ARTICLE

Methods, Tools, and Technologies

ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

Integrating individual tracking data and spatial surveys to improve estimation of animal spatial distribution

Valentin Lauret[1,2] | **Nicolas Courbin**[1] | **Olivier Scher**[3] | **Aurélien Besnard**[4]

[1]CEFE, University of Montpellier, CNRS, EPHE, IRD, Montpellier, France

[2]France Energies Marines, Plouzané, France

[3]CEN Occitanie, Montpellier, France

[4]CEFE, University of Montpellier, CNRS, EPHE PSL University, IRD, Montpellier, France

Correspondence
Valentin Lauret
Email: valentin.lauret@gmail.com

Handling Editor: Sunshine A. Van Bael

Abstract

Tracking data and spatial surveys (e.g., counts) contribute to understanding animal distribution despite highlighting complementary aspects of habitat selection, from detailed insights on few individuals to raw inferences for the population, respectively. Here, we showcased how to combine individual tracking and count data to estimate habitat selection at the population level. We developed an integrated model that provides a joint estimation of habitat selection for tracking data fitted with a resource selection function (RSF) and count data fitted with a Poisson generalized linear model (GLM), both respecting the statistical conditions for converging with an inhomogeneous Poisson point process. We tested our integrated habitat selection model using simulated movement data and a real case study of GPS-tracked Sandwich terns (*Thalasseus sandvicensis*) in the French Mediterranean Sea. Simulations showed that the integrated model correctly estimated habitat selection coefficients and benefited from both data sources with better accuracy and precision than RSF and Poisson GLM alone, especially when data are limited. Overall, our study formalized an easy-to-use approach for the integration of tracking and count data to estimate habitat selection, contributing to a promising research avenue, since individual tracking and spatial survey monitoring are abundant in many ecological contexts.

KEYWORDS

data integration, habitat selection, inhomogeneous Poisson point process, movement ecology, resource selection function, species distribution models

INTRODUCTION

Telemetry tracking and spatial surveys (e.g., population counts) provide insightful spatial information to understand the ecological drivers shaping animal distributions. A key process underlying animal space use is habitat selection, that is, the process by which an individual disproportionately uses environmental features (Boyce & McDonald, 1999; Manly et al., 2002; Muff et al., 2020; Northrup et al., 2022). Within the last two decades, the increasing accessibility of animal tracking devices (e.g., GPS) and of population monitoring techniques (e.g., transect surveys) makes tracking data and counts widely available to estimate habitat selection and patterns of spatial distribution (Matthiopoulos et al., 2022; Northrup et al., 2022). Such data may also coexist in ecological studies, for example,

management of wild ungulate populations (Chandler et al., 2022; Fradin & Chamaillé-Jammes, 2023; Gamelon et al., 2021) and characterization of habitat selection of seabirds (Yamamoto et al., 2015). Importantly, the availability of both spatial surveys and individual tracking data for the same study area may improve our understanding of animal space use as they might allow us to capture complementary aspects of habitat selection. Individual animal tracking data and spatial surveys indeed focus on different ecological scales due to fundamental differences in the nature of the data they record (Lindenmayer et al., 2022; Nathan et al., 2022). Individual tracking typically provides a large amount of data at the individual scale and on a limited sample of the population (few individuals) while spatial survey monitoring programs theoretically sample all individuals of the population despite collecting fewer data. Given their potential complementarity to study habitat selection, combining individual tracking and population count data sources in a joint statistical framework to simultaneously estimate habitat selection is appealing (Blackwell & Matthiopoulos, 2024; Buderman et al., 2025; Matthiopoulos et al., 2022).

Integrating (or combining) multiple datasets is an intense research avenue of current statistical ecology (Kéry & Royle, 2020; Matthiopoulos et al., 2022; Zipkin et al., 2021). The root principle of most integrated models is that multiple datasets describe a common underlying ecological process in which at least one parameter is shared between these datasets, which are linked via a joint likelihood to each single dataset analysis (Fletcher et al., 2019; Kéry & Royle, 2020). The two main advantages of integrated models are (1) increased precision of parameters of interest, and (2) in some situations, the ability to estimate parameters that could not be inferred from each data source alone (Zipkin et al., 2019). Methodological developments on data integration have made it possible to combine many different data types (reviewed in Kéry & Royle, 2020). Counts, presence-only, and/or presence–absence data collected from multiple monitoring programs have been used to fit integrated species distribution models estimating a joint abundance or distribution (Matthiopoulos et al., 2022; Miller et al., 2019; Strebel et al., 2022). Recent methodological developments have permitted presence-only data to be combined with population analysis that classically relies on standardized protocols (Farr et al., 2020; Koshkina et al., 2017). Individual tracking data have also been frequently considered for data integration, for instance within spatial capture-recapture models to better inform the spatial distribution of the individuals' activity centers (Converse et al., 2022; Tenan et al., 2017), or into integrated populations models to assess survival and demographic parameters (Kendall et al., 2023; Regehr et al., 2018).

More recently, individual tracking data and opportunistic data from fisheries-bycatch have been combined to assess the intensity of space use (Liang et al., 2023).

To date, two studies have formally integrated spatial survey and tracking data to estimate habitat selection (Blackwell & Matthiopoulos, 2024; Buderman et al., 2025). They both provide a framework to deal with the fundamental differences in the nature of estimated habitat selection from tracking data and survey data, that is, analyzing animal movement at a local scale does not necessarily align with population distributions at a larger scale (Fieberg et al., 2021; Michelot et al., 2019). Notably, Blackwell and Matthiopoulos (2024) built on the methodological developments that have enabled scaling up habitat selection estimated by movement models with tracking data, such as step selection functions (SSFs, Fortin et al., 2005; Muff et al., 2020) to population distributions (Michelot et al., 2019). Indeed, while SSFs allow researchers to make the most of all tracking data by explicitly modeling movement characteristics, extra steps are mandatory to translate movement dynamics into habitat selection at the population level. These extra steps are highly technically challenging (solving the master equation: Potts et al., 2014, Gibbs sampling: Michelot et al., 2019, partial differential equation model: Potts & Schlägel, 2020), especially for their implementation by ecological researchers. Alternatively, the resource selection function (RSF) constitutes a straightforward approach to model habitat selection at the population scale (Boyce & McDonald, 1999; Manly et al., 2002), provided that serial dependency has been taken into account (Fieberg et al., 2021). Considering that RSF is one of the approaches used by most ecological researchers to estimate large-scale habitat selection from tracking data, we developed a straightforward integration of count surveys and tracking data using RSF rather than SSF.

