precision and Reducing Bias in Biological Surveys: Estimating False-Negative  
 728 Error Rates. *Ecological Applications*, 13(6), 1790–1801. <https://doi.org/10.1890/02-5078>

729 Mackenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A.  
 730 (2002). Estimating site occupancy rates when detection probabilities are less than one.  
 731 *Ecology*, 83(8), 8.

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

735 Johnson, D. H. (1980). The Comparison of Usage and Availability Measurements for  
 736 Evaluating Resource Preference. *Ecology*, 61(1), 65–71. <https://doi.org/10.2307/1937156>