Bridging the Gap Between Lagrangian and Eulerian Species Distribution Models for Abundance Estimation-A Simulation Experiment

Charlotte Lambert, Anne-sophie Bonnet-Lebrun, David Grémillet

Journal of Biogeography

WILEY

Journal of Biogeography

RESEARCH ARTICLE OPEN ACCESS

Bridging the Gap Between Lagrangian and Eulerian Species Distribution Models for Abundance Estimation—A Simulation Experiment

Charlotte Lambert[1,2,3] | Anne-Sophie Bonnet-Lebrun[2] | David Grémillet[3,4]

[1]LIENSs, UMR 7266 CNRS-La Rochelle Université, La Rochelle, France | [2]CEBC, UMR 7372 CNRS-La Rochelle Université, Villiers-en-bois, France | [3]CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France | [4]FitzPatrick Institute of African Ornithology, Department of Biological Sciences, University of Cape Town, Rondebosch, South Africa

Correspondence: Charlotte Lambert (charlotte.lambert@univ-lr.fr)

Received: 25 September 2024 | **Revised:** 16 December 2024 | **Accepted:** 18 December 2024

Funding: This work was supported by Horizon 2020 Framework Programme.

Keywords: abundance estimation | distance sampling | habitat modelling | species-habitat association | telemetry data | visual survey

ABSTRACT

Aim: In mobile species, individual movement decisions based on biotic and abiotic conditions determine how individuals interact with the environment, heterospecifics and conspecifics. Accordingly, these decisions underpin all ecological principles and structure broader spatial patterns at the population and species level. Species distribution models (SDMs) are therefore paramount in ecology, with implications for both fundamental and applied studies. There are many robust SDM techniques, from individual-scale (Lagrangian) to population-scale (Eulerian) models. Their outputs routinely support wildlife management, conservation, or risk assessments. Yet, it remains unclear whether SDMs built at individual and population scales infer the same processes, and whether the spatial distributions they predict are comparable. Here, we address this key question with a simulation exercise.

Location: Virtual environment.

Taxon: Virtual species.

Methods: First, we simulated the individual movements of two highly mobile species, one central-place forager and one free ranger. Second, we surveyed the species at the individual-scale, replicating Lagrangian studies by tracking individual movements, and at the population-scale, replicating Eulerian surveys by censusing the study area with standardised protocols. The resulting data were analysed following well-established statistical methods to assess species abundance distribution. We used Resource Selection Functions (RSFs) for Lagrangian data and Density Surface Models (DSMs) for Eulerian data.

Results and Main Conclusions: Both Lagrangian and Eulerian SDMs adequately estimated the species' relationship with environmental conditions. Although some fine-scale differences occurred, both perspectives yielded highly correlated spatial distributions (correlations of 0.8–1.0 between pairs of models), and successfully predicted true abundance distributions (correlations of 0.6–0.7 with the true abundance distribution). Our results demonstrate that Lagrangian and Eulerian SDMs are statistically consistent and directly comparable, which is of great importance for conservation science. This provides crucial guidance for the combination of predictions from both model types to inform spatial planning within a wide range of management contexts.

1 | Introduction

Predicting the spatial distribution of species is a central goal in ecology. In mobile species, individuals are constantly making movement decisions (Mueller and Fagan 2008), which depends on environmental conditions, resource availability, previous knowledge, biological status, but also on inter- and intraspecific interactions, potential disturbance and fear of predators (Courbin et al. 2022; Palmer et al. 2022). From such individual-scale (Lagrangian) space-use choices emerge population-scale space-use patterns (DeCesare et al. 2012; Winter et al. 2024), which ultimately drive the spatial distribution of occupancy and abundance of populations and species (Eulerian process; Steenweg et al. 2018). By determining where and when individuals interact with their environment, as well as with heterospecifics and conspecifics, spatial processes at the individual scale ultimately underpin all ecological principles: from population and community dynamics to trophic relationships, disease spread and exposure to anthropogenic threats, which in turn affects individual movement decisions. Hence, to get a complete picture of ecosystem functioning, it is crucial to describe the spatial ecology of species from the individual and the population perspective (Truchy et al. 2019).

When studying and predicting the drivers of spatial distributions, such as resource and habitat use, occurrence or abundance distribution, we mostly rely on correlating individual location with environmental conditions experienced at these locations (Franklin 2010; Matthiopoulos, Fieberg, and Aarts 2020). Such correlative approaches form the core of species distribution modelling (SDM, also termed habitat modelling) and of habitat-selection analyses. The SDM terminology encompass many robust techniques that rely on a wide array of data types and condition the position of the analysis along the individual-to-population continuum (Figure 1). A majority of SDMs are based on observations recorded at discrete sample sites (Guillera-Arroita et al. 2015) and provide an Eulerian perspective. Such data include any type of direct (visual, acoustic) or indirect (faeces, tracks, DNA) observation of the species at a site. They range from opportunistic presence-only data (citizen science, museum collections, photo-ID catalogues), to presence–absence and count data, collected with standardised protocols (distance sampling, camera traps; Buckland et al. 2023). Eulerian SDMs are used to provide occurrence, usage or abundance maps. However, most wild populations are not easy to survey. This is especially the case of wide-ranging, highly mobile species that live at remote locations or that engage in continent- or ocean-wide migrations. For these, an individual perspective is often preferred, which is generally achieved through animal-borne tracking devices recording movements over time. This type of data can be used to build Lagrangian SDMs, which provide information on the probability of usage (Northrup et al. 2022; Hays et al. 2019).

Each approach comes with its limitations and drawbacks. Individual-based monitoring is typically limited to a relatively small number of tracked animals, due to economic, ethical or accessibility constraints. These tracking studies often provide a non-random sample of a population, and the data set might be further limited to particular demographic subsets (only breeders; only males or females). Due to differences between these subsets, this individual-based approach might lead to biased space-use estimates (Carroll et al. 2019; Pettex et al. 2019; Phillips et al. 2019). By contrast, Eulerian approaches provide a random sample of the population, as they sample all demographic subsets (if an adequate sampling scheme is used). However, they are limited in their spatio-temporal extent and resolution and are further subject to detection issues (Lahoz-Monfort, Guillera-Arroita, and Wintle 2014; Lambert et al. 2019). Also, the openness of a population (individuals moving in and out the study area during the survey) can bias abundance estimations. Yet, this bias can be reduced by carefully designing surveys (basin-scale surveys designed to cover the full range of populations; Hammond et al. 2017; Panigada et al. 2021) or by using dedicated statistical tools (N-mixture models; Ketz et al. 2018).

Both Lagrangian and Eulerian SDMs are extensively used to establish risk assessments, spatial planning, conservation management or mitigation strategies. They are also used to assess ecosystems status, often combining outputs from both SDM types (Marshall, Glegg, and Howell 2014). To answer to these needs, the SDM field has recently experienced a fast development of models that integrate heterogeneous observation data of various sources, types, quality and content information (Fletcher et al. 2019; Miller, Pacifici, et al. 2019). These models effectively combine various types of Eulerian data (standardised surveys, citizen science, camera trap or harvest data) to provide population-scale models of unprecedented reliability. However, although theoretical work has demonstrated that many Eulerian and Lagrangian SDMs are actually identical (Aarts, Fieberg, and Matthiopoulos 2012), and that individual-level Resource Selection Functions (RSFs) can be scaled up to population-level estimates of habitat selection (DeCesare et al. 2012; Winter et al. 2024), the integration of Lagrangian and Eulerian data into a single integrated SDM (iSDM) has, to the best of our knowledge, not been accomplished yet. Such an integration faces specific challenges (Matthiopoulos et al. 2022) because the two data types sample the spatial distribution in fundamentally different ways. This is particularly the case when the output of interest is the abundance distribution rather than the space-use or presence probability. It is also the case for wide-ranging species, whose movements occur at a greater spatial scale than Eulerian surveys.