A promising avenue to perform a joint analysis of individual tracking and spatial survey data (e.g., population counts) relies on the statistical framework of inhomogeneous Poisson point process (IPP). IPP describes the intensity function of a spatial pattern of points as a log-linear function of environmental covariates, being widely used tools to model spatial ecological processes and estimate habitat selection using individual tracking or population count data (Aarts et al., 2012; Illian, 2019; Johnson et al., 2013; Renner et al., 2015; Warton & Aarts, 2013). RSF contrasts environmental predictors at observed individual locations (e.g., GPS, Argos, or GLS locations) with those at available locations using logistic regression and estimates the exponential of the linear predictor (Boyce & McDonald, 1999; Manly et al., 2002; Muff et al., 2020; Northrup et al., 2022). Therefore, RSF

estimates the relative density of observed locations in space depending on environmental predictors for individuals (Aarts et al., 2012; Fieberg et al., 2021; Johnson et al., 2013; Muff et al., 2020). Under some circumstances (see next section), RSF and IPP likelihoods converge and regression coefficients are asymptotically equivalent (Aarts et al., 2012; Fieberg et al., 2021; Fithian & Hastie, 2013; Warton & Shepherd, 2010). Count data can also be analyzed with an IPP using log-linear regression of environmental covariates and a Poisson draw, hereafter Poisson generalized linear model (GLM) (Aarts et al., 2012). In this situation, the point pattern represents the density of individuals from the studied population (Matthiopoulos et al., 2022; Renner et al., 2015). Then, specifying RSF and Poisson GLM as the intensity function of an IPP gives a similar significance to the regression coefficients (except the intercept), also called habitat selection coefficients (Aarts et al., 2012; Fieberg et al., 2021), and provides the same information on the relative intensity of use by individuals for the studied population. Building on the work of Aarts et al. (2012) and Matthiopoulos et al. (2022) on the IPP formulation of RSF and Poisson GLM models, and considering their equivalence, we developed an integrated habitat selection model combining spatial survey and individual tracking data, complementary to those recently formalized (Blackwell & Matthiopoulos, 2024; Buderman et al., 2025).

In this article, we detailed how RSF and Poisson GLM must be designed as the intensity function of an IPP to appropriately estimate habitat selection coefficients and proposed a general formulation for an integrated model. Second, we assessed the performances of our integrated model using simulated movement data and applied it to estimate habitat selection of Sandwich terns (*Thalasseus sandvicensis*) in the Northwestern Mediterranean Sea. Finally, we discussed the conditions of application of our modeling approach while suggesting future extensions.

TOWARD AN INTEGRATED HABITAT SELECTION MODEL

Inhomogeneous Poisson point process as a unifying framework

An IPP is a model that describes 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 S with an intensity function λ_s that returns the expected number of points at every infinitesimal location s in S. The expected intensity λ is written as a log-linear function of environmental predictor X:

$$\log(\lambda_s) = \beta_0 + \sum \beta X_s, \qquad (1)$$

where parameters to be estimated are: β_0, the density intercept; and βs, the regression coefficients of the environmental predictor X.

The spatial ecological processes describing habitat selection can be modeled via an IPP using tracking or count data (Aarts et al., 2012; Johnson et al., 2013; Renner et al., 2015). Since both methods estimate habitat selection as a linear combination of environmental (biotic or abiotic) predictors, one can compare the relative habitat selection strength estimated through the regression coefficients of covariates, the βs.

Count data—Poisson generalized linear model

To estimate habitat selection from count data, we fitted a generalized linear model with a Poisson link function over a discrete geographical space where the expected number of individuals in a grid-cell follow the λ_s intensity function of an IPP (Aarts et al., 2012). We considered count data collected over multiple grid-cells as a count dataset N. N_s represents the expected counts in grid-cell s and is estimated as a Poisson draw in λ_s the IPP intensity at s.

$$N_s \sim \text{Poisson}(\lambda_s). \qquad (2)$$

One can extend the Poisson GLM to account for imperfect detection via a hierarchical model formulation (Royle & Nichols, 2003). Let us name Y_s the observed count at s. We estimated N_s the latent number of points in the grid-cell s through a binomial draw with individual detection probability p. Then, Y_s links with the IPP intensity λ through Equations (2) and (3).

$$Y_s \sim \text{Binomial}(N_s, p), \qquad (3)$$

where p can be grid-cell specific and estimated as a logistic regression of covariates (see Sandwich tern application for an example). If grid-cells are geographically closed (i.e., no individual moved in and out of the grid-cells), the intercept of the Poisson GLM, β_0, represents the expected number of individuals where all covariates equal 0 (or equal to mean if covariates are standardized).

However, it is unlikely the case for seabirds because their use of space is driven by ecological and environmental factors, and occupied locations are used only temporarily by individuals during travels or feeding.

Individual tracking data—RSF

To estimate habitat selection from individual tracking data, RSFs are classically modeled using a logistic regression to estimate the exponential of the linear predictor based on used-available locations (Boyce & McDonald, 1999; Manly et al., 2002; Muff et al., 2020). Animal locations being coded as 1 and available locations as 0, let us name K the dataset that stores the use-available data. The RSF convergence with IPP holds when RSF is specified in line with the weighted distribution theory, that is, when approximating an infinite number of available locations (Aarts et al., 2012; Fieberg et al., 2021; Fithian & Hastie, 2013; Warton & Shepherd, 2010). In practice, the weight w must be sufficiently large and we chose $w = 1000$ for available locations, and w is 1 for used locations (Fieberg et al., 2021; Muff et al., 2020). RSFs operate in a continuous geographical space, and the K_s of each point s in S is modeled as a weighted logistic regression of the IPP intensity, that is, a binomial draw of size w and probability equal to λ of Equation (4).

$$\text{logit}(\lambda_s) = \beta_0 + \sum \beta X_s, \quad (4)$$

$$K_s \sim \text{Binomial}(w_s, \lambda_s). \quad (5)$$

The regression coefficients of the RSF indicate the relative selection for covariate X. Overall, predictions of an RSF written as $\Theta = \exp(\beta X)$ is equivalent to the relative (not absolute) intensity of an IPP without its intercept. To account for individual heterogeneity in habitat selection in the population, we followed recommendations of Muff et al. (2020), and we included an individual random effect on regression parameters βs (i.e., random slopes) and used a random intercept with a large fixed variance with

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

$$\beta_{0,i} \sim \text{Normal}(\beta_0, 1e6), \quad (7)$$

for $i \in [1, I.]$ With β being the parameter of interest describing the population mean relative habitat selection strength of covariate X, τ represents the between-individuals variance, that is, population heterogeneity

regarding habitat selection of covariate X, and β_0 the population intercept that is biologically meaningless in the RSF.

RSF can lead to biased habitat selection coefficients and unreliable (narrow) CIs due to serial dependence in tracking data (Alston et al., 2023). Autocorrelation can be reduced with data thinning (Hooten et al., 2014; Northrup et al., 2013) or explicitly considered in RSF using likelihood weighting of animal locations (Alston et al., 2023).

Data integration—RSF-Poisson model

The integrated model keeps the structure of the Poisson GLM and the RSF, hence including two IPP-like models with specific intercepts β_0^{counts} for the Poisson GLM submodel and β_0^{RSF} for the RSF submodel and shared habitat selection coefficient, β.

Poisson GLM submodel

$$\log(\lambda_{\text{counts}}) = \beta_0^{\text{counts}} + \sum \beta X, \quad (8)$$

$$N \sim \text{Poisson}(\lambda_{\text{counts}}). \quad (9)$$

RSF submodel

$$\text{logit}(\lambda_{\text{RSF}}) = \beta_0^{\text{RSF}} + \sum \beta X, \quad (10)$$

$$K \sim \text{Binomial}(w, \lambda_{\text{RSF}}). \quad (11)$$

Finally, we obtained maps of relative habitat selection intensity by representing $\exp(\beta X)$ for RSF, Poisson GLM, and integrated model (Figure 2A). For details, see R codes (https://zenodo.org/records/15260611).

The convergence of Poisson GLM and RSF within the IPP framework allows estimating the same pattern of relative (and not absolute) habitat selection. The exponential predictor $\exp(\beta X)$ aligns, without the intercepts that are biologically specific to each data source (Northrup et al., 2022). We measured relative intensity of use, and the exponential predictor is "proportional to the probability of use" (Fieberg et al., 2021; Manly et al., 2002). If grid-cell closure assumption is respected, the Poisson GLM produces an absolute estimation of λ^{counts} that reflects the absolute intensity of use (i.e., expected number of individuals per grid-cell). Then, conditional on the geographic closure of the grid-cells, one could be tempted to obtain absolute estimation of density reconstructing

Equation (8) with the specific intercept β_0^{counts} of the Poisson GLM submodel. Although theoretically possible, we emphasize that such consideration is rarely compatible with mobile species equipped with tracking devices.

APPLICATIONS

Simulations

To test the validity and robustness of our models, we generated a discrete landscape of two random habitat types A and B equally available with low spatial autocorrelation on a study area of 1000×1000 cells. We simulated a population of 1000 individuals who moved in this landscape during 300 steps and set the movement parameters to obtain a higher use of habitat type A over type B. We simulated tracks based on a biased correlated random walk to mimic a central place forager behavior (see R code [https://zenodo.org/records/15260611], Benhamou, 2014), but we could have relaxed this assumption. To be more realistic, we assigned to each individual some heterogeneity in the relative use of A over B. Due to the implementation of stochasticity in individual use and the central place forager constraint (individuals return to the colony at the end of their trip), the observed relative selection of habitat type A over type B was $\widehat{\Delta} =$ number of steps ending in habitat A/number of steps ending in habitat B $= 3.22$ for the 1000 individuals (Appendix S1: Figure S3), which is logically slightly different from the parameter set in the movement model (4). $\widehat{\Delta}$ approximates the relative selection of habitat type A over type B (i.e., the ratio of selection ratios) in our context of equal availability of both habitat types and low spatial correlation. Therefore, the regression coefficient for the habitat type covariate to be estimated was $\beta = \log\left(\widehat{\Delta}\right) = 1.17$.

To mimic tracking data, we randomly selected I individuals among 1000 to fit the RSF model described in the previous section. We generated 10 available points for each used location within a negative exponential bivariate distribution centered on the colony from where the 1000 individuals started, as recommended by Benhamou and Courbin (2023) to account for the central place forager behavior. Then, to mimic the sampling of count data, we simulated zigzag line transects using {dsims} R-package (Marshall, 2023). To build our grid for Poisson GLM analyses, we segmented the line transects into cells of the same size as habitat grid-cells. We summed the number of locations of the 1000 individuals that fell into each grid-cell to constitute our count dataset N. We did not account for any imperfect detection and considered $p = 1$ in the simulations. Based on these count data, we fitted Poisson GLM to estimate the relative selection of A over B.

We explored the effect of the quantity of each data source to be included in the model on the estimation of the habitat selection coefficient β. Among the 1000 individuals, we kept subsets of $I = 2, 5,$ and 20 individuals (i.e., 0.2%, 0.5%, and 2% of the population) to fit RSFs. The very low spatial autocorrelation in the simulated landscape results in habitat selection being performed without serial dependence from previous locations. Then, no autocorrelation issue applies to the simulation case study and RSF can be fitted without data thinning. We did not consider random slopes because of the low number of individuals for the first two subsets. We fitted Poisson GLMs on line-transect length ranging from 2000 to 15,000 cell-units sampling effort (i.e., sampling 0.2%–1.5% of the study area). We fitted the integrated RSF-Poisson model using an increasing amount of data from count and RSF data (Figure 1).

We ran 40 scenarios of simulations to verify that Poisson GLM, RSF, and the integrated model produced unbiased estimation of β and checked the performance of the integrated model in terms of precision. These results are shown in Appendix S1: Figure S5. In the main text, we only presented the results for a single random scenario of simulations to observe the effect of data integration on both precision and bias in a specific situation (as with empirical data).

Empirical case of Sandwich tern

We applied our modeling development to estimate at-sea habitat selection of Sandwich tern (*T. sandvicensis*) during the breeding period (June–August) in the Gulf of Lion, French Mediterranean Sea.

Twenty-two Sandwich terns were equipped with GPS-GSM in Etang de Thau nesting colony (43°23′13″ N; 3°37′17″ E) near Montpellier (Southern France) in late May 2021 and 2022 (i.e., 0.9% of the nesting population of 2269 individuals, GISOM, 2022). We considered only data between June and August 2022 and removed all locations on land. We assessed that autocorrelation among step lengths vanished beyond 50 min and retained one location every hour (Appendix S1: Figure S4). After filtering our dataset, we obtained 8490 GPS locations from 16 individuals, other individuals being dead, out of the study area, or the tracking device stopped functioning (Appendix S1: Figure S2). To model the RSF, we generated 10 available points for each used location within a negative exponential bivariate distribution centered on the breeding colony to account for the central place forager behavior of terns during the breeding period (Benhamou & Courbin, 2023). We restricted our study area to the continental shelf of the Gulf of Lion, considering seafloors above 200 m depth. We considered