Hence, regardless if we want to couple outputs from Eulerian and Lagrangian SDMs for management purposes, or if we want to build abundance-orientated iSDMs that couple Eulerian and Lagrangian data, we need to empirically evaluate that the spatial patterns predicted from the two approaches are comparable. This is the aim of our current study and we address this by answering the following two long-standing questions: (1) Do models built with Lagrangian and Eulerian data provide consistent information about the abundance distribution of species? and (2) Do these models provide reliable abundance distributions for highly mobile species?

Testing these hypotheses with empirical data is difficult. Instead, we implemented a virtual individual-based simulation experiment (Zurell et al. 2010). Using a dedicated R package

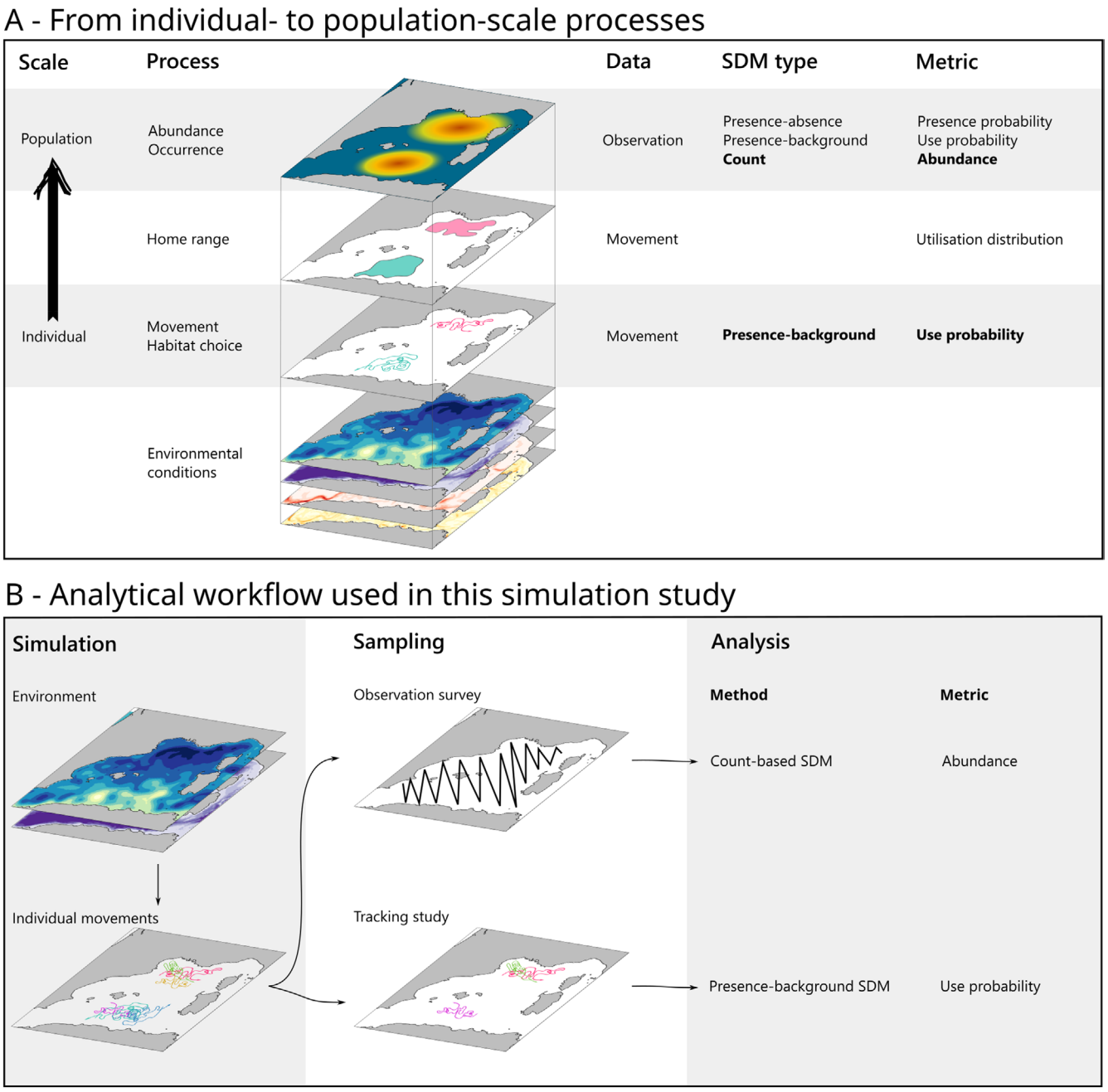

FIGURE 1 | (A) Individual scale (Lagrangian) processes are defined by movement choices of individuals based on the surrounding environmental conditions. They scale up to population-level (Eulerian) processes defining broad-scale patterns of occurrence and abundance distributions. Each data type samples the spatial processes at a different scale, limiting the type of SDMs available to model the spatial distribution of a species, and, ultimately, the metrics that can be inferred from it: From the probability of usage to abundance estimates. SDM refers to all techniques that infer the spatial distribution of a species from its relationships with environmental conditions, thus spanning from regression models (GLM, GAM and their mixed-effect counterparts) to deep-learning methods. SDM types and metrics that are compared in the present study are highlighted in bold. (B) Details of the analytical workflow used for our simulation study and overview of conducted analyses.

(`virtualecologist`; Lambert 2024a), we set a virtual environment in which we simulated the movements of two wide-ranging virtual species that either behave as central-place forager (hereafter the CPF species, which we assumed live in two colonies, with a total of 20,000 individuals) or as a free ranger (hereafter the FR species, 8000 individuals). The system was sampled from the Lagrangian perspective with a tracking study of both species and from the Eulerian perspective with two standardised surveys that were conducted from an aircraft (fast platform) and from a ship (slow platform).

We then analysed Lagrangian and Eulerian data using state-of-the-art statistical methods. Movement data were analysed with resource selection functions (RSFs; Northrup et al. 2022) to compute a map for the probability of usage, which we scaled-up to abundance. Using the survey data, we estimated

the abundance map with a Density Surface Model (DSM; Miller et al. 2013). We then assessed whether Lagrangian and Eulerian approaches arrived at similar estimates for the spatial distribution of both species, by comparing the abundance maps predicted either by RSF or by DSM (Pearson's correlation). Finally, we determined the absolute reliability of each method by comparing the inferred responses to environmental conditions and the predicted spatial abundance distributions with the factual situation.

2 | Methods

All analyses were conducted using R, version 4.2.3 (R Core Team 2024) and a dedicated package to set up the virtual environment, the species movements, and the surveys (`virtualecologist`; from which come all functions mentioned below, unless stated otherwise; Lambert 2024b).

Since our objective was to compare the efficacy of Lagrangian and Eulerian methods in retrieving the spatial distribution of abundance of a species, we simulated an ideal system. To this end, we controlled for the most common sources of statistical noise (species rareness, protocol bias, detection issues), by simulating abundant species and using line-transect protocol without detection bias for the surveys (all individuals were available to detection, i.e. they did not hide or dive). Such a setting ensured that the potential differences in model outputs would primarily arise from the performance of each method and not from statistical biases.

We used two types of species for our investigation. First, a Central Place Forager (CPF) that has to start each foraging trip from a given point (its colony) to which it has to return by the end of the day. This species moves across the study area at a medium pace, with a speed of about 30–50 km/h when travelling and of about 10–20 km/h when foraging. The species is highly mobile but cannot roam too far from the colony, as it has to return by the end of the day. As a consequence, individuals should maximise their search behaviour and focus on favourable patches during the time spent away from the colony. Hence, these individuals do not remain too long in one place (no more than 15 min in a 1 km radius). This type of behaviour is typical of the movements and ecology of many seabirds. Second, an FR that is not spatially constrained by a colony. It moves at a similar pace than the CPF, but it is more vagrant and spends less time at a particular place (no more than 15 min in a 2 km radius). Hence, it uses a larger area during a single day than the CPF. Such behaviour can be observed in cetaceans, for example.