FIGURE 1 Estimation of habitat selection coefficient for habitat type A over type B from simulated data. Coefficient of habitat selection estimated with (A) Poisson GLM with an increasing amount of data, (B) resource selection function (RSF) with an increasing number of individuals, (C) integrated RSF-Poisson GLM model, based on five individuals for the RSF submodel and an increasing amount of line-transect data (2000, 5000, and 10,000 cell-units) for the Poisson GLM submodel, and (D) integrated RSF-Poisson GLM model, based on 2000 cell-units line transect for the Poisson GLM submodel and an increasing amount of individuals (2, 5, and 20) for the RSF submodel. The black dotted line represents the expected value (i.e., the value used in the movement simulations).

linear and quadratic effects of bathymetry and distance to coastline as environmental covariates. We did not find a strong correlation between distance to coastline and bathymetry ($r = -0.51$). Covariates were scaled before the analysis. We fitted an RSF accounting for individual heterogeneity in Sandwich tern habitat selection, with both random intercepts and random slopes. Then, for each individual i and each location s in the study area, the RSF model was:

$$\text{logit}\left(\lambda_{i,s}^{RSF}\right) = \beta_{0,i} + \beta_{1,i}\text{Bathymetry}_s + \beta_{2,i}\text{Bathymetry}_s^2$$
$$+ \beta_{3,i}\text{Distance to coastline}_s, \quad (12)$$

with

$$\beta_{0,i} \sim \text{Normal}(\beta_0, 1e6),$$

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

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

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

where β_0, β_1, β_2, β_3, τ_1, τ_2, τ_3 are to be estimated.

For the count model, we benefited from a yearly monitoring program with large-scale line transects targeting small pelagic fisheries in the Gulf of Lion in July but also seabirds between 2017 and 2021 (PELMED, Laran et al., 2017; Appendix S1: Figure S1). We divided our study area into 711 contiguous hexagonal grid-cells of 21.65 km^2 (Figure 2). We fitted a Poisson GLM using count data collected during 2017–2021 summers, hence considering $J = 5$ sampling occasions, assuming no temporal variation in the expected number of individuals per grid-cell. We obtained 281 detections for a total count of 584 Sandwich terns. We included sampling effort calculated as the transect length (in kilometers) prospected in each grid-cell s during each sampling occasion j as an observation covariate affecting detection probability $p_{s,j}$.

$$\text{logit}\left(p_{s,j}\right) = \rho_0 + \rho_1 \text{Effort}_{s,j}. \quad (13)$$
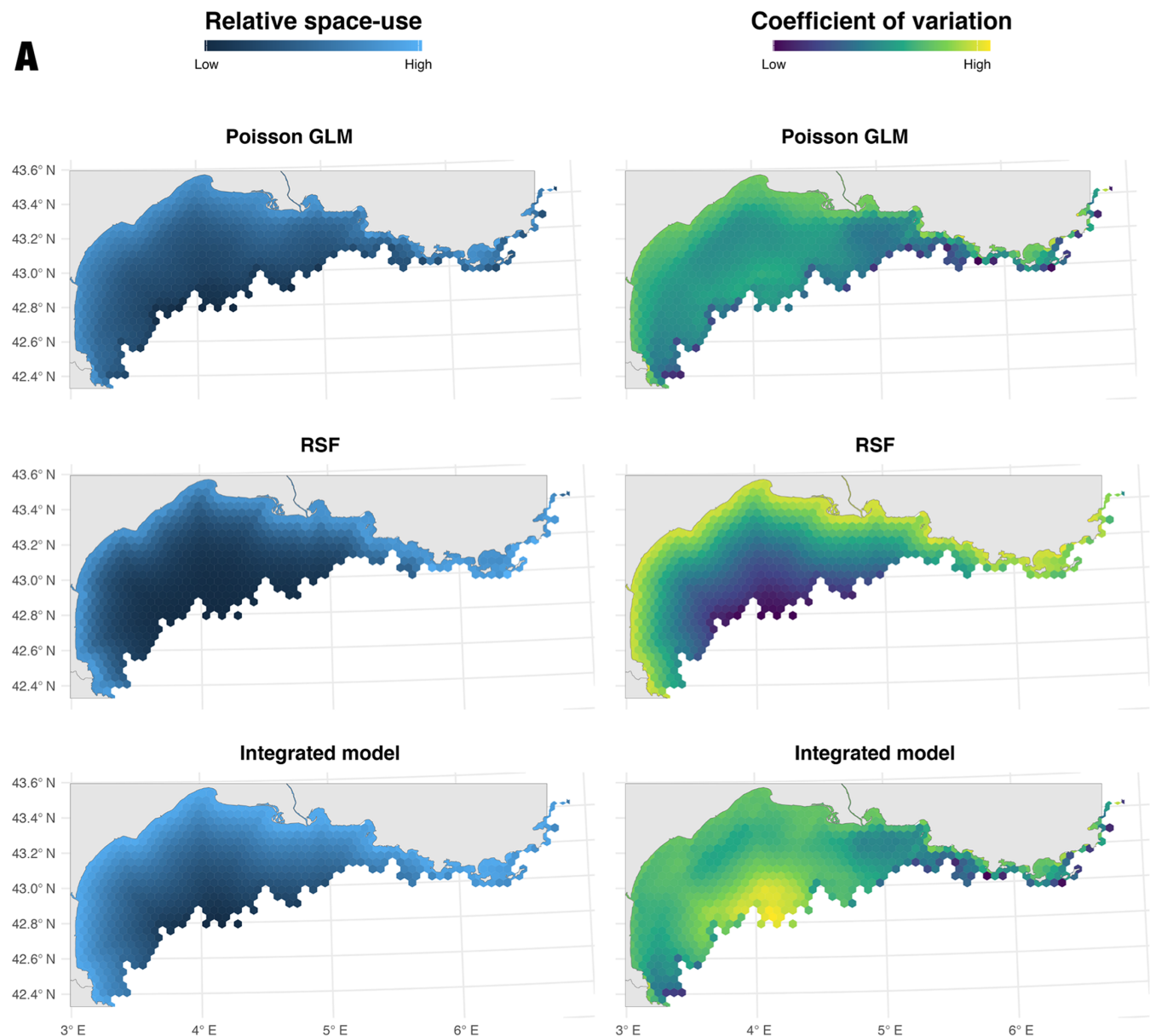
FIGURE 2 Habitat selection of Sandwich tern (*Thalasseus sandvicensis*) during the breeding period in the Gulf of Lion, French Mediterranean Sea, fitted with resource selection function (RSF), Poisson GLM, and an integrated habitat selection model combining telemetry and count data. We estimated habitat selection as a function of a linear and quadratic effect of bathymetry, and a linear effect of distance to coastline. (A) Spatial relative intensity of use for Sandwich tern and associated coefficient of variation predicted by RSF, Poisson GLM, and the integrated model. (B) Distribution of the habitat selection coefficient estimated by RSF, Poisson GLM, and the integrated model. (C) Distribution of the coefficient of variation for Sandwich tern space-use predictions among the grid-cells of the study area for RSF, Poisson GLM, and the integrated model.
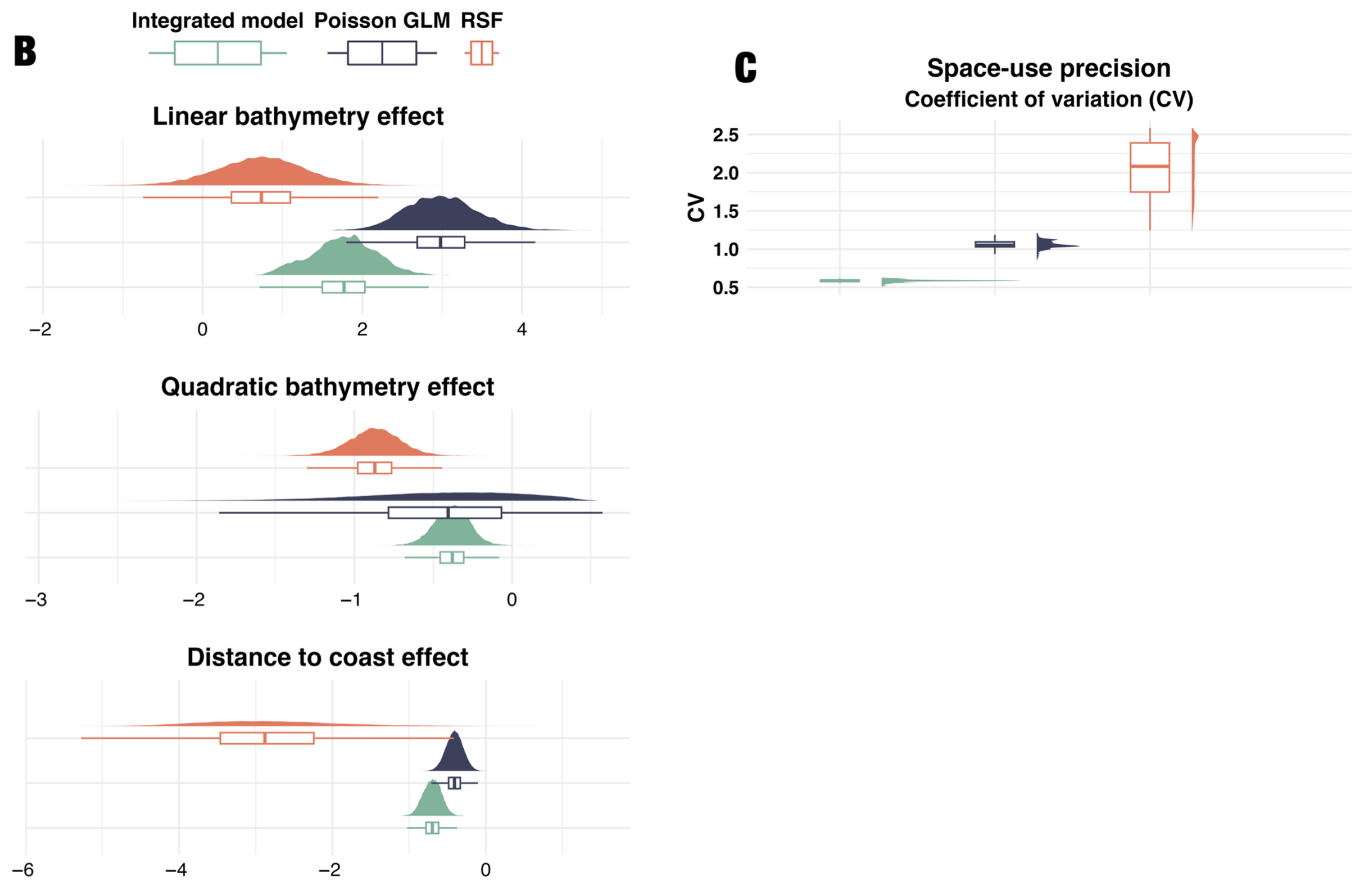
F I G U R E 2   (Continued)

We estimated the IPP intensity λ^{counts} using the same environmental covariates as the RSF model and the Poisson GLM submodel was:

$$\log\left(\lambda_s^{\text{counts}}\right) = \beta_0 + \beta_1 \text{Bathymetry} + \beta_2 \text{Bathymetry}^2 + \beta_3 \text{Distance to coastline}. \tag{14}$$

Finally, we implemented the integrated model joining the RSF and the Poisson GLM as described in *Data integration—RSF-Poisson model*.

Bayesian implementation

We fitted all models with NIMBLE R-package (de Valpine et al., 2017). We ran two MCMC chains with 110,000 iterations each, 10,000 iterations burn-in, and we assessed convergence, ensuring that Gelman test provided $\widehat{R} < 1.1$ (Gelman et al., 2013).

We displayed the posterior distribution of the βs for the RSF, the Poisson GLM, and the integrated model in Figures 1 and 2B for the simulation case and the empirical Sandwich tern case study, respectively. We predicted

the relative intensity of use by plotting the exponential of the linear predictor (without intercept)

$$\Theta = \exp\left(\beta_1 \text{Bathymetry} + \beta_2 \text{Bathymetry}^2 + \beta_3 \text{Distance to coastline}\right) \tag{15}$$

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 Θ calculating the coefficient of variation $\text{CV}(\Theta) = \text{SD}(\Theta)/\overline{\Theta}$ in each grid-cell for each model (Figure 2A) and representing the distribution of the precision over all grid-cells (Figure 2C).

RESULTS

Simulations

Over 40 scenarios of simulations, both Poisson GLM and RSF, and the integrated model produced unbiased estimates of the regression coefficient β, with the overall posterior distribution centered on the expected value 3.22

(Appendix S1: Figure S5). Apparent bias of β in Figure 1 was simply due to sampling stochasticity, which showed the benefit of the integrated model in a context of limited data. Both Poisson GLM and RSF resulted in better precision and lower bias of β with increasing number of cell-units sampled along the line transects and the increasing number of individuals, respectively (Figure 1A,B). Importantly, the integrated model was able to estimate unbiased β by combining medium-sized datasets. Integrating increasing count data to the RSF fitted with only five individuals significantly reduced bias and increased precision of the β estimate (Figure 1C). Similarly, integrating increasing number of individuals in an RSF to a Poisson GLM with a low sampling effort of 2000 cell-units line transect increased the precision and reduced the bias of β compared to the Poisson GLM fitted with 5000 cell-units in isolation (Figure 1D). Besides, the intercept of the Poisson GLM single model and the intercept of the Poisson GLM submodel β_0^{counts} converged, with increasing precision when more line-transect data were included in the model (Appendix S1: Figure S5).