2.1 | Data Simulation

2.1.1 | Environment

We set up a grid with a 0.5 × 0.5 km resolution (including 32,761 cells), using the `create_grid` function. Two Gaussian random fields were predicted across this space to simulate environmental conditions (using the `generate_env_layer` function; Figure 2A). From this, we built species-specific environmental suitability layers (using the `generate_resource_layer` function) by combining the two available resources as follows: for the CPF, suitability = 2 × Resource 1–1.5 × Resource 2; for the FR, suitability = −2 × Resource 1 + 0.5 × Resource 2 for the FR. We normalised the environmental conditions and the suitability layers.

2.1.2 | Individual Movements

2.1.2.1 | Central Place Forager. We simulated the movements of the CPF within the virtual environment (using the `simulate_trajectory_CPF` function). Individuals were assigned to two colonies (8000 and 10,000 individuals, respectively) from which individuals initiated central-place foraging trips on two consecutive days.

Individuals were set to launch foraging trips within 2 h of sunrise (departure time was randomly chosen within this interval) and to return to the colony before sunset. They departed the colony with a bearing drawn from a Von Mises distribution centred on 270° for Colony 1 and on 90° for Colony 2 (Table 1). Individuals could engage in two movement types: travelling (directed movement patterns: large steps with low variability, low angles with low variability) and foraging (area-restricted movement: short steps with low variability, large angles with large variability).

For each position (computed every minute), 10 potential subsequent positions were randomly sampled, using movement parameters from the activity at the previous step (travelling or foraging). During the first four positions of the track following colony departure, the ensuing step was randomly selected among these 10 potential positions. Thereafter, residence time was checked for every position. If residence time had not been reached (i.e., the individual did not spend more than 15 min in a 1 km radius of its current position), the subsequent position was chosen as the potential position with the highest values of environmental suitability. By contrast, if residence time had been reached (i.e., the individual had spent more than 15 min within a 1 km radius of its current position), only potential positions outside the 1 km radius were retained and the position with the best environmental condition was chosen. If all potential positions remained within the radius (which may happen when the individual is foraging), the subsequent position was chosen as the point situated at the greatest distance from the current position, irrespective of its environmental suitability, to force the individual moving away.

Then, we updated the behaviour of the individual based on the environmental suitability at the selected position: if suitability exceeded 0.7, the individual switched from travelling to foraging; otherwise, the individual continued to travel.

When individuals reached a distance from the colony further than 40 km, or when the trip duration reached 12 h, individuals started their homeward journey. The ensuing positions were chosen as travelling movements and were selected based on the minimum distance to the colony. If several potential positions met the selection criteria, the next step was randomly chosen

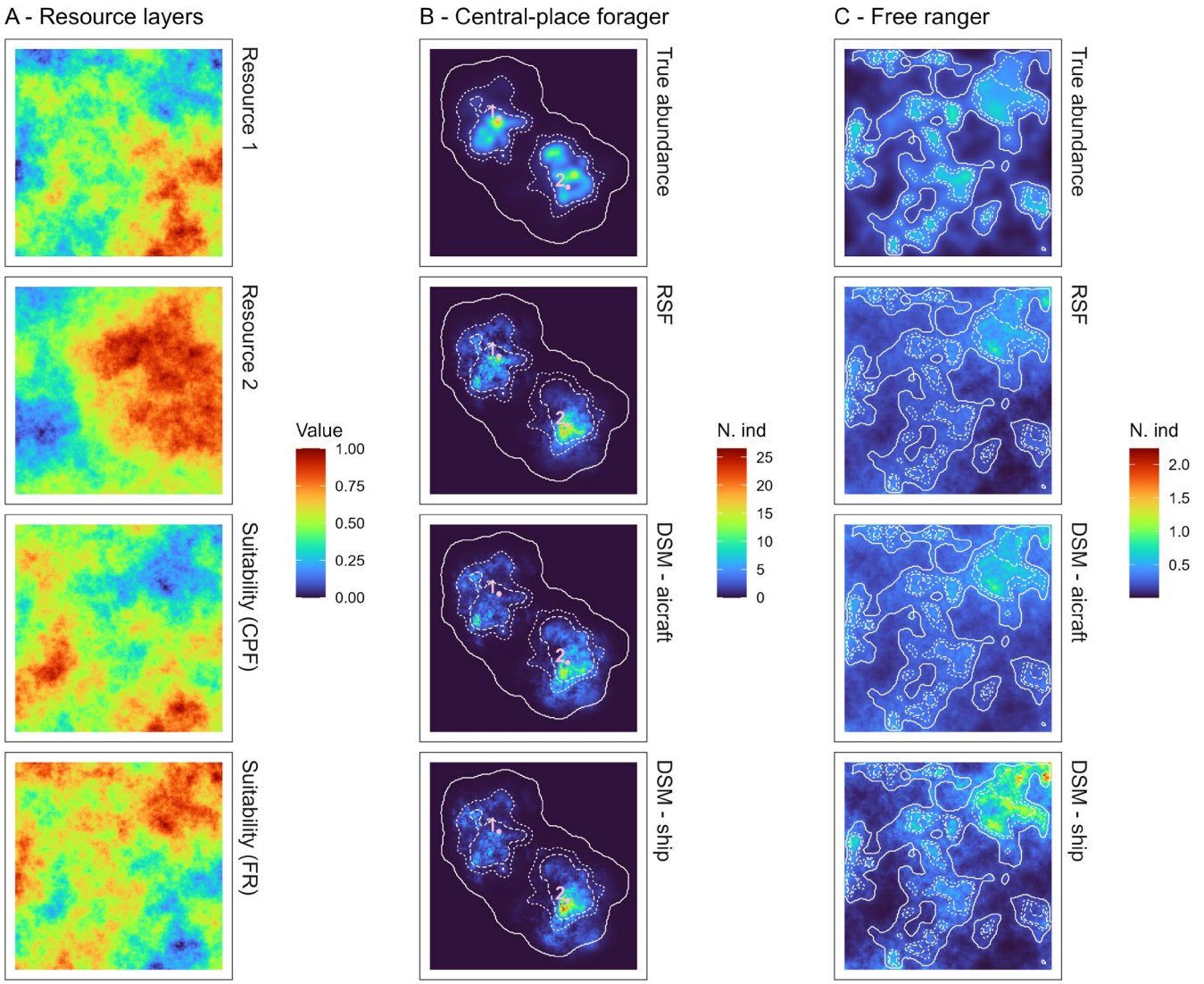

FIGURE 2 | (A) The virtual environment simulated for the study, showing from top to bottom the spatial distribution of Resources 1 and 2, the environmental suitability for the central-place forager (CPF) and the free ranger (FR). A greater value indicates a greater abundance of resources or suitability. All maps were normalised. (B) and (C) from top to bottom: True abundance maps, abundance maps predicted from the Resource Selection Function (RSF) model, abundance maps predicted from the Density Surface Model (DSM) computed either from aircraft survey data or from ship survey data, for the Central-place forager (B) and the Free ranger (C). Colour codes in (B) and (C) indicate the number of individuals. In (B) and (C), the solid white lines delimit the area that contains 100% of the population, whereas the dotted lines and the dashed lines indicate the area that delimit 80% and 50% of the population, respectively. Colonies in (B) are indicated by numbers.

TABLE 1 | Angle and step parameters used to simulate movements of the CPF.

Movement type	Colony	Angle	Step
Departing the colony	Colony 1	Von Mises ($\mu = 270$, $\kappa = 10$)	Gamma (scale = 0.9, rate = 3)
	Colony 2	Von Mises ($\mu = 90$, $\kappa = 10$)	
Travelling movement	Colonies 1, 2	Von Mises ($\mu = 0$, $\kappa = 20$)	Gamma (scale = 0.7, rate = 3)
Foraging movement	Colonies 1, 2	Von Mises ($\mu = 0$, $\kappa = 0.5$)	Gamma (scale = 0.2, rate = 3)

from these positions. An individual was considered to have returned when it was within 1 km of the colony.