Sandwich tern in the French Mediterranean Sea

RSF, Poisson GLM, and integrated models showed that Sandwich terns were more likely to select areas closer than far from the coast (Figure 2B). The linear effect of bathymetry was more similar between Poisson GLM and the integrated model than with the RSF. In contrast, the RSF and integrated model estimated a significant quadratic effect of bathymetry, while Poisson GLM estimated a quadratic effect of bathymetry with low precision (Figure 2A). RSF found a strong negative effect of distance to coastline; for example, terns selected habitat close to the shore, but with low precision (Figure 2B). In contrast, Poisson GLM and the integrated model supported a weak negative effect with high precision (Figure 2B). Then, the integrated model benefited from both datasets to produce precise estimates of all covariates, while Poisson GLM and RSF fitted in isolation produced low precision for one covariate. In terms of relative intensity of spatial use (Figure 2A), the integrated model predicted a high use of coastal and mid-range depth seas, while Poisson GLM and RSF predicted a high use of shallower seas compared to mid-range depth and pelagic seas, which reflected the combined significant effect of both the quadratic bathymetry effect and distance to coastline for the integrated model (Figure 2B). When comparing the coefficient of variation associated with relative space-use selection (Figure 2C), the integrated model displayed a better precision than the Poisson GLM model, RSF showing the lower precision (i.e., higher coefficient of variation).

DISCUSSION

Based on the IPP framework, we developed an integrated model RSF-Poisson GLM combining individual tracking and count data to estimate relative habitat selection and more generally animal space use, complementary to other recent approaches formalizing an integrated model SSF-Poisson GLM, for example (Blackwell & Matthiopoulos, 2024; Buderman et al., 2025). We showcased the benefit of such an integrated model by applying it to simulated movement data and an empirical Sandwich tern case study. Our results demonstrated that in the absence of an important amount of count data or tracking data, the integrated habitat selection model offers an appealing solution to produce unbiased and precise estimates of habitat selection coefficients.

Simulations and Sandwich tern study case

The integrated model correctly retrieved the expected value of relative selection of habitat type A over type B not only when the data are abundant but also when the data quantity is limited or when individual heterogeneity in habitat selection occurred (Figure 1A,B). Simulations outlined in the integrated models showed increased or equivalent precision and reduced bias compared with single-data models (Figure 1C,D), a well-known property of integrated models used for species distribution or population dynamics modeling (Lauret et al., 2021; Riecke et al., 2019). Although we highlighted the ability of the integrated model to produce a correct estimation of habitat selection, we encourage future studies to investigate the relevant conditions of applications. Caution has to be taken when performing data integration, as adding more data without explicitly accounting for a potential source of sampling bias may eventually lead to erroneous or inaccurate ecological estimations, sometimes providing poorer inferences compared to using a single dataset coming from an unbiased sampling (Simmonds et al., 2020). Integrating two datasets that include sources of bias in the same direction (i.e., both under or over estimating regression coefficients) would also be unlikely to improve the accuracy of the estimates, as shown when comparing the integrated models with 5-individuals RSF and 5000 cell-units versus 2000 cell-units for the Poisson GLM (see *Results*).

The Sandwich tern case study also showed that integrating individual tracking and spatial survey data improved the precision estimation of habitat selection

coefficients for all covariates compared to the Poisson GLM or RSF model alone (Figure 2). Bathymetry is known to affect the behavior of the Sandwich tern and its spatial use; however, the size of its effect remains unclear (Fijn et al., 2022) and seems to depend on the local context and the spatial scale studied (Fauchald et al., 2021; Fijn et al., 2022; Pratte et al., 2021; van Bemmelen et al., 2023). Although the results of the three models are coherent with existing literature, we suggest that data integration has potential conservation implications for the Sandwich tern study. Notably, the integrated model predicted a more—although limited—pelagic habitat selection compared to the Poisson GLM and RSF model alone, which would not have been uncovered without data integration (Figure 2A).

Conditions of use and extensions

Availability of both individual tracking and spatial survey data should not lead to a systematic application of integrated modeling. One should wonder whether it is worth integrating multiple datasets. When one dataset is abundant and produces precise estimates, engaging in data integration might not significantly improve the outputs (Figure 1). We expect that integration would be valuable, for instance, when interested in a large-scale habitat selection assessment where both individual tracking data and spatial surveys covered distinct areas, or when gaps in time series of spatial surveys can be complemented by individual tracking data. Similarly, when individuals equipped with tracking devices display a specific behavior that is not representative of the population, integrating unbiased sampling of count data should improve the accuracy and precision of habitat selection estimations (Farr et al., 2020). In such situations, integrated models can benefit from both data sources and capture spatial or temporal complementary aspects of the targeted population (Figure 2; Lauret et al., 2022). Yet, further investigation will be needed to confirm our hypothesis.

While we built our integrated model based on RSF and Poisson GLM, possible extensions are worth considering to make the most of tracking loggers that provide high-frequency telemetry data. SSFs constitute a more and more commonly used method to estimate habitat selection strength while accounting for movement characteristics (Fortin et al., 2005; Muff et al., 2020). Integration of movement models and spatial surveys has been dealt with in two recent papers (Blackwell & Matthiopoulos, 2024; Buderman et al., 2025). Regarding spatial surveys, we focused on population counts, while other population-level data types can be considered. Species distribution models widely use 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 log-log link function favor the convergence of the coefficients of the logistic regression fitted to presence-absence data with the coefficients estimated with an IPP fitted to count data (Aarts et al., 2012). Thus, the occupancy model represents another candidate to integrate population-level data with individual tracking data to estimate relative habitat selection. Whether it is possible or not to integrate RSF with a presence-only model, occupancy model, distance sampling, or integrated species distribution model, such methodological developments would however require further investigation (Farr et al., 2020; Mackenzie et al., 2002; Miller et al., 2019; Royle, 2004; Tyre et al., 2003).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data and code (Lauret, 2025) are available from Zenodo: https://doi.org/10.5281/zenodo.15260611.

ORCID

Valentin Lauret https://orcid.org/0000-0002-7739-4277
Nicolas Courbin https://orcid.org/0000-0002-6428-5012
Aurélien Besnard https://orcid.org/0000-0002-2289-9761

REFERENCES

Aarts, G., J. Fieberg, and J. Matthiopoulos. 2012. "Comparative Interpretation of Count, Presence-Absence and Point Methods for Species Distribution Models: Species Distribution as Spatial Point Process." *Methods in Ecology and Evolution* 3: 177–187.

Alston, J. M., C. H. Fleming, R. Kays, J. P. Streicher, C. T. Downs, T. Ramesh, B. Reineking, and J. M. Calabrese. 2023. "Mitigating Pseudoreplication and Bias in Resource Selection Functions with Autocorrelation-Informed Weighting." *Methods in Ecology and Evolution* 14: 643–654.

Benhamou, S. 2014. "Of Scales and Stationarity in Animal Movements." *Ecology Letters* 17: 261–272.

Benhamou, S., and N. Courbin. 2023. "Accounting for Central Place Foraging Constraints in Habitat Selection Studies." *Ecology* 104: e4134.

Blackwell, P. G., and J. Matthiopoulos. 2024. "Joint Inference for Telemetry and Spatial Survey Data." *Ecology* 105: e4457.

Boyce, M. S., and L. L. McDonald. 1999. "Relating Populations to Habitats Using Resource Selection Functions." *Trends in Ecology & Evolution* 14: 268–272.

Buderman, F. E., E. M. Hanks, V. Ruiz-Gutierrez, M. Shull, R. K. Murphy, and D. A. W. Miller. 2025. "Integrated Movement Models for Individual Tracking and Species Distribution Data." *Methods in Ecology and Evolution* 16: 345–361.

Chandler, R. B., D. A. Crawford, E. P. Garrison, K. V. Miller, and M. J. Cherry. 2022. "Modeling Abundance, Distribution, Movement and Space Use with Camera and Telemetry Data." *Ecology* 103: e3583.

Converse, S. J., B. T. McClintock, and P. B. Conn. 2022. "Special Feature: Linking Capture–Recapture and Movement." *Ecology* 103: e3770.

de Valpine, P., D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, and R. Bodik. 2017. "Programming with Models: Writing Statistical Algorithms for General Model Structures with NIMBLE." *Journal of Computational and Graphical Statistics* 26: 403–413.

Farr, M. T., D. S. Green, K. E. Holekamp, and E. F. Zipkin. 2020. "Integrating Distance Sampling and Presence-Only Data to Estimate Species Abundance." *Ecology* 102(1): e03204.

Fauchald, P., A. Tarroux, F. Amélineau, V. Bråthen, S. Descamps, M. Ekker, H. Helgason, et al. 2021. "Year-Round Distribution of Northeast Atlantic Seabird Populations: Applications for Population Management and Marine Spatial Planning." *Marine Ecology Progress Series* 676: 255–276.

Fieberg, J., J. Signer, B. Smith, and T. Avgar. 2021. "A 'how to' Guide for Interpreting Parameters in Habitat-Selection Analyses." *Journal of Animal Ecology* 90: 1027–43.

Fijn, R., C. Thaxter, G. Aarts, J. Adema, R. Middelveld, and R. Van Bemmelen. 2022. "Relative Effects of Static and Dynamic Abiotic Conditions on Foraging Behaviour in Breeding Sandwich Terns." *Marine Ecology Progress Series* 692: 137–150.

Fithian, W., and T. Hastie. 2013. "Finite-Sample Equivalence in Statistical Models for Presence-Only Data." *The Annals of Applied Statistics* 7: 1917–39.

Fletcher, R. J., T. J. Hefley, E. P. Robertson, B. Zuckerberg, R. A. McCleery, and R. M. Dorazio. 2019. "A Practical Guide for Combining Data to Model Species Distributions." *Ecology* 100(6): e02710.

Fortin, D., H. L. Beyer, M. S. Boyce, D. W. Smith, T. Duchesne, and J. S. Mao. 2005. "Wolves Influence Elk Movements: Behavior Shapes a Trophic Cascade in Yellowstone National Park." *Ecology* 86: 1320–30.

Fradin, G., and S. Chamaillé-Jammes. 2023. "Hogs Sleep like Logs: Wild Boars Reduce the Risk of Anthropic Disturbance by Adjusting where they Rest." *Ecology and Evolution* 13: e10336.

Gamelon, M., C. R. Nater, É. Baubet, A. Besnard, L. Touzot, J.-M. Gaillard, J.-D. Lebreton, and O. Gimenez. 2021. "Efficient Use of Harvest Data: A Size-Class-Structured Integrated Population Model for Exploited Populations." *Ecography* 44: 1296–1310.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis*. London: Chapman and Hall/CRC.

GISOM. 2022. "Recensement National des oiseaux marins nicheurs en France." Technical Report.

Hooten, M. B., E. M. Hanks, D. S. Johnson, and M. W. Alldredge. 2014. "Temporal Variation and Scale in Movement-Based Resource Selection Functions." *Statistical Methodology* 17: 82–98.

Illian, J. B. 2019. "Spatial and Spatio-Temporal Point Processes in Ecological Applications." In *Page Handbook of Environmental and Ecological Statistics*. London: Chapman and Hall/CRC.

Johnson, D. S., M. B. Hooten, and C. E. Kuhn. 2013. "Estimating Animal Resource Selection from Telemetry Data Using Point Process Models." *Journal of Animal Ecology* 82: 1155–64.

Kendall, C. J., C. Bracebridge, E. C. Lynch, M. Mgumba, A. Monadjem, A. Nicholas, and A. Kane. 2023. "Value of Combining Transect Counts and Telemetry Data to Determine Short-Term Population Trends in a Globally Threatened Species." *Conservation Biology* 37: e14146.

Kéry, M., and J. Royle. 2020. *Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in r and Bugs: Volume 2: Dynamic and Advanced Models*, First ed. Cambridge: Elsevier.

Koshkina, V., Y. Wang, A. Gordon, R. M. Dorazio, M. White, and L. Stone. 2017. "Integrated Species Distribution Models: Combining Presence-Background Data and Site-Occupancy Data with Imperfect Detection." *Methods in Ecology and Evolution* 8: 420–430.

Laran, S., E. Pettex, M. Authier, A. Blanck, L. David, G. Dorémus, H. Falchetto, P. Monestiez, O. Van Canneyt, and V. Ridoux. 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.

Lauret, V. 2025. "Valentinlauret/integratedIPP: Upload Rcodes and Files for Accepted Manuscript." Zenodo.

Lauret, V., H. Labach, M. Authier, and O. Gimenez. 2021. "Using Single Visits into Integrated Occupancy Models to Make the Most of Existing Monitoring Programs." *Ecology* 102: e03535.

Lauret, V., H. Labach, D. Turek, S. Laran, and O. Gimenez. 2022. "Integrated Spatial Models Foster Complementarity between Monitoring Programmes in Producing Large Scale Bottlenose Dolphin Indicators." *Animal Conservation* 26(2), 228–238.

Liang, D., H. Bailey, A. L. Hoover, S. Eckert, P. Zarate, J. Alfaro-Shigueto, J. C. Mangel, et al. 2023. "Integrating Telemetry and Point Observations to Inform Management and Conservation of Migratory Marine Species." *Ecosphere* 14(1): e4375. https://doi.org/10.1002/ecs2.4375.