2.1.2.2 | **Free Ranger.** A similar approach was used for the FR (using the `simulate_trajectory_FR` function), simulating the movements of 8000 independent individuals (i.e., not moving in coordinated groups). On the first day, the initial position of each individual was randomly selected within the study area. On the second day, individuals started to move from their last position recorded during the previous day. Starting bearings and steps were randomly drawn from Von Mises' and Gamma distributions, centred on 0 and 4.5, respectively.

Movements were simulated for each day independently. Individual movements were simulated following the same logic as for the CPF, although without returning to the colony. The suitability threshold used to cause a switch from travelling to foraging and the residence time were identical to those used for the CPF. However, we used a larger residence radius for the FR (2 km) than the CPF (1 km).

2.1.2.3 | Analysis, Formatting and Sampling Tracking Data. We quantified the track durations and the proportions of time spent travelling and foraging for each species and colony separately. The reference abundance map for the CPF population was quantified as follows: (1) we estimated the utilisation distributions (UDs, i.e., home range; using the `adehabitatHR` package; Calenge & contributions from Scott Fortmann-Roe 2023) for each individual, and normalised the map so that the sum of its values equalled one (Morris, Proffitt, and Blackburn 2016); (2) we averaged these individual UDs for each colony; (3) we multiplied the colony-specific maps by the number of individuals in each colony and (4) we summed both colony-specific maps to provide the reference ('true') map of abundance, which provides the average number of individuals present in each cell at any moment. The reference map for the FR was obtained in a similar way: the normalised individual UDs were averaged for the entire population and multiplied by the number of individuals in the population.

2.1.3 | Surveys

The surveys were simulated using the `generate_survey_plan` and `launch_survey_on_movement` functions. The virtual aerial survey was set up with 21 parallel transects (each 80 km in length), which were segmentised into 1-km-long segments. The survey spanned 1 day (from 6:00 to 16:16) and transects were sampled at 160 km/h (a 1-km-long segment was sampled in 22 s), which is the typical aircraft speed in aerial surveys (Lambert et al. 2019). The observation area around transects was set to 1 km^2 (a 500 m wide strip on each side). All individuals whose trajectories crossed a segment of the survey during the period it was sampled were considered available for detection by the survey as long as the centroid of the matching step was within the observation area. We stored the distance between the step centroid and the transect line for each individual. The survey was carried out using a line-transect protocol, where the probability for an animal to be detected is a function of its distance to the track line. This distance-dependent detection probability was modelled using a half-normal distribution with a 200 m effective strip half-width (ESW; Buckland et al. 2015). The number of individuals sighted per survey segment was computed as the sum of individuals detected within each segment.

A similar approach was used to simulate the ship survey. Since ships are slower than aircrafts (10 knots, or 18.5 km/h; a 1-km-long segment was sampled in 3.2 min), we set up eight parallel transects with a length of 50 km each. The ship survey was conducted over 2 days (4 transects sampled per day). The same logic as for the aerial survey was used to carry out the ship survey, but the ESW was set up to 500 m (ESW is typically greater for ship surveys, when compared to aerial surveys).

2.2 | Lagrangian SDMs

For the Lagrangian modelling, we considered that 1% of the population was tracked per colony (for the CPF: 80 individuals for Colony 1, 120 individuals for Colony 2; for the FR: 80 individuals), with devices sampling one position every 15 min. This sampling rate is intermediate to rates commonly used in marine tracking studies, where positions are frequently sampled at 2–30 min intervals.

Based on these sampled tracks, we mapped space-use by using an RSF based on a use-availability design (presence-background; Fieberg et al. 2021; Northrup et al. 2022). We modelled the distribution using the environmental variables (the two resource layers) instead of using the suitability map directly. This was done to simulate realistic situation, where knowledge of the exact suitability of the environment is rarely available.

The first step was to define available points from the environment to which the track locations (used points) could be compared. For the CPF, we randomly sampled 10 available points for each location point, within a buffer around the colony of origin. The latter was colony-specific and constructed by adding 5 km to the maximum distance between the furthest sampled track point and the colony. For the FR, available points were sampled from the whole study area. Geographic coordinates of each point, used/available, were used to extract the values for Resource 1 and Resource 2 (using `extract` from the `terra` package; Hijmans 2023). For the CPF, we also computed the distance to the colony of origin for every point (`st_distance` from the `sf` package; Pebesma 2018).

We fitted Generalised Additive Models (GAMs, using the `mgcv` package; Wood 2011) with 'usage' as the response variable and the binomial family as link function. 'Used points' were given a weight of 1, whereas 'available points' received a weight of 5000. The three environmental conditions (Resource 1, Resource 2, Distance to Nearest Colony) were included as covariates for the CPF, while only the first two conditions were included for the FR. To avoid overfitting, the curve complexity was restrained to a maximum of 4 degrees of freedom. We retrieved the smoothed terms for each component of the linear predictors to draw the responses of the species to environmental conditions (see `predict.gam` and `plot.gam` functions in `mgcv`).

The spatial distribution of the probability of usage was predicted from each model and normalised so that the sum of its values equalled one (Morris, Proffitt, and Blackburn 2016). For the CPF, this prediction was computed for each colony separately and only within the colony-specific buffers used to sample availability points. Hence, accessibility constraints are taken into account (cells outside the buffers were assigned a value of zero, since no individual used these areas). For the FR, a single prediction was computed for the entire study area.

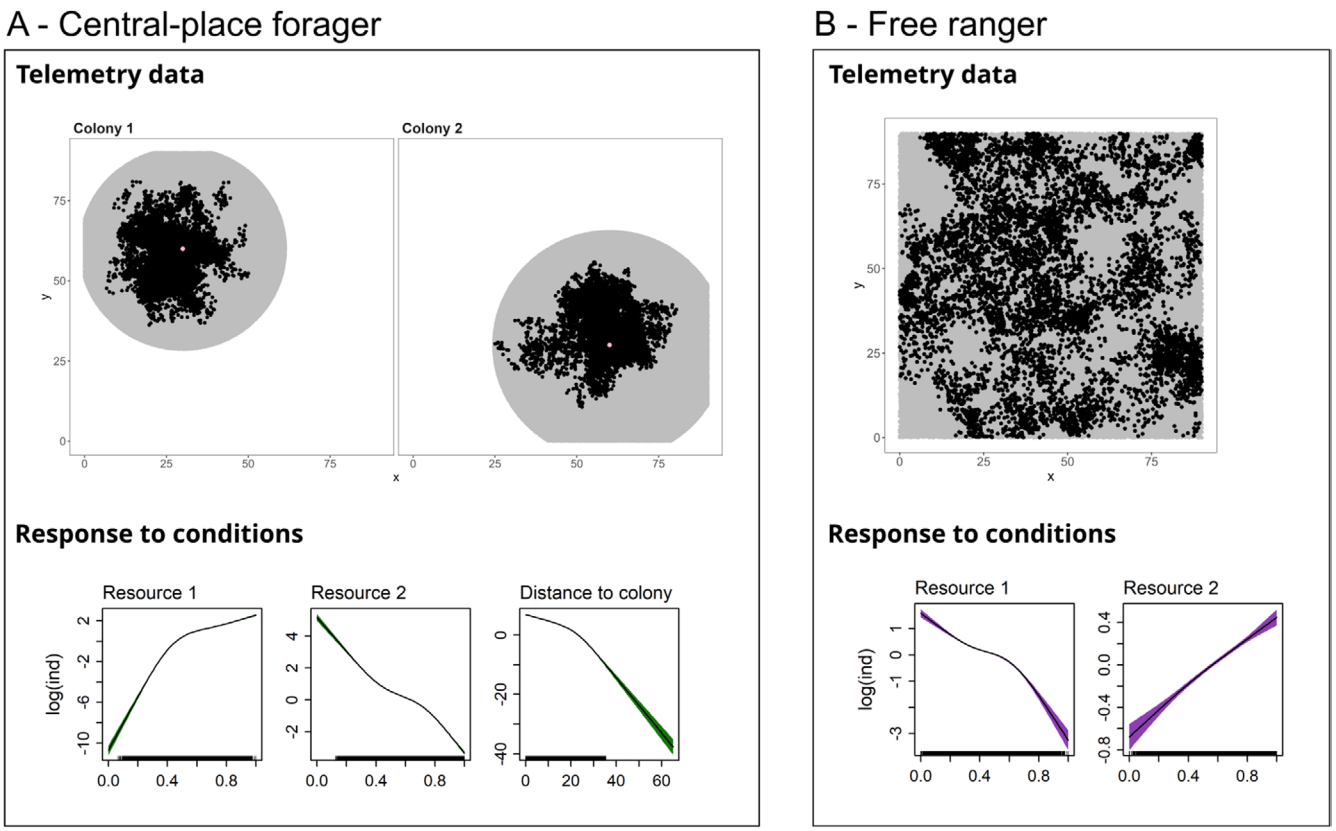

FIGURE 3 | Lagrangian perspective results for the central-place foraging (CPF; A) and the free-ranging (FR; B) species. (Top) The locations sampled from the tracks are shown in black (used locations), whereas the locations sampled from the available environment are shown in grey. For the CPF, available points were sampled within a buffer zone around the colony (see text), whereas for the FR, they were sampled over the entire study area. (Bottom) Responses to conditions estimated with the RSF for both species, showing the relationships between the linear predictor (log(usage)) on the *y*-axis and the covariates on the *x*-axis. The rugs on the *x*-axes indicate the data used to fit the model (note the ranges of sampled values for each variable differ between species).

We converted these maps into abundance maps by assuming that population sizes were known from an analysis independent of the data at hand. Abundance maps were obtained by multiplying each prediction by the number of individuals living within the considered population (Colonies 1 and 2 separately for the CPF; the whole population for the FR). For the CPF, the final abundance map was subsequently obtained by summing the maps from both colonies. These final species-scale maps represent the average number of individuals that occur in each cell at any moment.

2.3 | Eulerian SDMs

We used a classical DSM to estimate the spatial distribution of abundance (Miller et al. 2013). Environmental conditions associated with each survey segment were retrieved based on their centroid coordinates (using `st_centroid` from the `sf` package and `extract` from the `terra` package). Because the colony of origin for sighted individuals is unknown during an observation survey, we used the distance to the closest colony as a proxy (from the segment centroids).

We fitted a GAM using the count of individuals per segment as the response variable and the Tweedie distribution as link function. We included the sampled area per segment as an offset ($2 \times \text{ESW}$), where the ESW was estimated with Conventional Distance Sampling (CDS; Miller, Rexstad, et al. 2019). As for the RSF, we included the three environmental conditions (Resource 1, Resource 2, Distance to Nearest Colony) as covariates for the CPF, but only the first two for the FR. To avoid overfitting, the curve complexity was limited to a maximum of 4 degrees of freedom. We retrieved the smoothed terms for each component of the linear predictors to draw the responses to environmental conditions. The spatial distribution of abundance was subsequently predicted from these models for each species, and the total population abundances were computed as the sum of all cells from the study area.

2.4 | Assessing Prediction Accuracy

We evaluated the accuracy of model predictions with Pearson's correlation coefficients (with the `rcorr` function from the `Hmisc` package; Harrell Jr 2024) between all pairs of models, across data types and surveys, as well as against the true abundance distributions (derived from the UD, see Section 2.1.2) and the environmental suitability maps. To quantify the dependency of species space-use on environmental conditions, we also calculated Pearson's correlation coefficients between

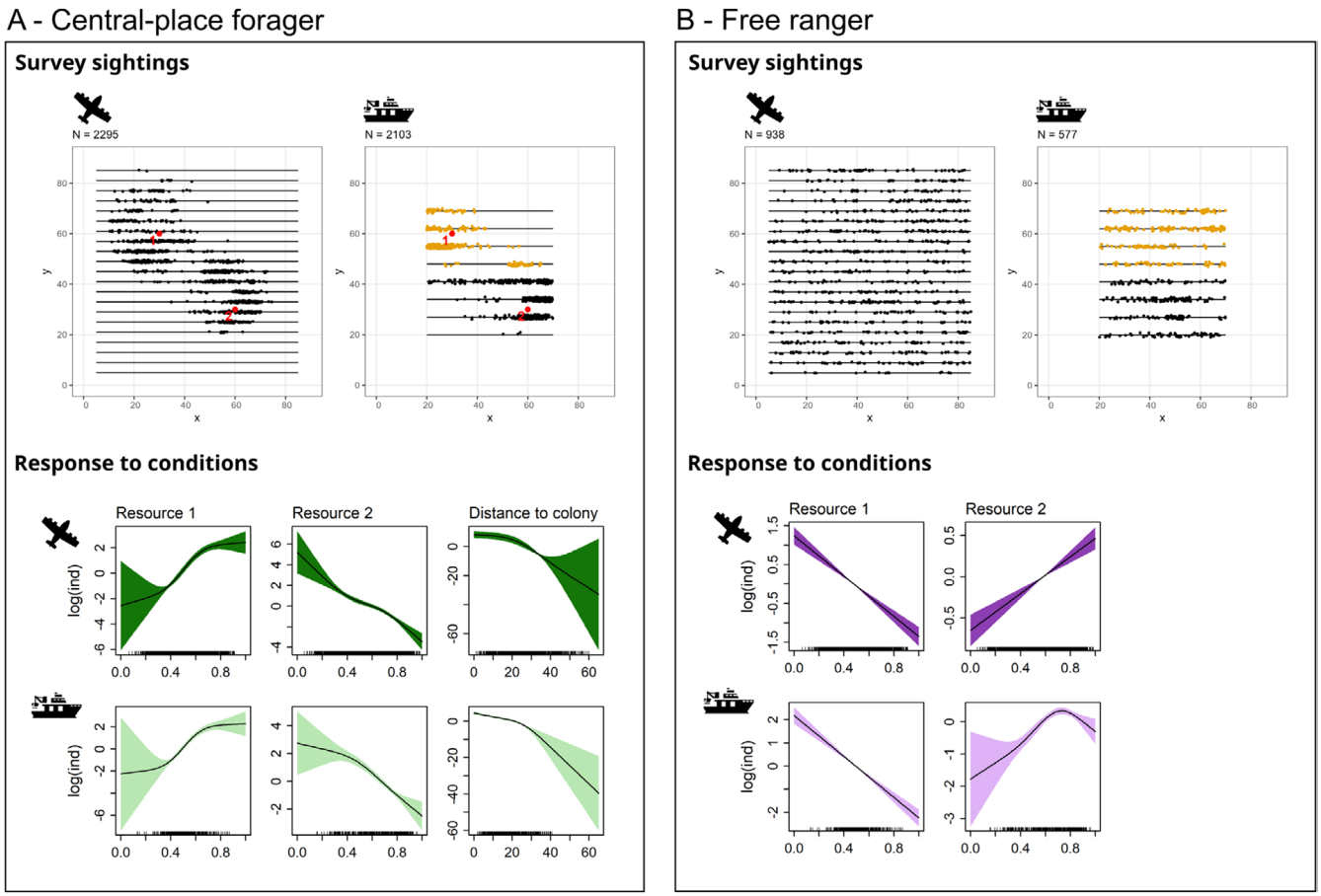

FIGURE 4 | Eulerian perspective results for the central-place foraging (CPF; A) and the free-ranging (FR; B) species. (Top) The aerial (left) and ship (right) survey designs are illustrated, where transects are depicted as black lines, and sightings are overlaid as black dots (first day of survey), and in the case of the ship-based survey, as golden dots for the second day of survey. For the CPF, colony locations are indicated by red dots. The total number of individuals sighted per survey is indicated on the top-left of each graph. (Bottom) The responses to environmental conditions estimated with the DSM for both species are shown, for aircraft (top) and ship (bottom) surveys. The relationship between the linear predictor (log(individuals); *y*-axis) and the covariates (*x*-axis) is shown. The rugs on the *x*-axes display the data used to fit the model (note the ranges of sampled values for each variable differ between aircraft and ship surveys).

the environment suitability maps and the true abundance distributions.