Lindenmayer, D. B., T. Lavery, and B. C. Scheele. 2022. "Why we Need to Invest in Large-Scale, Long-Term Monitoring Programs in Landscape Ecology and Conservation Biology." *Current Landscape Ecology Reports* 7: 179–187.

Mackenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. "Estimating Site Occupancy Rates when Detection Probabilities Are Less than One." *Ecology* 83: 8.

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

Marshall, L. 2023. "Dsims: Distance Sampling Simulations." R, Distance Development.

Matthiopoulos, J., E. Wakefield, J. W. E. Jeglinski, R. W. Furness, M. Trinder, G. Tyler, A. Mccluskie, S. Allen, J. Braithwaite,

and T. Evans. 2022. "Integrated Modelling of Seabird-Habitat Associations from Multi-Platform Data: A Review." *Journal of Applied Ecology* 59: 909–920.

Michelot, T., P. G. Blackwell, and J. Matthiopoulos. 2019. "Linking Resource Selection and Step Selection Models for Habitat Preferences in Animals." *Ecology* 100: e02452.

Miller, D. A. W., K. Pacifici, J. S. Sanderlin, and B. J. Reich. 2019. "The Recent Past and Promising Future for Data Integration Methods to Estimate Species' Distributions." *Methods in Ecology and Evolution* 10: 22–37.

Muff, S., J. Signer, and J. Fieberg. 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: 80–92.

Nathan, R., C. T. Monk, R. Arlinghaus, T. Adam, J. Alós, M. Assaf, H. Baktoft, et al. 2022. "Big-Data Approaches Lead to an Increased Understanding of the Ecology of Animal Movement." *Science* 375: eabg1780.

Northrup, J. M., M. B. Hooten, C. R. Anderson, Jr., and G. Wittemyer. 2013. "Practical Guidance on Characterizing Availability in Resource Selection Functions under a Use–Availability Design." *Ecology* 94: 1456–63.

Northrup, J. M., E. Vander Wal, M. Bonar, J. Fieberg, M. P. Laforge, M. Leclerc, C. M. Prokopenko, and B. D. Gerber. 2022. "Conceptual and Methodological Advances in Habitat-Selection Modeling: Guidelines for Ecology and Evolution." *Ecological Applications* 32: e02470.

Potts, J. R., G. Bastille-Rousseau, D. L. Murray, J. A. Schaefer, and M. A. Lewis. 2014. "Predicting Local and Non-local Effects of Resources on Animal Space Use Using a Mechanistic Step Selection Model." *Methods in Ecology and Evolution* 5: 253–262.

Potts, J. R., and U. E. Schlägel. 2020. "Parametrizing Diffusion-Taxis Equations from Animal Movement Trajectories Using Step Selection Analysis." *Methods in Ecology and Evolution* 11: 1092–1105.

Pratte, I., R. Ronconi, S. Craik, and J. McKnight. 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.

Regehr, E. V., N. J. Hostetter, R. R. Wilson, K. D. Rode, M. S. Martin, and S. J. Converse. 2018. "Integrated Population Modeling Provides the First Empirical Estimates of Vital Rates and Abundance for Polar Bears in the Chukchi Sea." *Scientific Reports* 8: 16780.

Renner, I. W., J. Elith, A. Baddeley, W. Fithian, T. Hastie, S. J. Phillips, G. Popovic, and D. I. Warton. 2015. "Point Process Models for Presence-Only Analysis." *Methods in Ecology and Evolution* 6: 366–379.

Riecke, T. V., P. J. Williams, T. L. Behnke, D. Gibson, A. G. Leach, B. S. Sedinger, P. A. Street, and J. S. Sedinger. 2019. "Integrated Population Models: Model Assumptions and Inference." *Methods in Ecology and Evolution* 10: 1072–82.

Royle, J. A. 2004. "N-Mixture Models for Estimating Population Size from Spatially Replicated Counts." *Biometrics* 60: 108–115.

Royle, J. A., and J. D. Nichols. 2003. "Estimating Abundance from Repeated Presence–Absence Data or Point Counts." *Ecology* 84: 777–790.

Simmonds, E. G., S. G. Jarvis, P. A. Henrys, N. J. B. Isaac, and R. B. O'Hara. 2020. "Is More Data Always Better? A Simulation Study of Benefits and Limitations of Integrated Distribution Models." *Ecography* 43(10): ecog.05146. https://doi.org/10.1111/ecog.05146.

Strebel, N., M. Kéry, J. Guélat, and T. Sattler. 2022. "Spatiotemporal Modelling of Abundance from Multiple Data Sources in an Integrated Spatial Distribution Model." *Journal of Biogeography* 49: 923–935.

Tenan, S., P. Pedrini, N. Bragalanti, C. Groff, and C. Sutherland. 2017. "Data Integration for Inference about Spatial Processes: A Model-Based Approach to Test and Account for Data Inconsistency." *PLoS One* 12: e0185588.

Tyre, A. J., B. Tenhumberg, S. A. Field, D. Niejalke, K. Parris, and H. P. Possingham. 2003. "Improving Precision and Reducing Bias in Biological Surveys: Estimating False-Negative Error Rates." *Ecological Applications* 13: 1790–1801.

van Bemmelen, R. S., J. J. Leemans, M. P. Collier, R. M. Green, R. P. Middelveld, C. B. Thaxter, and R. C. Fijn. 2023. "Avoidance of Offshore Wind Farms by Sandwich Terns in the North Sea Increases with Turbine Density." *Ornithological Applications* 126(1): duad055.

Warton, D., and G. Aarts. 2013. "Advancing our Thinking in Presence-Only and Used-Available Analysis." *Journal of Animal Ecology* 82: 1125–34.

Warton, D., and L. C. Shepherd. 2010. "Poisson Point Process Models Solve the 'Pseudo-Absence Problem' for Presence-Only Data in Ecology." *The Annals of Applied Statistics* 4: 1383–1402.

Yamamoto, T., Y. Watanuki, E. L. Hazen, B. Nishizawa, H. Sasaki, and A. Takahashi. 2015. "Statistical Integration of Tracking and Vessel Survey Data to Incorporate Life History Differences in Habitat Models." *Ecological Applications* 25: 2394–2406.

Zipkin, E. F., B. D. Inouye, and S. R. Beissinger. 2019. "Innovations in Data Integration for Modeling Populations." *Ecology* 100: e02713.

Zipkin, E. F., E. R. Zylstra, A. D. Wright, S. P. Saunders, A. O. Finley, M. C. Dietze, S. Itter, and M. W. Tingley. 2021. "Addressing Data Integration Challenges to Link Ecological Processes across Scales." *Frontiers in Ecology and the Environment* 19: 9–16.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.