3 | Results

3.1 | Simulated Individual Trajectories

The simulated individual movements and surveys for each species are presented in detail in the Data S1. Both CPF and FR individuals spent more time travelling than foraging. CPFs from Colony 1 spent more time travelling than individuals from Colony 2 (Figure S3), while the activity budget (time spent travelling and foraging) of individuals from both colonies did not differ between the days. By contrast, FRs were relieved from the need to commute and spent more time foraging on day 2 than on the first day because a greater number of individuals had reached favourable areas (Figures S1–S4). The same patterns were found in our movement data analysis, which sampled 1% of the population every 15 min (Figures S5–S8).

The distribution of abundance differed between species (Figure 2B,C). Abundance of the CPF was centred around the two colonies, with most individuals from Colony 1 (top left in Figure 2B) spread to the south of the colony, whereas individuals from Colony 2 (bottom right in Figure 2B) spread to the north of their colony. The distribution of the FRs was widespread, with high abundance patches scattered throughout the study area (Figure 2C).

3.2 | Lagrangian SDMs

The variance explained by RSF models differed greatly between species (13.6% for the CPF, 1.1% for the FR). All variables were significant in each case ($p < 2 \times 10^{-16}$) and the orientation of responses to environmental conditions were correctly identified for both species (Figure 3). Predicted abundance maps were consistent with the true patterns (Figure 2B,C, second and first rows, respectively). The CPF model correctly predicted the aggregation around the two colonies, but overestimated the abundance to the north of Colony 1, whereas it underestimated the abundances to the north of the Colony 2. The FR model correctly identified high abundance areas, but tended to underestimate the absolute values in the southern part of the study area, while overestimating the northern part.

3.3 | Eulerian SDMs

The aerial survey sighted 2295 CPFs and 938 FRs, whereas the ship survey yielded 2103 CPF and 577 FR detections (Figure 4, Figures S1–S12). The vast majority of detected individuals were sighted only once. However, some individuals were sighted several times throughout the survey (Figure S15; up to four times for the FR during the ship survey).

For the CPF, aircraft- and ship-based surveys achieved similar explained deviances for the CPF (66.8% and 66.5%, respectively). They also correctly estimated the orientation of the responses to environmental conditions (Figure 4A), with all relationships being highly significant ($p < 2 \times 10^{-16}$). Both models predicted a similar distribution of abundance in the study area (Figure 2B,C, third and fourth rows), which matched well the true distribution of the species (Figure 2B,C, top row). However, they overestimated the abundance to the north of Colony 1 and to the south of Colony 2 while underestimating abundance to the north of Colony 2 (so did the RSF). Both models estimated the absolute true abundance of the species correctly (18,620 ± 2614 individuals for aircraft-based model, 17,471 ± 3784 individuals for ship-based model; values are mean ± standard deviation). By contrast, CDS estimation (Figure S16) largely overestimated abundance for ship survey, while this is not the case for the aerial survey.

Results for the FR differed somewhat. Here, the aircraft-based model explained a larger proportion of the deviance than the ship-based survey (29.2% vs. 7.7%, respectively). Yet, both models found the two covariates to be highly significant ($p < 2 \times 10^{-16}$) and also successfully identified the responses to environmental conditions (though the ship-based model identified an inflection point for Resource 2 at greater values; Figure 4B). The predicted abundance distributions were rather similar (Figure 2C, third and fourth rows) and matched the true distribution (Figure 2C, top rows). However, the ship-based model locally overestimated the abundance in the north-eastern part of the study area. Nevertheless, this latter model achieved better results in reproducing the spatial gradients in abundance, especially for areas where the species was absent. The aircraft-based model slightly overestimated the absolute true abundance (9259 ± 517 individuals), whereas the ship-based model was correct (8811 ± 970 individuals). One should note that the uncertainties associated with the estimated abundances were lower for the FR than the CPF species (Figure S16).

3.4 | Assessing Prediction Accuracy

Pearson's correlations were high between RSF and DSM, all exceeding 0.8 for both the CPF and the FR (0.89–0.94 for the CPF; 0.90–0.98 for the FR; Figure 5). Correlation intensity with true abundance and with environmental suitability, however, differed between species. For the CPF, the RSF and DSM reached correlation values with the true abundance map of about 0.65 (0.61–0.71), while correlations with the suitability layer were poor (0.11–0.19). Furthermore, the correlation between true abundance and environmental suitability was also fairly low (0.05) for the CPF, which is due to the constraints imposed by central-place foraging that drive most of the spatial patterns

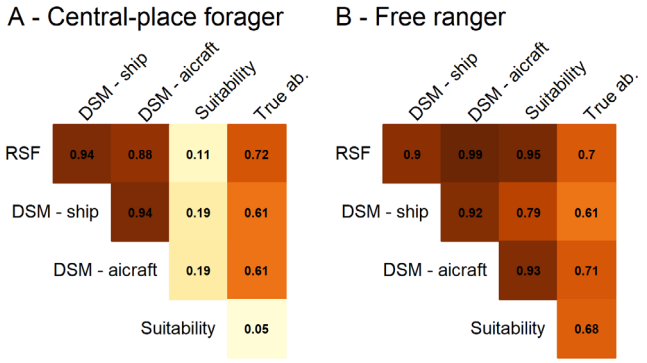

FIGURE 5 | Pearson's correlation coefficients between pairs of predicted and species-specific suitability and true abundance maps, for the central-place forager (A) and the free ranger (B). Correlation intensity is depicted by colour shading (a darker shade representing a greater correlation coefficients) and actual values.

with respect to abundance. By contrast, for the FR, true abundance and environmental suitability were highly correlated (0.68), and the RSF and DSM prediction maps reached high correlation levels with both the suitability (0.79–0.94) and the true abundance maps (0.69–0.71).

4 | Discussion

In this study, we investigated whether models built from Lagrangian and Eulerian data provide consistent information about species abundance distribution for highly mobile species. To address these issues, we conducted a simulation exercise, reproducing the whole SDM analysis process. We (1) simulated individual movements for two populations and (2) conducted dedicated surveys to monitor various parts of the system. To assess species abundance distributions, we subsequently analysed collected data, following well-established statistical methods. Our results provide important insights into the performance and reliability of SDM techniques.

Identifying the spatial distribution of environmental suitability is the paramount goal for which SDMs have been conceived (Franklin 2010; Elith and Leathwick 2009). The agreement between predicted and true distributions has been previously confirmed by simulation-based studies (Koshkina et al. 2017; Lambert and Virgili 2023). However, these studies focused on stationary individuals or individuals whose movements occurred at spatial scales smaller than the observation process (i.e., home ranges were smaller than the size of sampling units). Whether such agreement in distribution holds true for individuals that move at a larger scale relative to that of the survey has so far remained an open question.

In real life, individual movement decisions depend upon many intrinsic and extrinsic factors, from individual energy balance to inter-specific competition. The consequence is that the actual choice of favouring a particular site seldom depends exclusively on the site's quality. For example, mobile individuals in the marine realm rely on memory to target patches where they have been successfully finding resources during previous trips (Robira et al. 2021). They might also use social learning to find

high-quality patches in such dynamic environment (Thiebault, Mullers, Pistorius, Meza-Torres, et al. 2014; Thiebault, Mullers, Pistorius, and Tremblay 2014). These processes have the potential to reduce search time considerably. Hence, individuals may spend more time within high-quality patches. They might also avoid particular patches that are of high environmental quality, but where biotic conditions are unfavourable.

Since individual trajectories rely less on the encountered environmental conditions, the true drivers of movement choice are generally not directly accounted for in spatially-explicit models (but see the particular case of step selection functions (SSF); Michelot, Blackwell, and Matthiopoulos 2019). This results in a blurring of the signal encompassed in the data. Consequently, SDMs typically predict the distribution of favourable habitats or the potential abundance/usage according to habitat characteristics, rather than the absolute abundance or space-use. Hence, such a model will predict the potential abundance or usage at a given site based on its characteristics, while the actual, realised abundance or usage depends on a multitude of other factors (Franklin 2010; Elith and Leathwick 2009; Peterson 2011). The less the movement of a species relies on environmental suitability, the harder it will be to estimate such suitability with a SDM.

In our simulated system, individuals were naive about resource locations, and no interaction (neither positive nor negative) occurred between individuals. Yet, we reproduced the separation between environmental suitability and site use by introducing stochasticity in the movement decision process for both species and by also adding a spatial constraint on the possible movement patterns of the CPF species. This approach successfully decoupled environmental suitability and space-use, as indicated by the low correlation value between the suitability map and the abundance distribution of the CPF (Figure 5). For this species, the need to return to the colony strongly constrained the distribution of individuals. By contrast, correlation values for the FR were considerably higher, since its distribution was more directly linked to environmental suitability.

Despite this, both Lagrangian and Eulerian SDMs successfully retrieved the orientation and intensity of the species response to environmental conditions (Figures 3 and 4). Correlation coefficients with true abundance distributions were also high, indicating that large-scale movements of species have little impact on model reliability. Predictions from Lagrangian and Eulerian SDMs were well correlated (0.8–0.98), confirming the results from previous studies that used real data. With camera traps, Popescu, de Valpine, and Sweitzer (2014) and Bassing et al. (2023) found that estimates agreed well between the two model types for most tested species. However, Ferrer-Ferrando et al. (2023) reported contrasting results. When they found was a poor correlation between the two data types, the divergence likely was a consequence of the models selecting different predictors, which resulted in different predicted distributions. Such differences may arise from how the used data actually sampled the environment (Braun et al. 2023). Carroll et al. (2019) and Phillips et al. (2019) have specifically examined the consistency between distributions inferred from movement data and from distance sampling surveys. In both cases, the consistency depended on the temporal and demographic alignment of data types. The two models provided more diverging estimates, when tracking data did not exactly cover the survey period (e.g., one data set included the pre-migration period, whereas the other did not) or when the survey sampled a wider demographic subset of the population of interest. This latter aspect is of utmost importance for species for which tracking data focus on a small subset of the population, such as adult breeders or one sex, and where different groups of the population show different space-use patterns (Pettex et al. 2019).

Such mechanisms, however, were of no relevance in our simulation study, where all individuals behaved similarly, following the same movement principles and drivers. The good match between data types we obtained suggested that most of the divergence observed between models in real life may arise from misalignment of the sampled population subsets rather than from statistical inconsistency between the two types of approaches. Therefore, when considering the coupling of predictions from SDMs of heterogeneous nature for conservation management purposes, these misalignments should be explicitly incorporated to ensure accuracy.

When combining Lagrangian and Eulerian data, one should also consider the potential impact of data-specific statistical biases. For Lagrangian data, most biases typically arise from the unrepresentative nature of the data (low sample size; population subsets) and the construction of availability points (Fieberg et al. 2021). When built on Eulerian data, SDM accuracy is directly affected by many factors, such as species rareness, observation protocol or detection bias (Virgili et al. 2018; Lambert et al. 2024). Although, in our simulation study, we modelled an ideal system, we tested two different survey configurations. This resulted in a different sampling regime in space and time (aerial vs. ship survey) and demonstrated that both models were capable to reliably retrieve and predict abundance maps, at least for abundant species and without further bias. However, real-life Eulerian data are often subject to detection bias (Buckland et al. 2015; Lambert et al. 2019; Hammond et al. 2021). This might be due to availability issues (when individuals are present but cannot be detected by observers because they are resting in dens or are diving, e.g.) or to perception issues (when individuals are present and available for detection, but circumstances prevent their perception by the observer; e.g., sun glare). These statistical limitations can explicitly be accounted for in state-of-the-art SDMs (Buckland et al. 2015; Lambert et al. 2019), but it remains unclear to which degree this might alter the capacity of SDMs to identify the true spatial pattern of a species.

The uncertainty in the capacity of Lagrangian and Eulerian data to provide similar information on the abundance distribution process of species was one of the main impediments to their effective combination. However, such combination might be important to enable informed management decisions, assess anthropogenic threats, plan conservation strategies or assess ecosystem status. It is also important for the integration of these data in abundance-orientated iSDMs. By confirming that distributions predicted by both SDM types are similar when alignment of data types is ensured, our study paves the way for effectively building such iSDMs. Such achievement would considerably advance our knowledge of species distribution

patterns at broad scales by overcoming the drawbacks associated with each data type.

Author Contributions

C.L. conceived the ideas, designed the protocol, conducted the analyses and led the writing of the manuscript. A.-S.B.-L. assisted in analysis conduction and interpreting the results. D.G. and A.-S.B.-L. contributed significantly to drafting the ideas and writing the manuscript, and gave final approval for publication.

Acknowledgements

We warmly thank Julien Collet for kindly sharing pieces of code without which this simulation exercise would have not come to life. We are also indebted to Dr. Manfred Enstipp for the thorough English editing of this manuscript. No fieldwork was involved in this work.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All R codes to generate the data used in the study and replicate the analyses are available at Zenodo: https://zenodo.org/records/11195482.

References

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

Bassing, S. B., M. DeVivo, T. R. Ganz, et al. 2023. "Are We Telling the Same Story? Comparing Inferences Made From Camera Trap and Telemetry Data for Wildlife Monitoring." *Ecological Applications* 33: e2745.

Braun, C. D., M. C. Arostegui, N. Farchadi, et al. 2023. "Building Use-Inspired Species Distribution Models: Using Multiple Data Types to Examine and Improve Model Performance." *Ecological Applications* 33: e2893.

Buckland, S., D. Borchers, T. Marques, and R. Fewster. 2023. "Wildlife Population Assessment: Changing Priorities Driven by Technological Advances." *Journal of Statistical Theory and Practice* 17: 20.

Buckland, S., E. Rexstad, T. Marques, and C. Oedekoven. 2015. *Distance Sampling: Methods and Applications*. New York: Springer.

Calenge, C., and S. Fortmann-Roe. Contributions from 2023. "Adehabitathr: Home Range Estimation r package version 0.4.21." https://CRAN.R-project.org/package=adehabitatHR.

Carroll, M. J., E. D. Wakefield, E. S. Scragg, et al. 2019. "Matches and Mismatches Between Seabird Distributions Estimated From At-Sea Surveys and Concurrent Individual-Level Tracking." *Frontiers in Ecology and Evolution* 7: 333.

Courbin, N., L. Pichegru, M. Seakamela, et al. 2022. "Seascapes of Fear and Competition Shape Regional Seabird Movement Ecology." *Communications Biology* 5: 208.

DeCesare, N. J., M. Hebblewhite, F. Schmiegelow, et al. 2012. "Transcending Scale Dependence in Identifying Habitat With Resource Selection Functions." *Ecological Applications* 22: 1068–1083.

Elith, J., and J. R. Leathwick. 2009. "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time." *Annual Review of Ecology, Evolution, and Systematics* 40, no. 1: 677–697.

Ferrer-Ferrando, D., J. Fernández-López, R. Triguero-Ocaña, P. Palencia, J. Vicente, and P. Acevedo. 2023. "The Method Matters. A Comparative Study of Biologging and Camera Traps as Data Sources With Which to Describe Wildlife Habitat Selection." *Science of the Total Environment* 902: 166053.

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: 1027–1043.

Fletcher, R., T. Hefley, H. Robertson, B. Zuckerberg, R. McCleery, and R. Dorazio. 2019. "A Practical Guide for Combining Data to Model Species Distribution." *Ecology* 100: e02710.

Franklin, J. 2010. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge: Cambridge University Press.

Guillera-Arroita, G., J. J. Lahoz-Monfort, J. Elith, et al. 2015. "Is My Species Distribution Model Fit for Purpose? Matching Data and Models to Applications." *Global Ecology and Biogeography* 24: 276–292.

Hammond, P., C. Lacey, A. Gilles, et al. 2017. "Estimates of Cetacean Abundance in European Atlantic Waters in Summer 2016 From the SCANS-III Aerial and Shipboard Surveys." Technical Report Wageningen Marine Research.

Hammond, P. S., T. B. Francis, D. Heinemann, et al. 2021. "Estimating the Abundance of Marine Mammal Populations." *Frontiers in Marine Science* 8: 735770.

Harrell, F. E., Jr. 2024. With Contributions from Charles Dupont, Many Others. "Hmisc: Harrell Miscellaneous R package version 5.1–2." https://CRAN.R-project.org/package=Hmisc.

Hays, G. C., H. Bailey, S. J. Bograd, et al. 2019. "Translating Marine Animal Tracking Data Into Conservation Policy and Management." *Trends in Ecology & Evolution* 34: 459–473.

Hijmans, R. J. 2023. "terra: Spatial Data Analysis R package version 1.7-18." https://CRAN.R-project.org/package=terra.

Ketz, A. C., T. L. Johnson, R. J. Monello, et al. 2018. "Estimating Abundance of an Open Population With an N-Mixture Model Using Auxiliary Data on Animal Movements." *Ecological Applications* 28: 816–825.

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.

Lahoz-Monfort, J. J., G. Guillera-Arroita, and B. A. Wintle. 2014. "Imperfect Detection Impacts the Performance of Species Distribution Models." *Global Ecology and Biogeography* 23: 504–515.

Lambert, C. 2024a. "Energyscapes: Toolbox For Energyscape Analyses R Package Version 0.1.0." https://github.com/CLambert1/energyscapes.

Lambert, C. 2024b. "Virtualecologist: Tools For Simulating Virtual Ecosystems R Package Version 0.1.0." https://github.com/CLambert1/virtualecologist.

Lambert, C., M. Authier, G. Dorémus, et al. 2019. "The Effect of a Multi-Target Protocol on Cetacean Detection and Abundance Estimation in Aerial Surveys." *Royal Society Open Science* 6: 190296.

Lambert, C., J. G. Cecere, F. De Pascalis, and D. Grémillet. 2024. "Correcting Detection Bias in Mapping the Abundance of Marine Megafauna Using a Mediterranean Seabird as an Example." *ICES Journal of Marine Science* 81: fsae058.

Lambert, C., and A. Virgili. 2023. "Data Stochasticity and Model Parametrisation Impact the Performance of Species Distribution Models: Insights From a Simulation Study." *Peer Community Journal* 3: e34.

Marshall, C., G. Glegg, and K. Howell. 2014. "Species Distribution Modelling to Support Marine Conservation Planning: The Next Steps." *Marine Policy* 45: 330–332.

Matthiopoulos, J., J. Fieberg, and G. Aarts. 2020. *Species-Habitat Associations: Spatial Data, Predictive Models, and Ecological Insights*. Minneapolis, USA: University of Minnesota Libraries Publishing.

Matthiopoulos, J., E. Wakefield, J. W. Jeglinski, et al. 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., 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.

Miller, D. L., M. L. Burt, E. A. Rexstad, and L. Thomas. 2013. "Spatial Models for Distance Sampling Data: Recent Developments and Future Directions." *Methods in Ecology and Evolution* 4: 1001–1010.

Miller, D. L., E. Rexstad, L. Thomas, L. Marshall, and J. L. Laake. 2019. "Distance Sampling in R." *Journal of Statistical Software* 89: 1–28.

Morris, L. R., K. M. Proffitt, and J. K. Blackburn. 2016. "Mapping Resource Selection Functions in Wildlife Studies: Concerns and Recommendations." *Applied Geography* 76: 173–183.

Mueller, T., and W. F. Fagan. 2008. "Search and Navigation in Dynamic Environments—From Individual Behaviors to Population Distributions." *Oikos* 117: 654–664.

Northrup, J. M., E. Vander Wal, M. Bonar, et al. 2022. "Conceptual and Methodological Advances in Habitat-Selection Modeling: Guidelines for Ecology and Evolution." *Ecological Applications* 32: e02470.

Palmer, M. S., K. M. Gaynor, J. A. Becker, J. O. Abraham, M. A. Mumma, and R. M. Pringle. 2022. "Dynamic Landscapes of Fear: Understanding Spatiotemporal Risk." *Trends in Ecology & Evolution* 37: 911–925.

Panigada, S., O. Boisseau, A. Cañadas, et al. 2021. "Estimates of Abundance and Distribution of Cetaceans, Marine Mega-Fauna and Marine Litter in the Mediterranean Sea from 2018–2019 Surveys (Monaco: ACCOBAMS - ACCOBAMS Survey Initiative Project)." 177.

Pebesma, E. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *R Journal* 10: 439–446. https://doi.org/10.32614/RJ-2018-009.

Peterson, A. T. 2011. *Ecological Niches and Geographic Distributions* Number 49 in Monographs in Population Biology. Princeton: Princeton University Press.

Pettex, E., C. Lambert, J. Fort, G. Dorémus, and V. Ridoux. 2019. "Spatial Segregation Between Immatures and Adults in a Pelagic Seabird Suggests Age-Related Competition." *Journal of Avian Biology* 50: e01935.

Phillips, E. M., J. K. Horne, J. E. Zamon, J. J. Felis, and J. Adams. 2019. "Does Perspective Matter? A Case Study Comparing Eulerian and Lagrangian Estimates of Common Murre (*Uria aalge*) Distributions." *Ecology and Evolution* 9: 4805–4819.

Popescu, V. D., P. de Valpine, and R. A. Sweitzer. 2014. "Testing the Consistency of Wildlife Data Types Before Combining Them: The Case of Camera Traps and Telemetry." *Ecology and Evolution* 4: 933–943.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Robira, B., S. Benhamou, S. Masi, V. Llaurens, and L. Riotte-Lambert. 2021. "Foraging Efficiency in Temporally Predictable Environments: Is a Long-Term Temporal Memory Really Advantageous?" *Royal Society Open Science* 8: 210809.

Steenweg, R., M. Hebblewhite, J. Whittington, P. Lukacs, and K. McKelvey. 2018. "Sampling Scales Define Occupancy and Underlying Occupancy–Abundance Relationships in Animals." *Ecology* 99: 172–183.

Thiebault, A., R. Mullers, P. Pistorius, et al. 2014. "From Colony to First Patch: Processes of Prey Searching and Social Information in Cape Gannets." *Auk* 131: 595–609.

Thiebault, A., R. H. Mullers, P. A. Pistorius, and Y. Tremblay. 2014. "Local Enhancement in a Seabird: Reaction Distances and Foraging Consequence of Predator Aggregations." *Behavioral Ecology* 25: 1302–1310.

Truchy, A., E. Göthe, D. G. Angeler, et al. 2019. "Partitioning Spatial, Environmental, and Community Drivers of Ecosystem Functioning." *Landscape Ecology* 34: 2371–2384.

Virgili, A., M. Authier, P. Monestiez, and V. Ridoux. 2018. "How Many Sightings to Model Rare Marine Species Distributions." *PLoS One* 13: e0193231.

Winter, V. A., B. J. Smith, D. J. Berger, et al. 2024. "Forecasting Animal Distribution Through Individual Habitat Selection: Insights for Population Inference and Transferable Predictions." *Ecography* 2024: e07225.

Wood, S. N. 2011. "Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 73: 3–36.

Zurell, D., U. Berger, J. S. Cabral, et al. 2010. "The Virtual Ecologist Approach: Simulating Data and Observers." *Oikos* 119: 622–635.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.